

SIDD: A Framework for 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 mitigate 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 a protected network. We propose a multilevel framework called SIDD (Sensitive Information Dissemination Detection) system which is a high-speed transparent network bridge located at the edge of the protected network. SIDD consists of three main components: 1) network-level application identification, 2) content signature generation and detection, and 3) covert communication detection. Further, we introduce a model implementation of the key components, demonstrating how our system can be deployed. Our approach is based on the application of statistical and signal processing techniques on traffic flow to generate signatures and/or extract features for classification purposes. The proposed framework aims to address methods to detect, deter and prevent deliberate and unintended distribution of sensitive content outside the organization using the organization's system and network resources by a trusted insider.

1. Introduction

An insider attack describes the damage that can occur to the interests of an organization by a trusted individual with legitimate access to its network and system resources. Such an attack can occur through an inadvertent security breach by an authorized user, a

planned security breach by an authorized user, or by an outsider through a compromised system. The planned insider attack can result in the exfiltration or destruction of sensitive data or it can compromise the communications network and various network servers and resources.

In today's widely-connected network environments, a successful insider attack could result in serious damage to the interests of an enterprise [1-3]. Government sectors, which provide access to classified information to authorized personnel, are vulnerable to insider attacks. The private sectors, such as drug companies that maintain valuable and highly-sensitive proprietary information, and banking institutions that manage the flow of monetary transactions, are also vulnerable to insider attacks. In all of these instances, the consequences of an insider attack are dire, resulting in the loss of revenue, public distrust, as well as legal ramifications. Furthermore, with greater accessibility and availability of sensitive information, insider attacks are likely to increase [1].

Compared to external threats, insider threats are more dangerous, devastating, and challenging to detect and prevent since trusted individuals have access privileges, know the networks, and also have specific information they wish to exfiltrate [4]. Within the broader goals of mitigating insider attacks, our work only addresses the detection, deterrence and prevention of deliberate and unintended distribution of sensitive content outside of the organization using the organization's system and network resources by a trusted insider. In particular, we will focus on the outbound network traffic detection rather than inbound network traffic (We refer to traffic traveling from the "protected network" to the "external network" as outbound, and to traffic in the other direction as inbound.). The key technical challenge is to detect

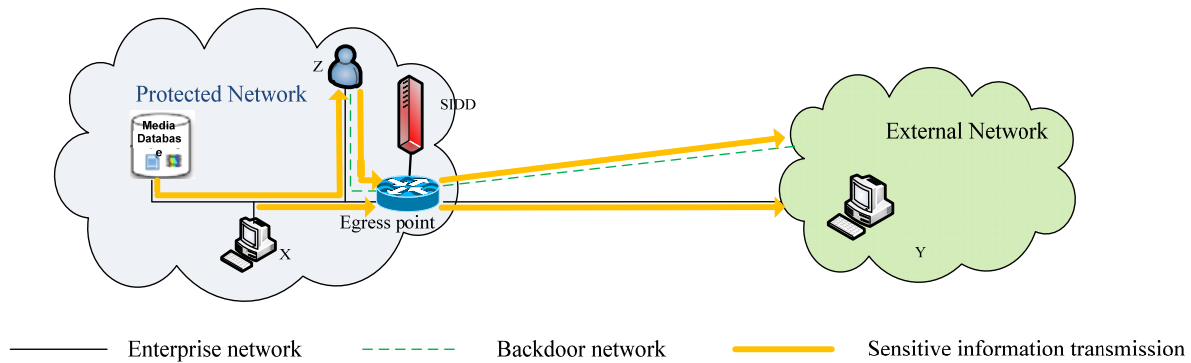


Figure 1. A motivating example for sensitive data exfiltration and detection.

these attacks despite transformations applied to the content, e.g. encryption, modulation by the communication protocol, slight modifications for the purpose of averting detection, and hiding of the content using steganography tools.

In this paper, we propose a multilevel system, called SIDD (Sensitive Information Dissemination Detection), to detect the dissemination of sensitive information by an insider. Unlike previous work which seeks to deal with specific methods that an insider can use to exfiltrate data (e.g., [3]), the proposed multilevel structure is a unified framework for detecting the occurrence of information leakage. We present the implementation of the key components of proposed framework, which rely upon statistical and signal processing techniques applied to outbound traffic flows to generate signatures and/or extract features. These signatures are then used for analysis by the system to detect and prevent the leakage of sensitive information. In addition, our scheme can apply to both anomaly and misuse detection [5]. Note that contrary to misuse detection which can only detect attacks by comparing outbound traffic activity against the expected actions of an intruder, anomaly detection approaches must first determine the normal behavior of the object being monitored, and then use deviations from this baseline to detect possible attacks. Therefore, following the anomaly detection approach, we assume the existence of general patterns in outbound traffic which can be used to characterize the normal traffic behavior and identify anomaly insider attacks.

