Design of an Automatic Password Protection Mechanism for Digital Documents

Shashank Kaushik and Thomas Way
Applied Computing Technology Laboratory
Department of Computing Sciences
Villanova University, Villanova, PA 19085

Abstract - *Maintaining digital security and privacy are critical issues in the modern workplace. Sensitive and proprietary data is frequently transmitted electronically, and with the large volume of such transmissions, it is inevitable that some material will be inadvertently sent to unintended recipients, leading potentially to the unintended revelation of private data. The paper reports on the design of an automatic mechanism for password protection of digital documents on a per-user basis. Recommended practices and details of architecture, approach and implementation are presented for the specific protection of Microsoft Excel documents, although the scheme can be generalized to other digital document management software.*

Keywords: Digital document security, information system security management, password protection.

1 Introduction

The corporate office setting is highly digital, and it is not uncommon to transmit tens or hundreds of sensitive documents each day within an organization, including its clients and customers [6,7]. With this large quantity of data sharing comes the potential for inadvertently sharing sensitive documents with incorrect recipients [1]. Although most unintended recipients will simply notify the sender of the error and delete the data that is not intended for them, others may choose to view the confidential material or even share it with others. The risk exists that sensitive documents could ultimately be revealed to competitors. Thus, protecting against this inevitable misdirection of material is a critical need in the modern workplace [1,6,7].

Password protection is a natural solution to this problem. With well-chosen and consistently applied passwords, digital documents can be encrypted so that unintended recipients cannot view, and possibly share, this sensitive data [17]. Ideally, each document will be individually password protected using a password known only to the sender and each recipient. In practice, this is a tedious and time-consuming approach, making it less likely to be used consistently. More frequently, a general password is used to share sensitive documents with a trusted group. Although this approach is better than no protection at all, widely shared and used passwords are likely to be known eventually by non-members of the trusted group [6].

Generating new passwords on an individual or group basis means that the passwords must be communicated to the recipients, and thus increases the chance that passwords may wind up in unintended hands in the same way that documents do [17]. Assuming that a secure mechanism exists for creating and transmitting individual passwords for recipients, the issue becomes one of how to manage and use a large list of recipients and their passwords. Such a list would be referred to each time a sensitive document is password protected and sent to a specific user on the list. The consistent use of this form of large database of individuals and passwords would be labor intensive and could therefore discourage consistent use.

There are a variety of approaches to digital data security in use [1,3,11,12,15]. One practical solution that can overcome the tyranny of password list management is to automate the use of that information. If documents can be easily password protected for specific users with only minimal effort, mistakes can be reduced and the likelihood that the password protection will be used can be increased.

In this paper, we present the design of a centralized, file-based, password protection system for securing digital documents. The design has the goal of allowing only an intended recipient of a document to view its contents in a digital office environment where many

sensitive documents are shared among many recipients. This system makes use of a master file of recipients and their corresponding passwords, and automatically performs password protection. Details of an implementation of this design that protects a commonly used office document format, Microsoft Excel Workbooks, are provided and the results of a performance evaluation are discussed. Best practices for password generation and use also are reviewed, and the advantages and disadvantages of this approach with suggestions for extending this work are provided.

2 Digital Document Security

Although the issue of digital document security has been well known and studied for a number of years [5], the recent tremendous explosion of connectivity and the corresponding rise in exchange of digital documents has increased the stakes. The security of these documents increasingly is an important issue, with a wide variety of approaches to digital document security in use [12]. Although the many approaches available and in use for a range of applications, such as digital watermarking or secure, end-to-end data streams, are valuable tools in the right circumstances [11,15], the use of password-based protection is nearly universal [3].

Password protection is widely used as a means to protect digital documents and guard sensitive data, ensuring that they are viewed only by those with the appropriate permission to do so by means of a password, whether the protection is used for individual documents or on a larger scale such as the Internet [3].

