COMPUTERS & SECURITY 30 (2011) 35–49

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Designing a cluster-based covert channel to evade disk investigation and forensics☆

Hassan Khan [a,b,*], Mobin Javed [b], Syed Ali Khayam [b], Fauzan Mirza [b]

[a] Computer Science Department, University of Southern California, Los Angeles, CA 90089 2560, USA
[b] School of Electrical Engineering and Computer Sciences, National University of Science and Technology, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 24 May 2010
Received in revised form
6 October 2010
Accepted 12 October 2010

Keywords:
Information hiding
Steganography
Covert channels
Disk forensics
Digital watermarking

ABSTRACT

Data confidentiality on a computer can be achieved using encryption. However, encryption is ineffective under a forensic investigation mainly because the presence of encrypted data on a disk can be easily detected and disk owners can subsequently be forced (by law or other means) to release decryption keys. To evade forensic investigation, intelligent information hiding techniques that support *plausible deniability* have been proposed as an alternative to encryption; plausible deniability allows an evader to hide data in a manner such that he/she can deny the very existence of the data.

In this paper, we present a new, plausible deniability approach to store sensitive information on a cluster-based filesystem. Under the proposed approach, a covert channel is used to encode the sensitive information by modifying the fragmentation patterns in the cluster distribution of an existing file. As opposed to existing schemes, the proposed covert channel does not require storage of *any* additional information on the filesystem. Moreover, the channel provides two-fold plausible deniability so that an investigator without the key cannot prove the presence of hidden information.

We derive the theoretical capacity of the covert channel and show that a capacity of up to 24 bits/cluster can be achieved on a half-empty disk. The proposed data hiding and recovery algorithms are implemented on FAT32 based disk drives and we show that the disk (read/write) access time of the algorithms is quite low as compared to the contemporary approaches. We also present statistics about the incidence of file fragmentation on actual file systems from 52 disk drives belonging to a diverse set of users. Based on these statistics, we present guidelines for selecting good cover files. Finally, we show that even if an investigator gets suspicious, he/she will incur an unreasonably high $O(m^2)$ complexity to reveal an m bit hidden message.

1. Introduction

Concealment of sensitive information on digital transmission and storage media is becoming increasingly difficult due to the growing sophistication of network and disk forensics tools. As data encryption using traditional cryptographic techniques is easily detected during forensic evaluations, information hiding provides an additional layer of security

☆ Based on the short paper "Evading Disk Investigation and Forensics using a Cluster-Based Covert Channel" by H. Khan, M. Javed, F. Mirza, S. A. Khayam which appeared in 16th ACM Conference on Computer and Communication Security (CCS), Chicago, IL, USA 2009.

* *Corresponding author*. School of Electrical Engineering and Computer Sciences, National University of Science and Technology, Islamabad, Pakistan.
E-mail addresses: hassan.khan@seecs.edu.pk, hassankh@usc.edu (H. Khan), mobin.javed@seecs.edu.pk (M. Javed), ali.khayam@seecs.edu.pk (S.A. Khayam), fauzan.mirza@seecs.edu.pk (F. Mirza).

0167-4048/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2010.10.005

over encryption by hiding sensitive information inside an innocuous medium (Petitcolas et al., 1999). In order to ensure evasion during forensic investigation, covert channels—a subclass of information hiding techniques—hide sensitive information in media that are neither designed for nor intended to transfer information (Simmons, 1984). Most of the contemporary covert channels are devised for communication protocols while a few information hiding techniques for storage devices have also been proposed (Huebner et al., 2006; Piper et al., 1984; Eckstein and Jahnke, 2005; Anderson et al., 1998; McDonald and Kuhn, 1999; Pang et al., 2003).

In this paper, we propose a novel information hiding approach to evade disk forensics. The proposed approach uses fragmentation patterns of an existing file in a filesystem to hide sensitive information. Under the proposed approach, the distribution of innocuous storage units (clusters in a filesystem) represent hidden information. Thus, even when a user is forced to release all the contents of his/her hard disk, a forensic investigator who is scanning the device for the presence of suspect information can neither detect the boundaries of hidden information nor its contents. Only a user with the correct key can extract the hidden information.