The paper is organized as follows. Section 2 provides example scenarios of insider threats and scopes the problem domain that we seek to address. In Section 3, we describe the architecture and the functionality of each phase of SIDD. In Section 4, we describe an example approach to implementing each component of SIDD system to illustrate the requirements in developing the system. Section 5

concludes the paper and points to the future research direction.

2. Motivating examples and design space

We have designed a SIDD system to ensure that information leaving a protected network has been approved for release. In this section, we first describe motivating examples for creating such a system by examining two threat scenarios. Then, we define the key elements of general insider threat problems that we take into consideration when developing the SIDD system.

2.1. Example threats

Consider the following scenario (see Figure 1): A company X outsources its customer service to another company Y by establishing a shared (enterprise) network connecting their corporate LANs. In the service process, X needs to provide proprietary documents and manuals to Y, but it does not wish to share some sensitive/proprietary information. A malicious insider Z seeks to 1) create backdoor networks to enable loss or damage of protected information and 2) exfiltrate sensitive/proprietary information using the enterprise's network resources.

To avoid detection, Z may transform the original content using compression, translation to another format, or encryption. Z might also use a higher bandwidth mechanism such as a peer-to-peer (P2P) [6] network to leak information. P2P protocols are usually blocked by firewalls on enterprise networks due to their known use for illegal file sharing and identity theft. However, P2P communication can be tunneled over an authorized protocol (e.g., HTTP) to cloak/mask its identity in order to penetrate the firewall [7]. In addition, a more clandestine way to conduct espionage over the enterprise network is to use a covert channel

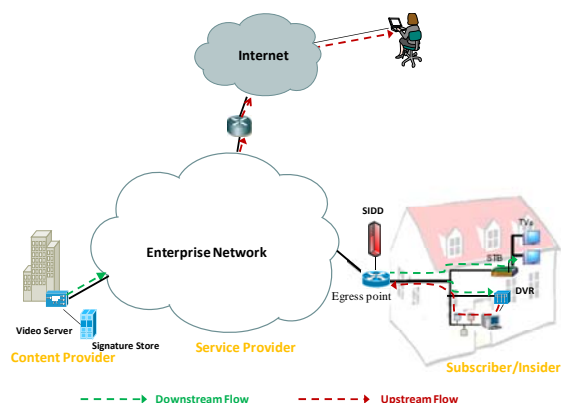


Figure 2. Sensitive data exfiltration and detection in content redistribution scenario.

to hide communication. For example, a communication protocol such as Skype [8] can be used to hide bits of information into the payload of the communication. While Skype samples and encodes voice calls into packets for transmission, a portion of the bits can be altered using steganography techniques [9] to embed information without any perceptual impact on the quality of the voice call. This covert channel can be used to send command and control messages, secret keys, as well as to exfiltrate sensitive content.

Another example of threat is content redistribution scenario (see Figure 2). Television service providers are rapidly deploying IPTV network [10] to deliver a wide variety of video content to subscribers over the Internet. Using Set-Top Boxes (STBs) or personal computer with enterprise network access, encrypted video content may be delivered to an authorized subscriber (insider) through an on-demand service request or live broadcast. The downloaded video stream may be legally decrypted upon receipt for viewing and stored in a Digital Video Recorder (DVR). Then, the subscriber can redistribute and pirate the decrypted content to millions of unauthorized viewers. Consequently, content providers risk loss of revenue and the authored content suffers losses from infringement of copyright laws and other restrictions.

2.2. Scope of the insider threat

To address the above threats, our goal is to architect a system that detects and prevents the leakage of sensitive information from a protected network to an external network. We illustrate our approach with two examples. However, there are many variations to these types of attack. Here we highlight the key elements that in our view are critical due to the varieties of threats, different forms of communication channels, insider types, or content types.

2.2.1. Communication channel. To disseminate sensitive information to someone outside the corporate network, the insider might use overt, tunneled, or covert communications. *Overt communication* is intended legitimate communication that is open, observable, and identifiable (e.g., HTTP, IRC). Additionally, privacy of its contents may be preserved from unauthorized users using communication protocols such as HTTPS, SFTP, and SSH. *Tunneled communication* is unauthorized communication (that would ordinarily be blocked) tunneling over an authorized overt channel. Additionally, its features are cloaked to masquerade as legitimate communication (e.g., P2P over HTTPS). *Covert communication* is clandestine communication embedded into the header or/and payloads of an overt channel using steganography techniques to hide the fact that communication is occurring.

