Applied Information Technology And Computer Science Vol. 2 No. 2 (2021) 155-164
© Universiti Tun Hussein Onn Malaysia Publisher's Office

AITCS

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/aitcs
e-ISSN :2773-5141

DocWIPE: Data Wiping Tool Using Randomized 512-Gram Technique

Nurul Atifah Mohd Borham, Kamaruddin Malik Mohamad*

Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400, MALAYSIA

DOI: https://doi.org/10.30880/aitcs.2021.02.02.011
Received 15 June 2021; Accepted 09 September 2021; Available online 30 November 2021

Abstract: Data wiping is a method that being used to delete data permanently from storage devices while overwriting is a process of write over existing data in hard drive to make it unreadable. Due to data breach, data privacy become a concern and the existent of data recovery software make the situation worst. Thus, this study proposed a data wiping tool named DocWIPE in which replacing the content of file using a random number algorithm. The tool called DocWIPE was developed to wipe Microsoft Words document of .docx format. The implemented algorithm replaced the original files content with sequence of random number and then replace the current overwritten content with selected character using Random Walk. Once rewriting phase is completed, the implemented algorithm repeats this rewriting phase again for every 512 bytes of data to make the file unrecoverable. This project creates an alternative way for file sanitization. Object-Oriented Software Development (OOSD) is used as the methodology to develop this tool and written using C# programming language. With this application, user can add, delete, display and wipe files. This tool is expected to securely dispose files while protect the privacy of the file content.

Keywords: Data Wiping, Overwriting, N-Gram, Random Walk

1. Introduction

In the past few years, data breach is one of the problems faced for data security [1]. As more people depends on digital devices to communicate or to their tasks which results in increasing amount of data stored, so data privacy become a concern. This issue mostly happened because users typically assumed that when simply delete the data, that data cannot be recovered. Simply deleting data by normal deletion like delete data in Recycle Bin is not reliable anymore because the data still available in the device storage [2]. The existent of data recovery software makes the situation worst as this application can recover the data from device storage in which could lead to data exposure.

Due to data breach issue, people need to aware about the important of data privacy. From this viewpoint, it is suggested that some precaution need to be taken which is to perform a proper deletion of data. Data wiping, on the other hands, is a method to delete data from storage devices permanently so that the data is unrecoverable [3]. One of the data wiping techniques is overwriting in which involve

**Corresponding author: malik@uthm.edu.my* write over existing data. This process may involve overwrite with several time of passes. The more it overwrites the data, the harder it is to recover adequate data to reassemble a deleted file.

This study is therefore proposed DocWIPE as one of the alternatives for user to permanently delete the files. This application focuses on secure file deletion instead of wiping the whole hard drive. The aim of this project is to develop a data wiping tool that wipe on .docx file with randomized 512-gram technique. In order to achieve the aim, three objectives have been set which are to design data wiping tool using randomized 512-gram overwrite technique, to implement DocWIPE and lastly to test the functionality of the DocWIPE.

The rest of this paper is organized as follows. Section 2 explains detailed about literature review on the related work. Section 3 discusses further about the methodology used throughout this paper. The analysis and design of this application will also be described in this section. Other than that, Section 4 presents the implementation and testing of the application. Lastly, conclusion and future works are presented in Section 5.

2. Background of Study

This section discusses the related terms to the application such as data wiping, examples of overwriting method being used, random walk algorithm and comparative study of the existing applications with the proposed application.

2.1 Data Wiping

Data wiping is a method that being used to delete data securely from storage devices and make it impossible to be recovered. Unlike degaussing or physical destruction which cause the storage media completely damage, overwriting leaves the hard drive operate and reusable [4].

A work by Wei *et al* [5] proposed that overwriting is a process of replacing previous data in computer storage with new data. Overwriting used algorithm to remove any part of existing data by writing over with new data. Previous works of [6], [7], [8] and [9] reported the implementation of different overwriting methods. There are many existing methods that have been used in overwriting process.

Firstly, Gutmann is a method that introduced by Peter Gutmann. Bennison *et al* [9] proposed Gutmann method to securely erase the data in hard drive. This method using 35 passes overwrite of random data and specific bit patterns. The general concept behind of overwriting scheme is to flip each magnetic domain on the disk without writing the same pattern twice in a row [4].

