DOCUMENT SECURITY

ABOUT THE AUTHOR

Ronald L. Mendell holds a Master of Science degree in Network Security from Capitol College in Laurel, Maryland. He also holds the Certified Information Systems Security Professional (CISSP) designation. He has also held the Certified Legal Investigator (CLI) designation from the National Association of Legal Investigators (NALI). A member of the Information Systems Security Association (ISSA), he is a Distinguished Visiting Lecturer in Network and Computer Security at Our Lady of the Lake University in San Antonio, Texas. A writer specializing in investigative and security topics, he has numerous published articles in magazines such as *Security Management* and *The ISSA Journal* with subjects ranging from business intelligence to financial investigations to computer security. This is his fourth book for Charles C Thomas Publisher, Ltd. He works for a high-tech company in Austin, Texas.

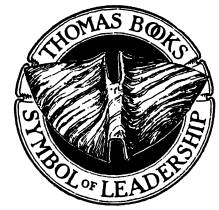
DOCUMENT SECURITY

Protecting Physical and Electronic Content

By

RONALD L. MENDELL, MS, CISSP, CLI

Master of Science in Network Security
Certified Information Systems Secutity Professional
Certified Legal Investigator
Member of the International Systems Security Association (ISSA)
Member of High Technology Crime Investigation Association (HTCIA)



CHARLES C THOMAS • PUBLISHER, LTD.
Springfield • Illinois • U.S.A.

Published and Distributed Throughout the World by

CHARLES C THOMAS • PUBLISHER, LTD.
2600 South First Street
Springfield, Illinois 62704

ISBN 978-0-398-07766-2 (hard)
ISBN 978-0-398-07767-9 (paper)

Library of Congress Catalog Card Number: 2007015249

With THOMAS BOOKS *careful attention is given to all details of manufacturing and design. It is the Publisher's desire to present books that are satisfactory as to their physical qualities and artistic possibilities and appropriate for their particular use.* THOMAS BOOKS *will be true to those laws of quality that assure a good name and good will.*

Printed in the United States of America
UB-R-3

Library of Congress Cataloging-in-Publication Data

Mendell, Ronald L.
Document security : protecting physical and electronic content / by Ronald L. Mendell.
p. cm.
ISBN 978-0-398-07766-2 (hard) -- ISBN 978-0-398-07767-9 (pbk.)
1. Computer security. 2. Computer networks--Security measures. I. Title.

QA76.9.A25M457 2007
005.8--dc22

2007015249

PREFACE

Several electronic layers exist in most documents, a fact overlooked by many writers. Probing these sublayers often reveals information not intended for release by the author. Documents in electronic formats create a "palimpsest" that even semiskilled investigators can probe for sensitive data.

Palimpsest seems like an exotic word. But literally, it means "scraped again" from the Greek word roots. In ancient and medieval Europe, writers often scraped off previous writing on a manuscript and wrote new text. (Writing media were in short supply and were expensive.) With modern forensic techniques like ultraviolet light and photography researchers uncover the original layer of writing.

Using computer forensic techniques, twenty-first century sleuths discover text and data in electronic documents thought erased by previous users. Modern electronic media are inherently palimpsestuous. Secrets become visible through metadata in documents, slack space in files, magnetic remanence, and other thorny ironies of information retention. They disclose information often, under the radar, by unintentionally making sensitive information Web-facing or not encrypting data on a laptop, which results in information leakage.

Overconfidence that one's sensitive data is not leaking through to the outside world will vex security professionals in the twenty-first century. Immense security resources go to prevent deliberate network intrusion. However, content security is not always on the forefront of security thinking. More information leaks out of organizations unintentionally than corporate America would like to think about. Many of the most recent headline-grabbers about security breaches involve documents or files leaked by a stolen laptop or by "misplaced" computer tapes or by being inadvertently Web-facing. The text identifies common pitfalls in document security and suggests remedies to prevent future headlines.

INTRODUCTION

The "hacker" culture dominated network security throughout the 1980s and 1990s. As the exploits of teenagers cracking into the systems of multibillion dollar corporations grew, basic countermeasures evolved to deal with the onslaught. As the twenty-first century arrived, the criminal sector caught on to the treasures lying in the data on those systems. While "hackers" have not disappeared, the dangerous attacks are now less thrill-motivated and more geared toward seizing valuable data.