2.2.2. Content type. Sensitive information ranges from personal identifiable information (e.g., social security numbers, credit card numbers, etc.) to intellectual property. This information could be contained in a static file (e.g., image, software program, and excel spreadsheet) or a multimedia session (e.g., telephone conversation and video conference). Sensitive information may be leaked to an outsider in its original, modified, or hidden format. Content in its *original* format has not been modified in any way. *Modified* content includes content that may be compressed, padded, encoded into a new file type, or encrypted. *Hidden* content includes content that has been embedded into other content or the communication protocol using steganography techniques.

2.2.3. Insider type. There are different types of insiders that result in information leakage: inadvertent, intentional, and malicious. An *inadvertent insider* is a trusted person with access to sensitive information who inadvertently discloses sensitive information. An *intentional insider* is a trusted person who knowingly discloses sensitive information and is aware of the security that it is purposefully bypassing. This person may try to manipulate the content or use overt communication that preserves privacy to avoid detection. A *malicious insider* is also a trusted person who knowingly discloses sensitive information. However, in addition to manipulating the content or using overt communication like an international insider, tunneled or covert communication is usually employed to avoid detection.

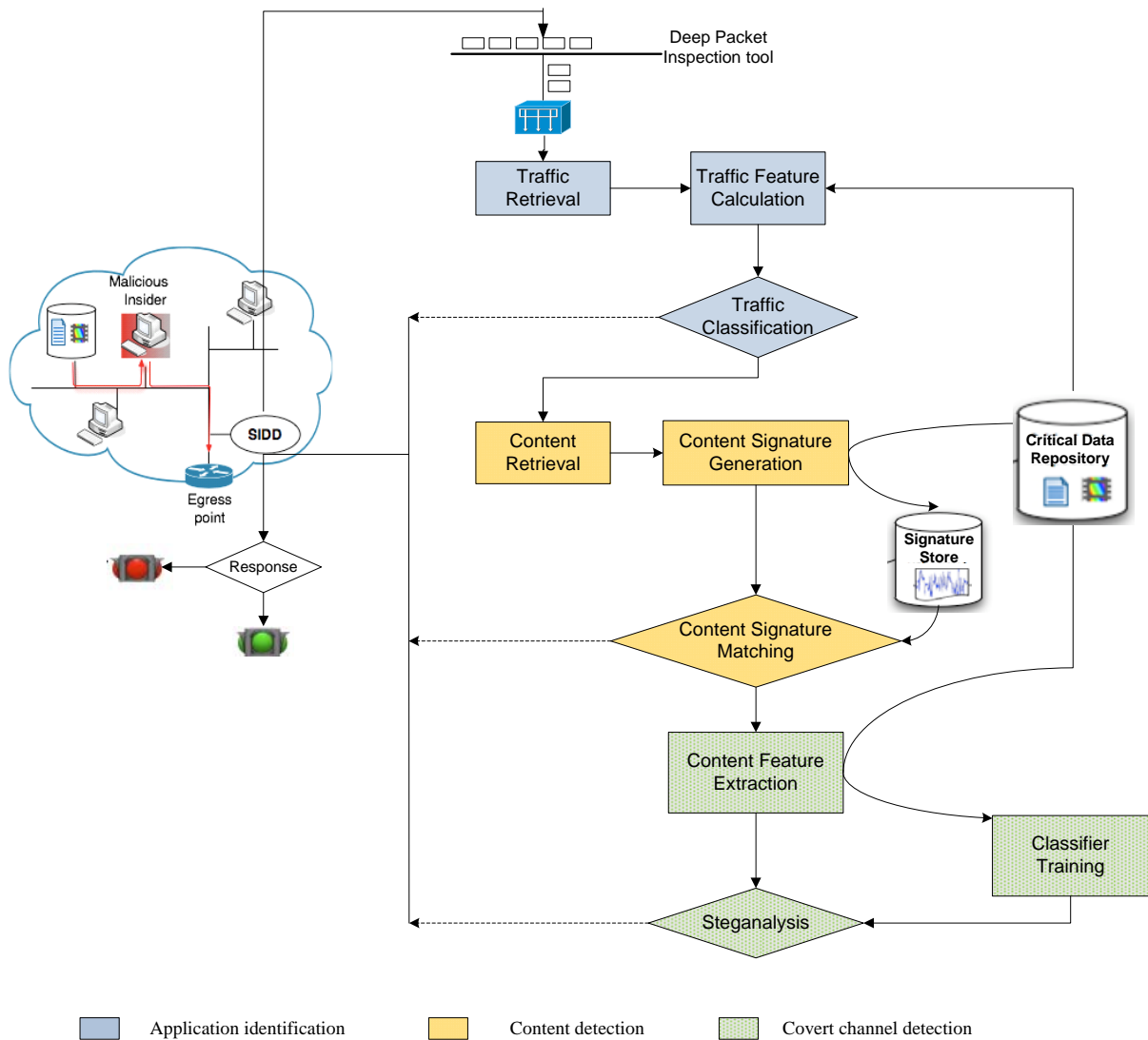


Figure 3. Sensitive Information Dissemination Detection (SIDD) system.

3. Structure of SIDD system

Figure 3 shows the architecture of the proposed SIDD system and its deployment in the network. To detect and prevent the leakage of sensitive information, SIDD is placed at the network egress to monitor the traffic flow outbound from the protected network. It provides an interface to network management to define the exit policy and take a responsive action (i.e. generate alarm, drop packets, flag offending traffic for investigation, etc.).

The Critical Data Repository shown in the Figure 3 represents sensitive content (e.g., files, audio, video, images, etc.) that needs protection. To begin, SIDD

retrieves traffic off the network via a commodity Deep Packet Inspection (DPI) tool. The captured network traffic is filtered into the application identification system (phase 1) to extract traffic features. Some traffic characteristics, such as the temporal patterns and sizes of packets, will be used to instantiate different application features. By these means, SIDD allows proper policing of application traffic to dictate what traffic is allowed across the network boundary. If the system confirms that the traffic is coming from a potential dangerous application, it may automatically exit and take a responsive action. Note that, in this stage, SIDD can also provide knowledge about the structure and semantics of the communication, type of the system, type of the client, network connectivity,

etc., all of which are useful attributes for the successive detection stages.