Next is Schneier's algorithm which developed by Bruce Schneier. He suggested the process of wiping information consists of seven passes to permanently delete data on a storage device. This process involves write with one, zero and lastly with a pattern of random character [10].

Other than that, DoD 5220.22 M is a method developed by US National Industrial Security Program (NISP). In this method, overwriting process involve three passes and seven passes. The process involves write zero and verify; write one and verify; write a random character and verify [11].

2.2 Random Walk

Random walk is a mathematical object, known as a random process in which describes a path that consists of a sequence of random steps [12]. The term of random walk was first introduced by Karl Pearson which described about drunkard's walk [12]. In the other word, random walk is used to describe the situation of an object that moves in random motion.

In this method, random number algorithm is implemented which is Pseudorandom Number Generator (PRNG). The PRNG is an algorithm that generates a sequence of random number. Actually, this sequence number is not completely random because the initial value in the algorithm has been determined [13].

Pseudorandom numbers are chosen with the equal probability from a finite set of numbers. The chosen numbers are not completely random as the mathematical algorithm is used to select the numbers but also meet the certain requirement of randomness [14]. For instance, given a pseudorandom number sequence initial value of three. It is means that the random number sequence consists of number zero, one and two. Then from that, the algorithm selects numbers within this range to generate sequence of random number.

Random walk is widely used for stochastic function in which outlined many outcomes due to its randomness. Nowadays, pseudorandom number algorithm are used in variety of areas like game design, simulation and modeling [13]. Random walk serves an essential of stochastic activity. A stochastic or random process has a collection of random variables in which each variable is associated with the element in the set of numbers. Hence, random walk is used in this project to presents its randomness in the file content. So that it can meet the requirement of randomized overwriting process.

2.3 Comparative Study of Existing Tools

This subsection discusses and compare the existing data wiping tools. There are several tools that function in the same way as the proposed application. There are three existing tools like CBL Data Shredder, Eraser and also Darik's Boot and Nuke (DBAN) are chosen as comparison with the proposed tool. In addition, the study on data recovery software has also been conducted.

Firstly, CBL Data Shredder is a data wiping tool that specifically wiping on entire hard drive [15]. This tool offers bootable media in which can be run as a floppy disk program or ISO disc image. This tool supports six different overwriting methods. It also allows user to determine custom deletion method for the hard drive. However, users need to aware with their action when using this tool as this tool perform wiping process without warnings prompt. Other than that, some text on the software is written in German and cannot be changed.

Besides that, Eraser also has been analyzed. Eraser is an open source software that offer securely delete files, folders, and entire hard drive [16]. It allows user to set up wiping task schedules which in form of daily, weekly, and monthly. This tool supports 13 different overwriting methods and the default method in this tool is Gutmann method. The tool seems easy to use and well-designed. However, this tool lacks feature such as the number of data sanitization methods. The tool only can be run within Windows operating system.

Darik's Boot and Nuke (DBAN) is another an open source software that specifically wiping on entire hard drive [17]. This tool provide feature of burn the program on a disc or create bootable removal media devices. So, user need to burn the ISO image of this software to a USB, DVD or CD with a DVD burning software and then run it from external hard drive when restarting the operating systems. This tool supports six different overwriting methods that user can choose. However, this tool cannot erase a certain partition on the disk as well as lacks the ability to erase certain files or folders. Other than that, this tool has an unapproachable user interface and it might look unfriendly for non-technical users.

DocWIPE is a proposed application in this project. It is an application that offer securely delete individual files. This tool specifically wiping on files in .docx format. The tool is easy to be use and well-designed, so it is approachable even for non-technical users. Most of the tools stated have been used the usual algorithm, which is either write zeros, ones as well as random characters. None of the tools have a function where the file content being rewritten with random walk method.

Hence, the DocWIPE application would be an application that implements an algorithm to generate sequence of random number for overwriting and also add on a new method which is replacing the file content with selected characters. After rewriting phase is completed, the implemented algorithm repeats this rewriting phase again for every 512 bytes of data to make the file unrecoverable. The user can use this tool without worries as the tool perform wiping and deleting process with warning prompt. This tool can be run on Windows only.