Password-based schemes can be highly secure as evidenced by their widespread use [1,2,10]. A password is a form of secret authentication data that is used to control access to a resource [2]. The password is kept secret from those not allowed access, and those wishing to gain access are tested on whether or not they know the password and are granted or denied access accordingly. The security of any system, including the one we describe in this paper, is greatly improved by making use of good practices for password selection and consistent use.

2.1 Good Password Practices

Password policies are generally known and often include advice on proper password management [2,4,10,14,16], with the most common including:

- Never share login information for a personal computer account, as such sharing can lead to reduced security through inadvertent or intentional additional sharing of login information.
- Never use the same password for more than one account, to reduce the risk of any one password being compromised.
- Never reveal a password to anyone, including people who claim to be from customer service or security, to prevent social engineering attempts to steal a password.
- Never write down a password, since written passwords may be inadvertently revealed to others.
- Never communicate a password via email or instant messaging, and only rarely via telephone, as these communication media can be compromised.
- Always log off before leaving a computer unattended, which prevents unintended access to sensitive data on one's computer which could lead to compromise of one's password.
- Change passwords whenever there is suspicion they may have been compromised, reducing the risk of exposure of confidential data.

The general goals of password management policies are to maintain the integrity, privacy and efficacy of a password-based security system. Among the most critical is the careful selection of a password.

2.2 Choosing an Appropriate Password

By following guidelines for selection of a password, the risk of a password being compromised through guessing or cracking is greatly reduced [4,10]. The level of password strength required depends, in part, on how easy it is for an attacker to submit multiple guesses. While some systems limit the number of times that a user can enter an incorrect password before a delay is imposed or the account is frozen, other systems allow virtually unlimited login attempts. An attacker can try passwords by guessing commonly used ones based on a user's name and other personal information. Common guidelines for appropriate password selection are:

- Make passwords as long as possible, as longer passwords are harder to guess.
- Use as many different characters, numbers and symbols as possible to make it harder to guess.
- Do not use personal information, as it is easily guessed.

- Do not use common dictionary words, as they are easier to guess than random characters.
- Change passwords on a regular basis to minimize the risk of compromised passwords.

While following such guidelines can reduce the chance that a password will be guessed and therefore compromised, a stronger justification for following these practices is to reduce the change that password cracking software will compromise a password.

2.3 Password Cracking

Some systems make use of a specially hashed or encrypted version of each user's password so that the system can verify that password upon login [3]. If this file of encrypted passwords is obtained, an attempt can be made to discover the original passwords using password cracking software.

Password cracking is the process of recovering secret passwords from data that has been stored in or transmitted by a computer system [13]. The approach makes use of lists of common passwords, dictionary words and typical algorithms that people use to create passwords to repeatedly try to guess a password. Although there are legitimate uses for password cracking, such as the recovery of a forgotten password, these are typically unnecessary. A system administrator can easily reset a user's forgotten password. By carefully selecting a hard to guess password, a user can reduce or eliminate the risk that an automated password cracker can expose the password.

3 System Design

Maintaining, managing and making consistent use of a large list or database of recipients and passwords can be a significant challenge [1,2,17]. Certainly, the use of well selected passwords is critical, but there is more to document security than simply choosing good passwords. As with most tasks in an office environment, if the use of a password scheme is cumbersome or overly time-consuming, it is less likely to be used with any regularity [7]. The use of an automatic mechanism for managing the use of a large database of passwords is a natural solution. In this section, we present the design of a systematic approach to the automatic password protection of sensitive documents in a busy digital office environment.

3.1 Architecture

The proposed design of an automatic password protection system is shown in Figure 1. The system makes use of a central database of recipients and corresponding passwords, together with a programmatic mechanism that uses the list to apply password protection the desired document. The major actors and components of the system design are:

Creator – the author of the document for which password protection is desired.

Password Generator – an administrator responsible for generating and maintaining the database of recipients and passwords.

Recipient – an authorized or intended recipient of a password protected document received via email or other means from the **Creator**.