After the traffic flow passes through the application identification checking process, it will pass into the Content Retrieval process and the content of application will be analyzed by the content detection stage (phase 2). The Content Signature Generator is used to generate the content signatures of sensitive content which are stored in the Signature Store. It will also generate the signature for content traversing the network to be compared using the matching algorithm with the stored signatures in order to detect dissemination of sensitive content. Similar to the first phase, if the SIDD system confirms a match in content detection phase, it may automatically exit and take a responsive action. Otherwise, the retrieved content will pass to the covert channel detection process (phase 3) to explore covert communication establishment.

To detect the existence of covert communication, some statistical features that can capture natural correlation inherent in media will be generated. The resulting features and the system classifier, which has been trained with the same features stored in the critical data repository, will be input to perform Steganalysis [11] to determine the presence of hidden information in the target content. In this final step, SIDD may only be able to detect the presence of hidden content and not fully recover the content for comparison. In this case, the traffic can be labeled as suspicious to warrant further investigation of the host and end-user that generated the traffic.

Overall, the SIDD system provides mechanisms for reviewing objects to determine if the object is approved for release. When the internal network traffic passes through, SIDD parses the flow and waits for the server response. It may perform up to three-phase checks in a response time to determine how to filter outgoing traffic to prevent exfiltration of sensitive information.

4. Methodology and performance

In Section 3, we described the key challenges and identified the multiple components of a tool that can be used to detect sensitive data exfiltration by insider attacks. In this section, we provide the detailed algorithm for each component of SIDD and some preliminary experiments to evaluate the algorithms.

4.1. Application identification

The goal of application identification phase is to identify the application generating traffic based on characteristics of its flow. Voluminous aggregated traffic, use of encryption, and use of cloaking software

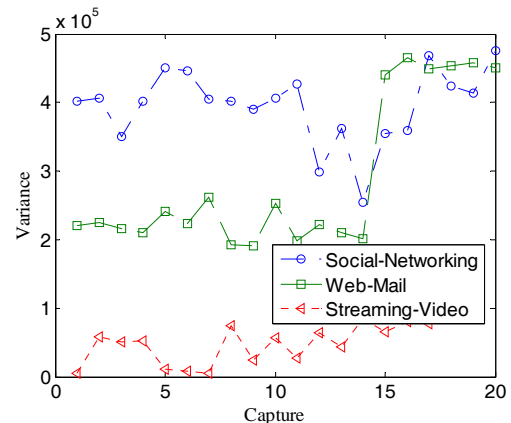


Figure 4. Identification of Web application based on the variance of incoming packet sizes.

to avoid identification make the analysis and classification of traffic a challenging task [12, 13]. Research has been done in the identification of network protocols such as SSH, HTTP, FTP, and SMTP [12-15]. As a result, some existing algorithms and tools can be easily implemented in SIDD to allow proper policing of application traffic to dictate what traffic is allowed across the network boundary. In particular, content signature and identification leverages application identification as a filtering mechanism and to gain awareness about the structure and semantics of the communication, type of the system, type of the client, network connectivity, etc.