Other than that, Recuva also has been analyzed. Recuva is a file recovery software that can recover files that have permanently deleted and also marked as free space in hard drive by the operating system [18]. The tool supports any kinds of file such as pictures, music, documents, and emails. For this project, Recuva has been used to test the functionality of wiping files. It is supposed that the overwritten file cannot be recovered. Hence, this is to ensure that DocWIPE does not allow file recovery.

Finding of the existing data wiping tools and comparison with the proposed DocWIPE is summarized as in Table 1 below.

Table 1: Comparative study of existing data wiping tools

	CBL Data Shredder	Eraser	DBAN	Proposed tool - DocWIPE
Platform	Unix/Linux And Windows	Windows Only	Unix/Linux And Windows	Windows Only
Hard Disk Sanitization	Yes	No	Yes	No
File Selection	No	Yes	No	Yes
File Sanitization	No	Yes	No	Yes
Algorithm Used	Peter Gutmann's Algorithm, Bruce Schneier's Algorithm, and German VSITR Standard	Bruce Schneier's Algorithm, Canadian RCMP TSSIT OPS-II, German VSITR Standard	US DoD Standard 5220.22-M, Peter Gutmann's Algorithm, Canadian RCMP TSSIT OPS-II	Random walk using Pseudorandom Number Generator (PRNG)

Based on Table 1, the similarity between CBL Data Shredder, Eraser and DBAN are the tools provide multiple of data sanitization methods as there are many algorithms have been used. These tools stated have been used the usual algorithm, which is either write zeros, ones as well as random characters. DocWIPE application implements Pseudorandom Number Generator (PRNG) algorithm to generate sequence of random number for overwriting and add on a new method which is replacing the file content with selected characters by using Random Walk method.

Besides that, both CBL Data Shredder and DBAN specifically support wiping on entire hard drive. Meanwhile, Eraser and DocWIPE offer feature of wiping files as well as support file selection mode. Other than that, CBL Data Shredder and DBAN can be run on Unix/Linux and Windows operating systems while both Eraser and DocWIPE can be run only on Windows operating system.

3. System Development: Methodology and System Analysis

Object-Oriented Software Development (OOSD) model has been chosen as the methodology to carry out this project. OOSD is a practical method for analyzing and designing an application that apply the object-oriented concepts [19]. It also develops visual models throughout the application development. There are five phases in OOSD as shown in Figure 1.

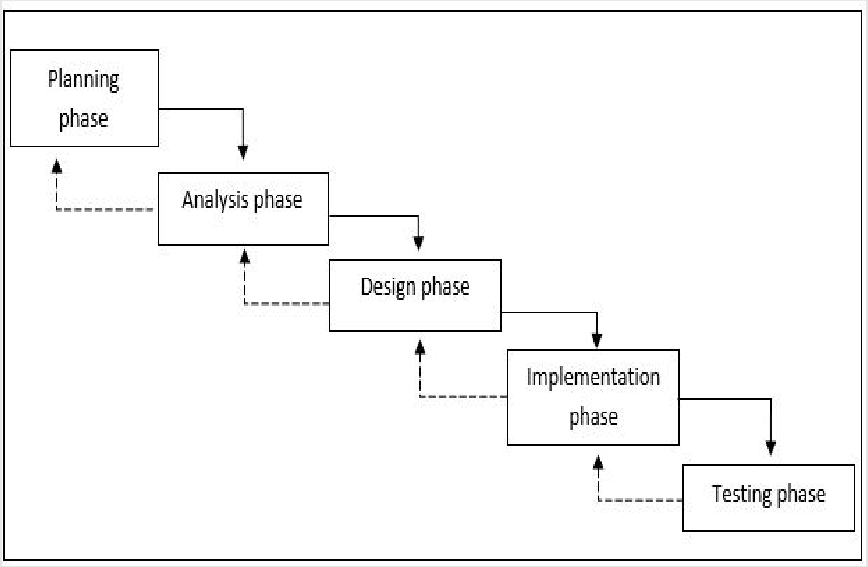


Figure 1: Phases of OOSD

Figure 1 shows all main phases that included in the Object-Oriented Software Development (OOSD). In the beginning of the development, the application is observed, analysed and the requirements are defined. After that, the objects in the application are identified.

Planning phase is the earliest stage in this methodology. In this phase, project objectives; problem statement; and expected result are proposed. The project activities and schedule also are determined. Other than that, some researches have been conducted which relates to the project. A comparative study of existing data wiping tools has been carried out and then the finding is compared with the proposed tool which is DocWIPE.