The proposed covert channel offers a significant advantage over existing schemes because it does not require storage of any additional information on the filesystem. Moreover, existing approaches provide plausible deniability by disclosing only a subset of covert files to the forensic investigator and successfully denying that any additional files exist (Anderson et al., 1998; McDonald and Kuhn, 1999; Pang et al., 2003). Since the proposed technique does not require storage of any random data on the disk, it provides an additional layer of plausible deniability. The user can deny that *any* hidden data exists on the disk, with the fragmentation being explained as a consequence of regular disk usage.

We analytically derive upper and lower bounds on the capacity of the proposed channel. We show that for the case where 50% of the clusters of the disk are occupied and 10%, 25%, and 50% of the occupied clusters belong to cover files, we are able to hide 2.194×10^4, 2.174×10^4, and 2.115×10^4 bits, respectively. While our approach is quite generic and can be extended to any cluster-based filesystem, we demonstrate a proof-of-concept of the proposed evasion technique on the widely-used FAT filesystem. We implement the data embedding and recovery algorithms on USB flash drives and show that the number and time of disk accesses (read and write) of algorithms is quite low. We also analyze fragmentation patterns on actual file systems from 52 disk drives belonging to a diverse set of users. To the best of our knowledge, this is the second largest fragmentation pattern study since Garfinkel's work (Garfinkel, 2007). During our study we gather statistical evidence that shows that our proposed covert channel can be easily embedded in majority of the disk drives without raising suspicion. Based on our analysis, we present guidelines for identifying good cover files and disk conditions. Finally, we show that the channel is secure as a forensics investigator trying to reveal m hidden bits will incur an unreasonably high $[O(m^2)]$ complexity.

The rest of this paper is organized as follows: Section 2 provides background and related work in this area. Section 3 discusses the threat model. Section 4 explains the basic concept and proposes two data hiding approaches based on the basic idea. Section 5 presents performance evaluation in terms of volatility of hidden data, capacity and access times. Section 6 provides the guidelines for selecting ideal cover files and the security of proposed covert channel. Section 7 discusses the design issues and limitations of the proposed approach. Section 8 summarizes key findings of this paper.

2. Background and related work

In this section, we first provide brief background on the information hiding concepts pertinent to this paper. We then discuss existing techniques for information hiding in filesystems. To maintain a logical flow of thought, background on cluster-based filesystems is deferred to subsequent sections.

2.1. Background

2.1.1. Information hiding versus encryption

Encryption has been the predominant method to secure sensitive information on digital transmission and storage media. However, encryption is ineffective in most forensic investigation scenarios because the presence of encrypted content can easily be detected by a forensic investigator. As an example, Fig. 1 shows the sample entropy of clusters on a USB disk containing different filetypes. It can be easily observed that clusters containing encrypted content have significantly higher entropies than those containing unencrypted content.[1] Thus a simple analysis can reveal the presence and boundaries of encrypted content to a forensic investigator. Subsequently, the disk (or traffic) owner can be compelled to release decryption keys. For instance, soldiers or intelligence officers may be tortured into revealing secret keys. Similarly, court orders may force a business organization to provide decryption keys and private data to investigating officials.

In such a scenario, the sensitive information is decrypted and compromised in its entirety. To avoid such a scenario, information hiding techniques are being proposed to evade network and disk investigation and forensics (Petitcolas et al., 1999). Information hiding techniques are classified by their evasiveness (ability to look similar to normal data) and capacity (volume of hidden information). In the evasiveness context, a good information hiding technique should not store/transmit any encrypted content.

2.1.2. Steganographic versus covert channels

While the two main information hiding classes of steganographic channels and covert channels are both designed for evasion, it is important to differentiate between them. Steganographic channels hide information in innocuous looking

[1] High entropies are observed because any effective encryption algorithm randomizes (i.e., removes the skew from) the underlying alphabet's symbol distribution in order to ensure robustness against frequency analysis.