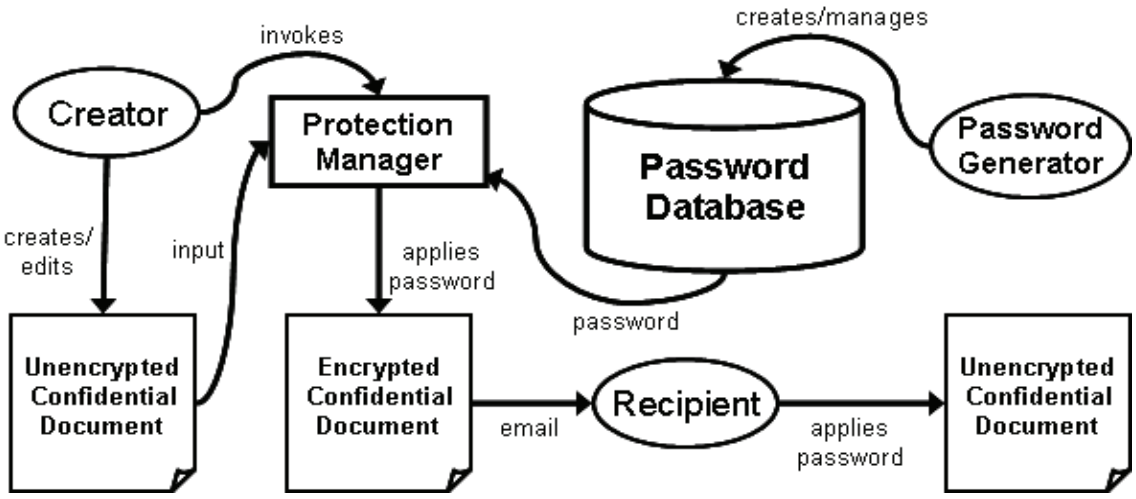

Figure 1. Design of general automatic digital document password protection system.

Protection Manager – software module that performs encryption of a document using a recipient name and password, invoked by the **Creator**.

Password Database – database, file or other collection of recipient names and corresponding passwords, managed by the **Password Generator** and used by the **Protection Manager** to automatically encrypt documents.

Unencrypted Confidential Document – the viewable document for which password protection is desired.

Encrypted Confidential Document – the encrypted version of the **Unencrypted Confidential Document**, viewable only by a recipient who applies the correct password.

The process by which a digital document is prepared, protected and later viewed proceeds as follows. First, a **Creator** authors or edits an **Unencrypted Confident Document** with the intent to distribute it securely to one or more recipients. The **Creator** then invokes the **Protection Manager** software, which retrieves from the **Password Database** a password for each intended recipient. The **Protection Manager** automatically encrypts the original document using the password for each intended recipient, resulting in an **Encrypted Confidential Document** for each of the recipients. The encrypted document is transmitted to the **Recipient**, who uses their own password, which is the same as is stored in the **Password Database**, to decrypt and view the original document.

The **Password Generator** is an administrator responsible for maintaining the database of recipients and passwords, including guiding the use of good password practices.

This system design has been implemented in prototype form for use in automatic password protection of Microsoft Excel Workbook documents. In addition to conforming to the proposed design, the prototype was designed to be easy to use and maintain, increasing the likelihood that it will be consistently used to maintain the confidentiality of sensitive data.

Figure 2 illustrates the specific components of the implemented system. The **PasswordManager.xls** file is the set of macros, in this case in an Excel document that makes use of a recipient password to perform encryption of a document. These macros reside on each user's computer, or alternately on a central server, and contain no sensitive information. The **PasswordTable.xls** is a straightforward table of recipient names and passwords.