The second phase is analysis phase. From the finding of the existing data wiping tools, the features of DocWIPE are determined. DocWIPE is a data wiping tool that implement overwriting process using randomized 512-gram technique. This is where user requirements have been analyzed. Other than that, the tool requirements are also defined which are consists of functional requirements and non-functional requirements. The software and hardware that needed in the application development have also been identified.

During the design phase, Graphical User Interface (GUI) are designed which consists of login, register, main menu, file selection, overwriting and file deletion. A database design is created using Microsoft Access in which important to store user's information during login and registration. This application is developed using C# programming language. The interaction between user and the application are presented and modelized using Unified Modelling Language (UML).

For implementation phase, the object's class and the interrelationships are perceived and then written in the C# programming language. The database is established, and the application is given functional design. Every module contained in the application involves the use of specific programming code to enable the module to interact with each other.

Lastly, testing phase is conducted to examine the functionality of DocWIPE. This is to ensure that all module function well as expected and also meet the requirements. The testing is carried out to identify whether designed modules in DocWIPE can be executed properly without any bug or error. The application tested based on their functionality and design of the interface.

3.1 Unified Modeling Language (UML)

Unified Modelling Language (UML) provides a standard way of visualizing the design of a system. The UML is presented in use case diagram, sequence diagram and activity diagram.