On the other hand, the traffic tunneling over the same protocol is indistinguishable, which may lead to potential security and performance issues. For example, HTTP has become the universal entry mechanism to the corporate network, penetrating through firewall protection. However, HTTP is a very vague classification since numerous applications are built on top of it. Unlike classical techniques [12-15], which have focused on course-grained categorization of traffic (HTTP vs. FTP), we seek to make a finer distinction within the broader categories (hotmail vs. gmail – both applications using HTTP). To classify applications we can leverage time domain, frequency domain and wavelet-based [16] analysis on attributes of the network traffic (such as the packet inter-arrival time and packet size,) generated by the application.

In order to accurately identify the overriding applications in a given network flow of outgoing traffic from a given host, we investigate the characteristics of different applications over the same protocol. Specifically, the captured network traffic with the DPI tool will perform Traffic Retrieval to get some useful information by gleaning the packet headers instead of requiring information from the payload. This is

important since our system can now work even when the payload is encrypted. The retrieved traffic will go to Traffic Feature Calculation block to obtain a set of features that remain relatively constant and unique to a particular type of web application. Particularly, different signals derived from the network flow, e.g., the packet inter-arrival times and packet sizes, will be extracted from the network flow. We have investigated if characteristics, such as the temporal patterns and sizes of packets, can instantiate a signature with a high degree of confidence. Figure 4 plots traffic features (the variance of incoming packet sizes) generated from incoming packet size of three different classes of applications, namely, social networking (MySpace and Facebook), web-mail (Gmail and Hotmail), and streaming video applications (Youtube and Veoh) running on HTTP. These categories cover the majority of the applications that are prevalent in today's networks and therefore are good representatives for the application examples. The particular statistic exhibits consistent behavior across different captures. More importantly, the results show that the applications do indeed have discernable characteristics for all the three analysis techniques. After sufficiently characterizing the traffic categories mentioned above, a Traffic Classification system will be able to automatically identify the applications of flows on a network.

Note that our technique can help to solve application identification problems in two different ways. On the one hand, it can be used to identify the application generating the traffic based on characteristics of its flow without relying on content level information. On the other hand, it can help to identify the application even if it attempts to mask its features by tunneling over another application or even encrypted protocol, e.g., POP over SSH. Although this paper reports only preliminary results related to the first class of applications, the methodology can be applied to the second class of problems as the dissimilar traffic behaviors driven from different applications are very useful for the detection of protocol tunneling problems [17].

4.2. Content signature generation and detection

Application identification can provide some useful information which allows proper policing of application traffic to dictate what traffic is allowed across the network boundary. However, deep analysis of the flow content may be necessary for detection, deterrence, and prevention of deliberate and unintended distribution of sensitive content outside the organization by generating a legal application channel

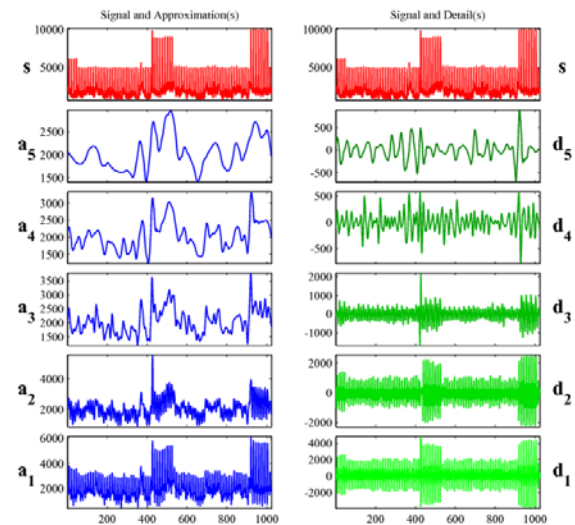


Figure 5. Wavelet-based video signature for the signal s of “Star wars IV”.

or using of an existing enterprise connection created by a trusted insider.