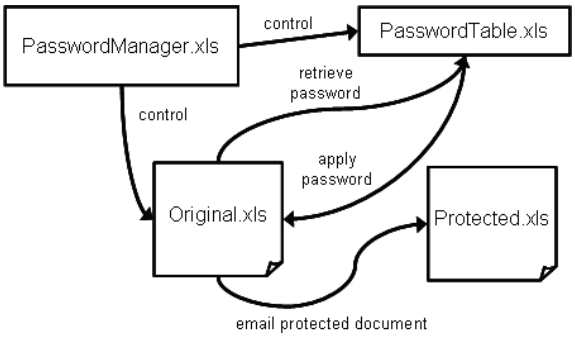

Figure 2. Diagram of prototype automatic password protection system for Microsoft Excel documents.

In the prototype implementation, passwords are stored in plaintext in the **PasswordTable.xls** file, so security of this file is critical. In a production version, passwords would be stored in a reversibly encrypted form, perhaps by using a public key system so that passwords would not be transmitted, even across an internal network, in plaintext form. The **Original.xls** and **Protected.xls** files are the original, unencrypted document and its encrypted counterpart, respectively.

When a workbook is selected for encryption by a user, the **PasswordManager.xls** macros will attempt to retrieve the password for the intended recipient from the **PasswordTable.xls** database file. The name of the recipient is either contained as the first item in the **Original.xls** file, or alternately entered in a simple, pop-up dialog. If a password is found, the password protection provided in Microsoft Excel is applied using the recipient's password. If no password is found for a recipient, the user is notified that password protection has failed and the file is not encrypted.

Once received by a recipient, the Protected.xls file is opened only if the correct password is known and entered. In the case where an unintended recipient receives the Protected.xls file, so long as good password selection practices were followed, the confidentiality of the document will be preserved.

3.2 Description of Files

Care must be taken is selecting locations for the files used to manage this password protection system. If these files are not kept in a consistent and available location, ease of use will be affected, and users will be less likely to make use of the system. Fortunately, all common versions of the Microsoft operating system provide standard user directories, making use of installation package software practical. The locations for password protection system files are:

PasswordManager.xls

C:\Documents and Settings\<user>\Application Data\Microsoft\Excel\XLSTART

This path is equally valid on single or multi-user computer systems. By placing the **PasswordManager.xls** in the folder "Excel\XLSTART," it will be able to control all new and existing Excel files that are opened [8,9]. The file should be set with a read-only attribute, primarily to protect it from inadvertent deletion.

PasswordTable.xls

C:\Documents and Settings\<user>\Application Data\Microsoft\Excel

Again, this path holds true in the case of systems which are shared by multiple users. Although this file could exist anywhere on a system, keeping it on a user's machine provides additional security against unintended password revelation. It could also be kept on a central server, although additional considerations of security and access should be addressed.

The specific form of the **PasswordTable.xls** file is a single Excel workbook of two columns. Column A has recipient names, which could be the names or email addresses of individual or company recipients. Column B contains the corresponding password for each recipient listed in column A. The constraint imposed by the **PasswordManager.xls** on passwords is that they must be at least six characters in length, to reduce the occurrence of insecure or easily guessed passwords. Passwords that are shorter result in an encryption failure notification to the user, with the result that the document will not be password protected. Additionally, the **PasswordTable.xls** cannot contain empty rows in between filled rows in the table.

Possible enhancements to the **PasswordTable.xls** and **PasswordManager.xls** include the use of individualized user encryption and description of the **PasswordTable.xls** by the **PasswordManager.xls** macros to increase security, and additional password analysis to assure the best possible practices are used in the selection of passwords.

3.3 Deployment and Use

The system is deployed by copying the files above into the designated directories on each user's computer. If a different location is used for the **PasswordTable.xls** file, the corresponding path for that file listed in the **PasswordManager.xls** file should be updated manually. Upon installation of the **PasswordManager.xls** file, the administrator should set the file's "hidden" attribute to reduce the chance that a user will inadvertently delete or modify the macros. Once the files are installed, the system is ready to automatically encrypt documents.

When a document **Creator** tries to save a new or modified workbook, the **PasswordManager.xls** triggers a **WorkbookBeforeSave** event, which starts the password protection process. The **PasswordManager.xls** refers to the **PasswordTable.xls** database, looking for a match for the desired recipient. If a recipient is matched, the corresponding password is retrieved and applied to the workbook being protected. When a recipient is not matched, the document simply is saved normally with no encryption.