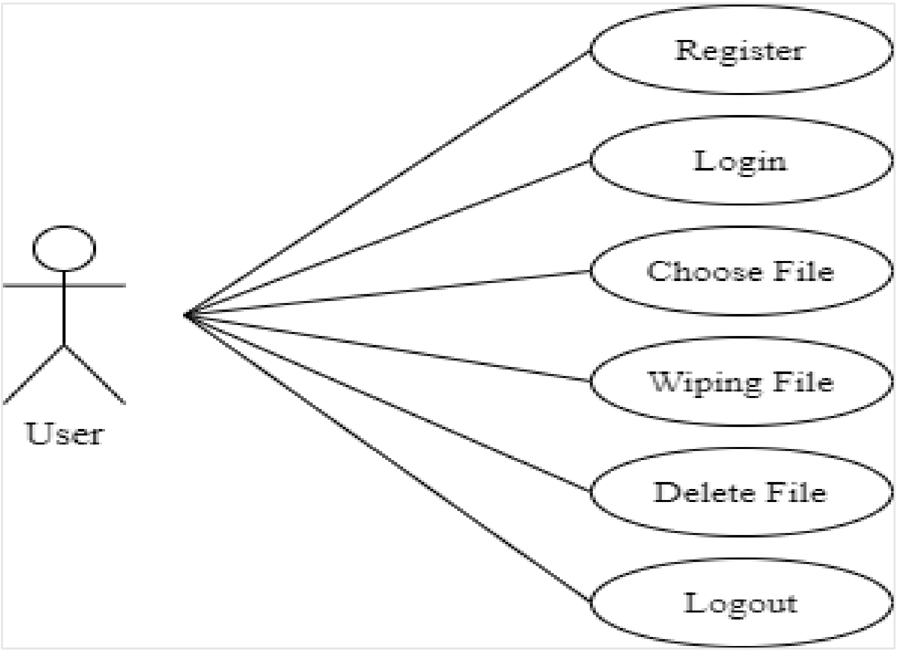


Figure 2: Use case diagram for DocWIPE

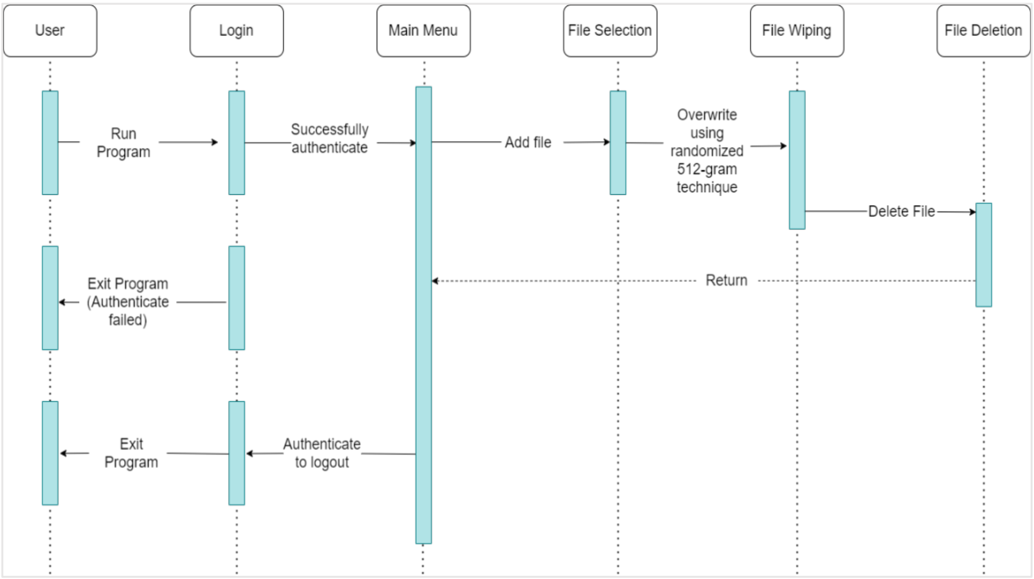


Figure 3: Sequence diagram for DocWIPE

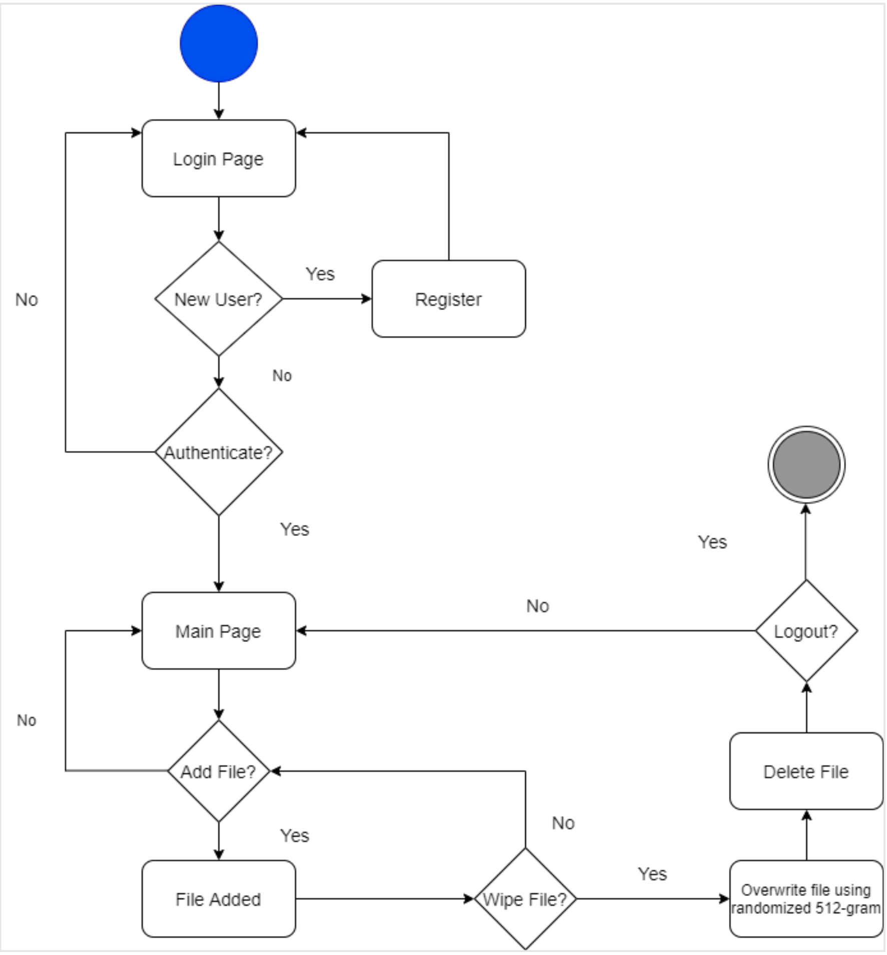


Figure 4: Activity diagram for DocWIPE

Figure 2 shows the interaction between user and the application. Figure 3 describes the associated sequence of events with user action while Figure 4 presents the flowchart of the application. Based on these diagrams, there are only an actor and six use case to show the requirements of the application. The user able to perform tasks like register and login to the application, choose file, wipe file, delete file and logout from the application.