Financially motivated crime continues to grow in cyberspace. The target is files or documents. Content, whether it be credit card numbers, social security numbers, banking information, customer lists, or trade secrets, has become "king." Some of the most notable headlines involve organizations losing databases, misplacing files or documents containing customer data, or having laptops stolen with, of course, confidential data on them.

Organized criminal rings target financial data online through a variety of schemes ranging from phishing to planting malicious code, such as Trojans, on PCs to simply researching public records available on the Web. Spies obtain proprietary data through finding Web-facing documents via search engines, and social engineering continues to trump the best of network security technology. Kevin Mitnick and Robert Schifreen acknowledge in their respective books, *The Art of Intrusion* and *Defeating the Hacker*, that social engineering often is the shortest and easiest route to most secrets.

In twenty-first century America, individuals and organizations leak information on a regular basis. In some cases, they hemorrhage data, albeit unintentionally. Protecting networks is essential, but due attention needs to go to protecting content, even when it is not residing just in electronic form on a network.

Information leakage or compromise happens in the following ways:

1. Web-facing documents contain sensitive or confidential data. Employees, however, place the documents on an "internal server," thinking the information will remain visible only within the internal network. Unfortunately, the information becomes visible to the external world through Internet access.
2. Documents undergo multiple drafts and then get sent to recipients in electronic form. Savvy readers can learn about the history of the document and even view redacted sections by accessing the metadata within the document.
3. Documents on laptops and PDAs containing sensitive data have no encryption protection, or they lack robust encryption protection. When the laptops and PDAs are lost or stolen, the critical data has little protection.
4. Storage media for documents in electronic format do not have proper markings as to content and sensitivity. Tracking procedures do not exist for the media. No encryption is in effect for the data. Such media are easily misplaced, lost, mislabeled, or stolen.
5. Documents, whether in paper, physical, or electronic format are not disposed of in a secure manner.
6. Reuse of electronic media occurs without following recommended secure procedures. Persons with a minimal understanding of computer forensics can read sensitive information remaining on the media.
7. Digital devices record all activity on the machine. Computer forensic examination recovers much of what the uninformed user thought he or she had deleted.
8. Web pages contain details about the hiring of technical staff, recent network infrastructure enhancements, and details about the enterprise's business organization. All of this available information aids corporate spies and hackers.
9. Disinformation on fraudulent Web sites compromise legitimate businesses' logos, branding, and services.
10. Credentials from business organizations can be easy to forge or to fake. These vulnerabilities permit fraud in gaining employment, in obtaining physical entry to the facilities, and in impersonating the business in the marketplace.

In other words, paying attention to documents and their content covers considerable security territory. Most of the leakage of sensitive information is not intentional. Workers and managers do not mean for it to happen. Often, the compromise of data arises from someone working extra hard. They take sensitive files home and before anyone realizes the problem the data becomes compromised. It is lost, stolen, or accidentally placed in the trash.

Thinking to help others, employees place information on the Web. When it is available online, information becomes easy to disseminate and to update. These advantages improve internal communication within an organization, but they also facilitate hacking and information theft against the organization.

The text strives to alert an audience of managers, security professionals, and workers who come in regular contact with sensitive information. Document security is not an accident. At any point in the life cycle of a document if it faces exposure to unauthorized eyes, compromise and loss of confidentiality occurs.

Recognition of how sensitive documents can violate the principle of confidentiality is the primary focus. Continuous protection requires understanding all of the possible avenues for compromise. Those avenues include the following:

A. Not understanding the information conveyed in metadata.
B. Not employing robust encryption protection.
C. Inadequate monitoring of business channels and subsequent filtering to reduce information leakage.
D. Inadequate erasure of magnetic media to reduce remanence.

Chapter 1 discusses metadata in documents. The most common metadata Microsoft Office documents are in the document properties section. The statistical information available there can reveal how long it took to create and to revise the document. In addition, previous revisions of the document may be discoverable. Paying attention to this issue can reduce unintentional release of sensitive information.