4 Evaluation

The prototype system was implemented as described and tested on a typical office workstation consisting of a 1.8GHz processor with 1 GB of memory and 120GB disk. The system was evaluated by timing the duration required to retrieve Excel workbooks using password protection with varying sizes of password tables and document lengths. This was done to measure scalability of the approach, and to identify areas of future improvement.

Documents of varying sizes were selected, ranging from a very small 13.5 KB document to very large 42.8 MB document. Preliminary experiments indicated that the performance of this password protection scheme was independent of the document size.

4.1 Results

The **PasswordManager.xls** was tested for a range of password table row matches, which indicates the row in which the matching recipient was found in the **PasswordTable.xls** file. The password file contained 65,536 rows for all experiments. The results of these experiments are shown in Table 1.

Table 1. Time needed to search and find a matching password for a specified recipient.

Matched row in password database	Elapsed time (secs.)
2	1.53
475	1.55
2570	1.56
3741	1.58
65536	2.22
not found	2.22

The critical measurement for our application was the scalability of the approach based on the elapsed time needed to search the **PasswordTable.xls** to locate a

password to be used. The results indicate that the approach scales very well. It is relatively unlikely that a given user will require a database of recipients and passwords as large as 65,536, although even at that size the performance is very close to that of a much smaller database. These results demonstrate that this approach is very scalable. Of course, the time required to encrypt a document once a password is retrieved will depend on the original document size, but that is a factor regardless of the password selection mechanism used.

In general, a user may expect an approximate delay of 2 seconds for the application of password protection using this system. It is likely, given the tendency toward creation of ever larger documents, that this small delay will be unnoticed relative to the time needed to perform the actual password protection. These experiments and the overall design and use of such a system suggest a number of advantages and disadvantages to using such a system.

4.2 Advantages

Transparency - The creator of the original Excel file does not have to worry if the file across to the wrong recipient, because password protection is applied automatically. The use of this transparent mechanism greatly increases the chances that sensitive data will remain secure, even if it ends up in the hands of an unintended recipient.

Maintainability - The **PasswordTable.xls** can be placed on a shared drive and have an administrator maintain it, so the typical user is not concerned with managing the password database.

Uncrackability - The current scheme ensures that passwords will be difficult to crack, so long as good password selection practices are used. The minimum length of the password in this system is currently configured to be six, although it can easily be changed if stronger security is needed by manually editing the **PasswordManager.xls** file. This flexibility in length ensures that passwords will be difficult to guess and crack, improving document security.

4.3 Disadvantages

Missing Passwords - The creator of the Excel file has to take care of certain conditions, such as missing passwords for intended recipients and poor password selection, without which the Excel workbook would behave as any general workbook. The result would be a document that is not be password protected. However, with brief training, this disadvantage can be overcome.

Scalability of Applications - As the number of companies and clients increases, the actual application of encryption once a password is retrieved by our system can become slower. This problem with scalability is recognized, so any commercial application of this approach will require additional investigation to improve scalability.

5 Conclusions & Future Work

This report describes the design and evaluation of a scalable password protection scheme for use in maintaining confidentiality of documents in the digital workplace. The system makes use of a centralized database of passwords, applying then on-demand to a sensitive file. Assuming good password selection practices, the prototype system that was implemented is easy to install and use, and scales very well to any expected database size. The issue of scalability of encryption for large documents remains but is beyond control of our password retrieval and application mechanism.

Considering the identified advantages and recognized disadvantages of this approach, an overall system that makes use of a password database to provide and apply password-based encryption to documents is feasible and well-advised. Although this research only analyzes an implementation for protection of Microsoft Excel workbooks, the general approach is more broadly applicable. The macros defined in the PasswordManager.xls file can be easily ported, with little or no modification, to other Microsoft Office products. Adapting the approach to other platform will require more effort, but the basic algorithms used are uncomplicated and easily understood.