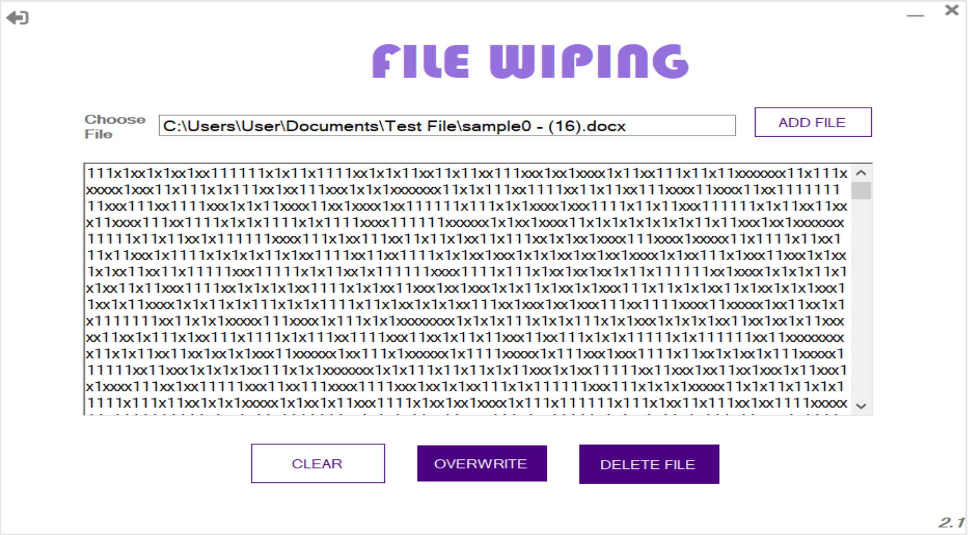
4. Result and Discussion

This section presents the implementation and testing that was carried out for DocWIPE application. The implementation phase involves programming code and graphical user interface development for the application. Other than that, the testing phase examines the functionality of DocWIPE itself.

4.1 Application Implementation

Implementation process starts from the interface design and database. The coding for the application was done using C# programming language and Microsoft Visual Studio 2019 was chosen for the Integrated Development Environment (IDE). The dataset was also prepared and collected from various collection of Microsoft Word files to test the functionality of file wiping module.

Initially, to start using the application, a user needs to create an account by registering in order to login to the application. Username and password of an active account must be provided to login the application. A message will be displayed if the username or password invalid and the field is empty. Once user successfully authenticated to the application, user is directed to main page. This is where user has to choose file wiping button in order to proceed with file wiping process.



C:\Users\User\Documents\Test File\sample0 - (16).docx

```
111x1xx1x1xx1xx111111x1x11x1111xx1x1x11xx11x11xx111xxx1xx1xxxx1x11xx111x11x11xxxxxxx11x111x
xxxxx1xxx11x111x1x111xx1xx111xxx1x1x1xxxxxxx11x1x111xx1111xx11x11xx111xxxx11xxxx11xx1111111
11xxx111xx1111xxx1x1x11xxxx11xx1xxxx1xx111111x111x1x1xxxx1xxx1111x11x11xxx111111x1x11xx11xx
x11xxxx111xx1111x1x1x1111x1x1111xxxx111111xxxxxx1x1xx1xxxx11x1x1x1x1x1x1x11x11xxx1xx1xxxxxxx
11111x11x11xx1x111111xxxx111x1xx111xx11x11x1xx11x111xx1x1xx1xxxx111xxxx1xxxxx11x1111x11xx1
11x11xxx1x1111x1x1x1x11x1xx1111xx11xx1111x1x1xx1xxx1x1x1xx1xx1xx1xxxx1x1xx111x1xxx11xxx1x1xx
1x1xx11xx11x11111xxx11111x1x11xx1x111111xxxx1111x111x1xx1xx1xx1x11x111111xx1xxxx1x1x1x11x1
x1xx11x11xxx1111xx1x1x1x1xx1111x1x1xx11xxx1xx1xxx1x1x11x1xx1x1xxx111x11x1x1xx11x1xx1x1x1xxx1
1xx1x11xxxx1x1x11x1x111x1x1x1111x11x1xx1x1x1xx111xx1xxx1xx1xxx111xx1111xxxx11xxxxx1xx11xx1x1
x1111111xx11x1x1xxxxx111xxxx1x111x1x1xxxxxxxx1x1x1x111x1x1x111x1x1xxx1x1x1x1xx11xx1xx1x11xxx
xx11xx1x111x1xx111x1111x1x111xx1111xxx11xx1x11x11xxx11xx111x1x1x11111x1x111111xx11xxxxxxxx
x11x1x11xx11xx1xx1x1xxx11xxxxxx1xx111x1xxxxxx1x1111xxxxx1x111xxx1xxx1111x11xx1x1xx1x111xxxxx1
11111xx11xxx1x1x1x1xx111x1x1xxxxxxx1x1x111x11x11x1x11xxx1x1xx11111xx11xxx1xx11xx1xxx1x11xxx1
x1xxxx111xx1xx11111xxx11xx111xxxx1111xxx1xx1x1xx111x1x111111xxx111x1x1x1xxxxx11x1x11x11x1x1
1111x111x11xx1x1x1xxxxx1x1xx1x11xxx1111x1xx1xx1xxxx1x111x111111x111x1xx11x111xx1xx1111xxxxx
```