In Chapter 2 the text explores Web-facing documents and how search engines like Google® can uncover sensitive data in those documents. This is a widespread problem, and it requires constant attention by security to reduce or eliminate the exposure.

Business channels range from e-mail to instant messaging to FTP transfers. Chapter 3 discusses how filtering these channels is feasible

with modern technology. However, the telephone and events like trade shows and professional meetings also provide business channels that are difficult to filter.

Chapter 4 covers the theft of digital devices such as personal data assistants (PDAs), laptops, and cellular telephones. These devices all contain documentary information. The chapter discusses the use of global tracking technologies and encryption to protect vital information from this growing problem.

Erasing most computer media does not completely remove the information. Special procedures are necessary to completely remove sensitive data. Chapter 5 discusses this issue and explains methods for disposal and reuse procedures.

In Chapter 6, paper and physical documents, such as information written on whiteboards or printed on boxes, pose unique control, disposal, and storage challenges. These documents bring the physical security force into the information security effort, if the organization uses the force properly. Protecting paper and physical documents forms the core of any document security program. Carelessness here is symptomatic overall of a weak information security effort.

Forensics involving computer-based documents looks at digital fragments on hard drives and on other computer media. These fragments tell a story about what a user thought was deleted or written over on the computer. Chapter 7 examines the whole issue of "slack space" on a computer and what security can do to make users aware that computers are the ultimate recording machines.

Chapter 8 continues the discussion by describing anti-forensics. These techniques minimize what forensic examination can uncover. Nothing is foolproof, but awareness goes a long way to preventing inadvertent passing of sensitive data on a data storage device.

Being deceived or fooled by documents is an important issue for security. Chapter 9 deals with the evaluation of online information. Bogus sites can imitate legitimate ones, and other Web sites can pass on disinformation to facilitate phishing and other scams. Learning to evaluate the validity and reliability of online information should be a part of the security training for all employees.

Chapter 10 discusses document forgeries. The increasing sophistication of desktop publishing programs, scanners, and printers means security has to be able to detect forged credentials and vital documents as a part of protecting an organization. Bogus documents necessary for

securing employment continue to proliferate. In addition, academic and business records are also the subject of growing forgery trends.

The basic principles of information security as to documents require understanding the vulnerabilities that information faces. Upon creation, an electronic document may leave unintentional clues as to its content. Even if the main document remains secure, metadata about its contents may be elsewhere on the PC or PDA. Mirrored images may reside in swap files or in backup storage.

In storage, an electronic document may face surreptitious copying or alteration. If not properly classified, confidential electronic documents may encounter unauthorized eyes. Paper documents, due to them being commonplace in work areas, tend to be ignored as a security red flag. Those individuals, however, with a need to know, albeit not a necessarily authorized need to know, will haunt accumulation centers for documents to skim for information. Physical documents written on chalkboards and whiteboards often convey sensitive information in a completely innocuous way. Unless procedures exist to erase this information timely, unwanted eyes may get to study it.

Reusing electronic media has its own special dangers. A disk, a hard drive, a USB drive, or a backup tape that contained confidential data may end up being recycled for nonsensitive use. The remanence of sensitive information compromises the data to unauthorized users. Unless stringent procedures guard against sloppy reuse expect proprietary and confidential data to go walking out the door.

Lastly, destruction of confidential documents requires careful planning and thought, whether those documents are paper-based, physical, or on electronic media. It is a difficult argument to make that someone stole your trade secrets when that person was able to recover them from your unlocked dumpster.

CONTENTS

DOCUMENT SECURITY

Chapter 1

METADATA

The Preface introduced the term, "palimpsest," to describe the texture of electronic documents. Much like an ancient or a medieval parchment, a hidden layer exists below the surface text. With proper forensic techniques this substrate becomes visible. Paintings sometimes have a layer of a previous drawing or painting underneath what our eye perceives. Building on these analogies, we understand that electronic documents often have an unintentional subtext, which, if ignored, may result in the leakage of sensitive information.