The goal of content signature generation and detection stage is to create network traffic based content signatures and a method to compare signatures to detect exfiltration of sensitive/protected content. Specifically, some content information such as volume of packets, will be operated on packets to first perform Content Retrieval. The retrieved content information is treated as a time series which allows us to use signal processing techniques and tools to analyze these signals. The premise of Content Signature Generation is to create a signature based on the content itself [18] such that attempts to modify the content to evade detection render the content useless to the outsider at the receiving end. Additionally, compact signatures will be developed to support faster detection especially for larger volume content.

In our work, we use video redistribution as an example. To create unique, compact, and robust signatures for video content, we extract a time series (or signal) s from the size of frames of a video stream. A key issue of using the content itself as a signature is that of scalability arising from the size of the signature. As a 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 [16] 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 which 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. Figure 5 illustrates the coefficients of the wavelet transform applied to a portion of the video traffic from the Star Wars IV movie [19]. The approximation and detailed coefficients are shown for up to 5 levels.

Another important focus of our work is on the Content Signature Matching technique. One 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. With the unique, compact, and robust signatures generated by the Content Signature Generation block, some statistics distance, i.e., cross-correlation, will be adopted to estimate the degree to which two series are correlated. This technique is effective when the monitored content is identical to the sensitive information in the media database and even when the monitored content has some distortion during the transmission process. For example, various possible video traffic rate adaptation scenarios [20] resulting from network congestion or server overload may cause the video traffic and its signatures captured from the network to differ from the ones saved in the signature store. The insider may also evade detection by transcoding [21] the video stream and/or injecting noise which may impact the short-term correlation. Figure 6 shows the robustness of the algorithm measured by the receiver operating characteristics (ROC) curve, which is a plot of the false positive rate versus the false negative rate in one rate-adaption case (see Figure 6). Here N is the length of a captured video clip by frames (the video clips is around 5 minutes when $N = 8192$). These preliminary results show that wavelet-based signatures have good performance both in terms of false positive and false negative probabilities compared to direct perform comparison of the signal in the time domain.

4.3. Detecting covert communication

Covert communication is clandestine communication which hides the fact that communication is indeed occurring. For example, steganography is a form of covert channel to hide the presence of communication through the embedding of a secret message in an innocuous carrier medium, such as digital audio, image, and video. Instead of direct exfiltration of sensitive content, exfiltration of sensitive information may be performed by using covert. The content may be hidden spatially, temporally, or even in the transform domain spaces; and the content may be encrypted before it is embedded in a carrier. By slightly altering the binary sequence of the distributed sensitive content samples with existing steganography

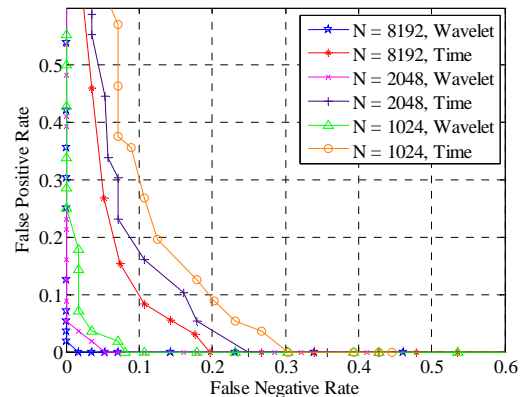


Figure 6. Comparison of Receiver Operating Characteristics (ROC) curves based on the time-domain signal and its wavelet-based signatures.

tools, the covert channel may be relatively easy to establish. The insider may also use the dedicated communication channel (of the enterprise network) to set up covert communication and then distribute sensitive content. For example, some features of the TCP/IP protocol suite can be used to send covert messages [22] and a timing-based covert channel can be set up by varying the rate of sending jobs [23]. Therefore, one key aspect of the SIDD system is the development of methods that can be used to detect the presence of covert communication channels. Particularly, some statistical features that can capture the change of the media/traffic characteristics by the distortion introduced with hidden information are extracted in the Content Feature Extraction process. As the data embedding method is typically unknown prior to detection, the signature needs to be independent of the steganography algorithms used. The defined feature vector will be used in classifier training with the available media files or communication information in critical data repository. The test content/communication will go through the same procedures of feature selection and then used to judge the existence of covert communication.

In general, there are many different types of carriers: 1) structured carrier such as networking protocols, and computer languages such as XML, 2) unstructured carriers such as still images, audio, streaming video, and natural languages, 3) dynamic/real-time carriers such as a live real-time multimedia session (i.e., VoIP session), and 4) timing carriers in which the information is hidden through the timing of packet arrivals. In this paper we assume unstructured carriers without loss of generality. In

particular, we focus on detecting digital audio medium as covert channels. We believe that the techniques presented here could be applied in detecting some other types of covert channels also.