There are a number of ways that this work can be extended to provide additional levels of security, or to transfer this mechanism to other document formats, such as:

Password guessing remains a problem with the current system. If a limit is places on the number of allowed attempts at entering the correct password, or a time delay is introduced between password attempts, password guessing can be made much less attractive to an unauthorized recipient.

Automatic renewal of passwords for an end client, such as when a password is forgotten or lost would alleviate the most common password database maintenance task. By allowing users to securely recover or reset a password, management overhead is reduced. Common approaches to this could be used, such as verifying the user's email address and including a secret question-answer key.

Time-based expiration of the Excel file could be incorporated, so that files can only be opened in a given

time period. By using a time-based expiration scheme, files that are likely to be needed for a short period of time can be given additional protection. This mechanism becomes particularly important when machines are used by multiple users, where some common exposure of the file system on that machine is unavoidable.

6 References

[1] M. Attaran and I. VanLaar, "Privacy and security on the Internet: how to secure your personal information and company data," Information Management & Computer Security, 7(5), 1999, pp. 241-247.

[2] M. Bishop, "Password management," Compcon Spring '91 Digest of Papers, 1991, pp. 167-169.

[3] W. R. Cheswick, S. M. Bellovin, and A. D. Rubin, "Firewalls and Internet Security: Repelling the Wily Hacker," Addison-Wesley, 2003.

[4] S. Furnell, "An assessment of website password practices," Computers & Security, 26(7-8), December 2007, pp. 445-451.

[5] M. Gasser, A. Goldstein, C. Kaufman, and B. Lampson, "The Digital Distributed System Security Architecture," Proc. 12th National Computer Security Conf., NIST/NCSC, Baltimore, 1989, pp 305-319.

[6] S. K. Katsikas, J. López, and , G. Pernul, "Trust and Privacy in Digital Business," First International Conference on Trust & Privacy in Digital Business, Zaragoza, Spain, 2004.

[7] P. J. McFadden, "Guarding Computer Data," *Journal of Accountancy*, 184, 1997.

[8] Microsoft online article, "About the XLSTART folder in Microsoft Excel," Accessed 9/13/07, Available: http://www.officearticles.com/excel/about_the_xlstart_folder_in_microsoft_excel.htm.

[9] Microsoft online article, "How to use Excel startup folders in Excel," Accessed Available at: http://support.microsoft.com/kb/291218, extracted on 09/23/2007

[10] Microsoft whitepaper, "Strong passwords: How to create and use them," 2006, Accessed Jan. 31, 2008, Available at: http://www.microsoft.com/protect/yourself/password/create.mspx.

[11] P. Papadimitratos and Z. J. Haas, "Secure data transmission in mobile ad hoc networks," Proceedings of the 2nd ACM Workshop on Wireless Security, San Diego, California, 2003, pp. 41-50.

[12] J. Park, R. Sandhu, and J. Schifalacqua, "Security architectures for controlled digital information dissemination," Computer Security Applications (ACSAC '00) 16th Annual Conference, 2000, pp. 224-233.

[13] Password Cracking Wikipedia article, Accessed Jan. 31, 2008, Available at: http://en.wikipedia.org/wiki/Password_cracking.

[14] Password Policy Wikipedia article, Accessed Jan. 31, 2008, Available at: http://en.wikipedia.org/wiki/Password_policy.

[15] C. I. Podilchuk and E. J. Delp, "Digital watermarking: algorithms and applications," Signal Processing Magazine, IEEE, 18(4), 2001, pp. 33-46.

[16] J. Yan, "A note on proactive password checking," New Security Paradigms Workshop, Proceedings of the 2001 workshop on New security paradigms, Cloudcroft, New Mexico, 2001, 127-135.

[17] J. Yan, A. Blackwell, R. Anderson, and A. Grant, "Password memorability and security: empirical results," Security & Privacy, 2(5), 2004, pp. 25-31.