In the BBC Web-based article of August 18, 2003, "The Hidden Dangers of Documents," Mark Ward offers several insights into this unique vulnerability. First, documents with numerous revisions, especially if there are multiple collaborators, are prone to information leakage via metadata not being removed after the document's drafting. (Metadata is information about the document itself: the authors, the number of revisions, the time required to produce the document, and so on. But most important, it includes text, tables, and graphics, the authors thought they obscured or deleted.) People do not realize that many word processing systems like MS Word® automatically record this production data and statistics. They fail to recognize that using the command to hide text or illustrations fails to prevent inquiring eyes from discovering the information later. Also, common techniques like whitening text or blackening a graphic or table often fall short in protecting sensitive data. (Numerous business and consumer software products, such as MS Office®, which include Excel® and PowerPoint®, possess this vulnerability.)

Second, the problem is widespread. Mark Ward cites a study by computer researcher, Simon Byers, where Byers gathered 100,000 Word documents from various Web sites. There was not a single document that did

not contain some kind of hidden information. With this shocking evidence, the conclusion that metadata results in the significant leakage of sensitive, confidential, or embarrassing information in both government or business is an information security nightmare that rears its head every day.

Finally, Ward discusses several incidents of metadata telling more than the authors intended. In the United Kingdom, the publicized Iraq "dodgy dossier" unintentionally contained the names of civil servants who worked on it. In America, during the period of the Washington sniper attacks, the *Washington Post* published a letter sent to the police that included confidential names and addresses. Ward notes a case where an employment contract received by an applicant contained previous revisions. The applicant used that sensitive information to negotiate a better deal.

IMPLICATIONS

Why bother about metadata? If all that business and technical writers ever did was print out what they wrote into hard copy and distribute their work product on paper, metadata would not be an issue. Electronic documents allow information to pass rapidly across great distances, and they facilitate twenty-first century commerce. Storing electronic documents uses little physical space compared to paper, and these documents permit searching for the phrase or section heading on the tip of your tongue. In other words, electronic formats for information will continue to stay on the forefront of business and governmental communication.

Awareness of what lies in the sublayers of electronic documents is an important security concern for now and the future. Having an electronic document say more than the author intended is not difficult. Vigilance against these information leaks requires user education, and that education process involves recognizing the main avenues for metadata telling too much.

Begin educating users by explaining that all electronic documents possess properties. Those properties include statistics about the document: editing time, the number of pages, the number of paragraphs, file dates, and how many revisions. While at face value, these numbers appear innocuous. Imagine, however, if a writer bills eight hours to a client for a document where the metadata indicates total editing time was only two hours. True, the writer took into account time to research and to plan the project, but the metadata raises doubts in the client's mind. Knowing the number of revisions may give clues about the difficulty and

complexity in the document's composition. Again, claims of an arduous drafting process may be questioned if the statistics suggest a less difficult composition effort.

Other general properties provide the names of authors, collaborators, and author's comments about the document. Custom properties include the document number, the group creating the document, the language used, the editor, and other facts about the text or file. Routing slips containing email addresses of reviewers or collaborators, when using the "File Send" function, also act as another "metadata trap." Document authors often forget that these internal properties exist as metadata behind or below the visible, overt data. While at face value, little harm results in most cases if a third party sees this data, yet in certain documents, one may not want outsiders to know all the collaborators on a project, or who reviewed the document prior to publication.

Most problems resulting from metadata information leakage arise when the user or author tries to hide portions of the document. Hiding equates to security in many writers' minds. But, security through obscurity often fails in practice. Common methods for hiding are:

1. Suppressing portions of text.
2. Hiding comments appended to a text, a spreadsheet, or a slide in a presentation.
3. Suppressing headers and footers or footnotes.
4. Whitening text on a white background.
5. Making text very small (usually on Web pages).
6. Hiding slides in a presentation.
7. Suppressing cells, data rows, and columns in a spreadsheet.
8. Suppressing embedded objects such as graphics or photographs in a document.
9. Suppressing hyperlinks or using text as an alias for the URL.
10. Redacting sensitive portions of a document by blackening or otherwise obscuring the area.

A majority of the items on the list (except for items 4, 5, and 10) occur during the drafting of the document and quickly get forgotten as being a hidden part of the final draft. If an author uses the "Track Changes" feature during the writing process, the history of changes to the document remain as a sublayer in the final draft. Many desktop suites like MS Office make the suppression of portions or a section in a document just a matter of a few keystrokes. Turning on the "Reveal Formatting"