With the rapid proliferation of Voice over Internet Protocol (VoIP) and other Peer-to-Peer (P2P) audio services such as Skype, covert channels using digital audio may be relatively easy to establish [9]. 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 make audio a good candidate for use as a “cover” for covert communications by hiding secret messages and also for exfiltration of sensitive and protected content. Therefore, in our work, we use audio steganalysis as an example to explain how to perform covert channel detection. Note that, although we use audio files as the target carrier medium, the methods can be generalized to other types of media files.

As hiding information in digital media requires alterations to the signal’s properties as well as introducing some form of degradation that can be measured by some standard audio quality metrics, distortion measures have been shown to be effective to test the presence of hidden messages. In our work, we propose an audio steganalysis scheme that measures audio content distortion using Hausdorff Distance [24]. Unlike previous work in audio steganalysis that used the traditional audio quality metrics [25], 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 covert communication (modifications to pure audio content) as follows.

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 segmentation¹, we apply wavelet decomposition to both x and x' to generate wavelet coefficients at different levels of resolution. Next, Hausdorff distances [26] 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 differences between known pure content and the modified audio files with covert channels.

Figure 7 plots the correct classification rates as a function of the hidden ratio with numerous audio

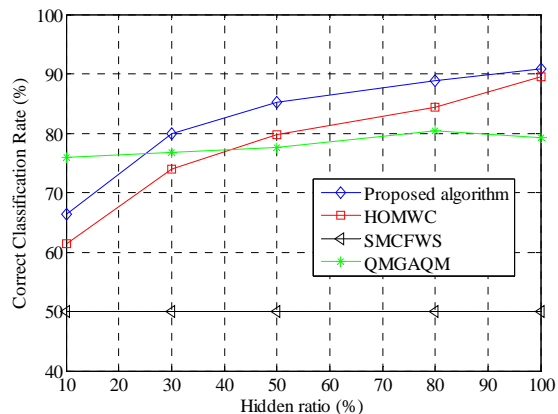


Figure 7. Comparison of correct classification rate among audio steganalysis algorithms.

sequences and compares the classification rate with three different reference algorithms, i.e., High-Order Moments of the Wavelet Coefficients of the audio signal (HOMWC) [27], Statistical Moments of the Characteristic Functions of Wavelet Sub-bands (SMCFWS) [28], Quality Measurement with General Audio Quality Metrics (QMGAQM) [25]. The result shows 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 can be easily extended to other steganography tools and algorithms.

5. Conclusion and discussion

Insider attacks are most damaging to the interests of an enterprise and government organizations. Mitigating insider attacks requires addressing multiple problems. This paper developed a systematic approach to address the key problems of detecting the exfiltration of sensitive content. Particularly, a multilevel framework that composed of application detection, content signature generation and detection, and covert channel detection was proposed. Based on our preliminary results, we feel encouraged that we will be able to significantly extend the state-of-the-art in addressing a critical problem in network security.

Although the experimental results are encouraging, the following are some of the issues that need to be addressed. First, in this work, we constrain ourselves to one specific case when providing the preliminary results in each component of the framework, i.e., 3 web-based applications for application identification, video redistribution detection for content signature generation and audio steganalysis for covert channel detection. The methodology and testing need to be

¹ To get a good local distortion estimation, the audio file is split into small segments so that the distortion analysis can be performed separately on the individual segments.

extended to other types of problems at each category, which will make them more applicable in practice. Second, the algorithm measurement is applied to different components of SIDD independently. Systematic performance reporting on a real prototype system that is subjected to rigorous tests employing a broad range of exfiltration attempts need to be investigated.

6. References

- [1] M. R. Randazzo, M. Keeney, E. Kowalski, D. Cappelli, and A. Moore, "Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector", CERT and the National Threat Assessment Center, Aug. 2004.
- [2] E. D. Shaw, K. G. Ruby, and J. M. Post, "The insider threat to information systems: The psychology of the dangerous insider", Security Awareness Bulletin, vol. 2-98, pp. 27-46, Sept. 1998.
- [3] L. Spitzner, "Honeypots: catching the insider threat", Proceedings of 19th Annual Computer Security Applications Conference, pp. 170-179, Dec. 2003.
- [4] N. Nguyen, P. Reiher, G.H. Kuenning, "Detecting insider threats by monitoring system call activity", IEEE Information Assurance Workshop, pp. 45-52, June 2003.
- [5] P. Gaonjur and C. Bokhoree, "Risk of Insider Threats in Information Technology Outsourcing: Can deceptive techniques be applied?", School of Business Informatics, University of Technology, Mauritius, 2006.
- [6] A. Oram, "Peer-to-Peer: Harnessing the Power of Disruptive Technologies", O'Reilly & Associates, 2001.
- [7] Y. Zhao, N. Li, and C. Wang., "Remote wireless transmission and error recovery of log data", Applied Geophysics, vol. 4, pp. 308-312, 2007.
- [8] skype, <http://www.skype.com>
- [9] J. Dittmann, D. Hesse, and 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, pp. 607-618, Magdeburg, 2005.
- [10] U. Jennehag and T. Zhang, "Increasing bandwidth utilization in next generation IPTV networks", Proceedings IEEE International Conference on Image Processing (ICIP), pp. 2075-2078, Oct. 2004.
- [11] N.F. Johnson, and S. Jajodia, "Steganalysis of images created using current steganography software", David Aucsmith (Ed.): Information Hiding, LNCS 1525, pp. 32-47. Springer-Verlag Berlin Heidelberg, 1998.
- [12] L. Bernaille, R. Teixeira, I. Akodkenou, A. Soule, and K. Salamatian, "Traffic classification on the fly", ACM SIGCOMM Comput. Commun. Rev. vol. 36, no. 2, pp. 23-26, 2006.
- [13] G. Bissias, M. Liberatore, D. Jensen, and B. Levine, "Privacy Vulnerabilities in Encrypted HTTP Streams", Proceedings of Privacy Enhancing Technologies Workshop, pp. 1-11, 2005.
- [14] T.T.T. Nguyen and G. Armitage, "A Survey of Techniques for Internet Traffic Classification using Machine Learning", to appear in IEEE Communications Surveys and Tutorials, 2008.
- [15] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "BLINC: multilevel traffic classification in the dark", ACM SIGCOMM Comput. Commun. Rev., pp. 229-240, 2005.
- [16] C. S. Burrus, R. A. Gopinath, and H. Guo, Introduction to Wavelets and Wavelets Transforms, a Primer, Upper Saddle River, NJ, Prentice Hall, 1998.
- [17] M. Crotti, M. Dusi, F. Gringoli, L. Salgarelli, "Detecting HTTP Tunnels with Statistical Mechanisms", IEEE International Conference on Communications (ICC) '07, pp.6162-6168, 2007.
- [18] A. Hampapur and R. M. Bolle, "Comparison of distance measures for video copy detection", Proceedings of IEEE Int. Conf. Multimedia and Expo (ICME), pp. 737-740, Aug. 2001.
- [19] "Video traces for network performance evaluation", [Online]. Available: <http://www.tkn.tuberlin.de/research/trace/trace.html>
- [20] A. R. Reibman, S. Sen, and J. V. der Merwe, "Network monitoring for video quality over IP", Picture Coding Symposium, 2004.
- [21] A. Vetro, C. Christopoulos, and H. Sun, "Video transcoding architectures and techniques: an overview", Signal Processing Magazine, IEEE, vol. 20, no. 2, pp. 18-29, Mar 2003.
- [22] C. Rowland, <http://www.firstmonday.dk/issues/issue25/rowland/>
- [23] S. Lipner, "A comment on the confinement problem," Fifth symposium on Operating systems principles, pp. 192-197, Nov. 1975.
- [24] Y. Liu, K. Chiang, C. Corbett, R. Archibald, B. Mukherjee, and D. Ghosal, "A novel audio Steganalysis based on high order statistics of a distortion measure with Hausdorff distance", 11th Information Security Conference(ISC) 2008, Lecture Notes in Computer Science, Vol. 5222, pp. 487-500.

[25] H. Ozer, I. Avcibas, B. Sankur, and N. Memon, "Steganalysis of audio based on audio quality metrics", Security and Watermarking of Multimedia Contents, pp. 55-66. Santa Clara, 2003.

[26] D. P. Huttenlocher, D. Klanderman, and W.J. Rucklidge, "Comparing images using the Hausdorff distance", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp. 850-863, 1993.

[27] J.H. Farid, "Detecting hidden messages using higher-order statistical models", IEEE International Conference on Image Processing, vol. 2, pp. 905-908, 2002.

[28] Y. Q. Shi, G. Xuan, C. Yang, J. Gao, Z. Zhang, P. Chai, D. Zou, C. Chen, and W. Chen, "Effective Steganalysis Based on Statistical Moments of Wavelet Characteristic Function", IEEE International Conference on Information Technology, pp. 1195-1198, 2005.