Figure 5: Overwritten file content

The user then can proceed to choose any .docx files that wants to be deleted. Once user chose the file, the file content will be automatically displayed in the box. Next, user has to choose overwrite button to proceed with overwrite process. This is where the file content will be overwritten randomly using randomized 512-gram technique. User able to view the current file content after overwriting process as shown in Figure 5 above. For file deletion, on the other hand, user must choose delete file button to remove file from the device. The deleted file is not supposed to be exists in the file location or recycle bin once it is being overwritten.

4.2 Application Testing

Once the application has been developed, a testing phase is conducted to examine the functionality of DocWIPE. Testing was conducted to identify any errors that occur when using this application. This is to ensure that all modules function as expected and meet the requirements. Table 2 shows the summary of the functional testing results for all modules.

Table 2: Functional test

No.	Module	Test Case	Expected Output	Actual Output	Remark
1	Register	i. Click Register Button	Account registration is carried out. All field must be filled in, valid username and password needs to be provided	As expected	
		ii. Password Validation	An error message will be shown if the password does not meet the complexity requirements. Password must be more than 8 characters, contains uppercase, lowercase, number, and special characters	As expected	
2	Login	i. Click Login Button	Account authentication is carried out	As expected	

Table 2: (cont.)

No.	Module	Test Case	Expected Output	Actual Output	Remark
3	File Wiping	i. Click Add File Button	Users can add Microsoft Word file from their personal computers	As expected	
		ii. Click Clear Button	Clear file content and file path	As expected	
		iii. Click Overwrite Button	Data Wiping process being carried out. Overwritten file is not supposed to be exists in the file location or recycle bin once it is being overwrite	As expected	Tested with file recovery tool named Recuva
4	File Deletion	i. Click Delete File Button	Delete file that has been overwrite	As expected	

The test plan is carried out for register module, login module, file wiping module and file deletion module which consist of different test case. This application performed as intended outcome based on the testing that is conducted.

5. Conclusion

This section concludes overall development, implementation and testing of the application. It includes objectives' achievement, advantages, and disadvantages of DocWIPE.

The modules designed for DocWIPE manage to deliver their purposed successfully. User can add file, delete file, display file before and after overwriting process, rewrite the original content of the file and lastly wipe the file without any remnants left at the file location as well as in the recycle bin. Thus, it can be concluded that DocWIPE has achieved the objectives of its development.

On the other hand, there are several limitations that can be found in the application. Firstly, this application can only allow user to wipe .docx file. This application also limited to overwrite and delete one file at one time only.

There are some aspects of the application that needed improvement in the future in order to enhance the limitation stated. This application is suggested to be able to support other files like Portable Document Format (PDF) and Microsoft Excel Open Xml Spreadsheet (.xlsx). Besides that, an overwrite and delete multiple files at one time features should also be considered.

Acknowledgement

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia for its support and encouragement throughout the process of conducting this study.

References

[1] P. Jain, M. Gyanchandani, and N. Khare, "Big data privacy: A technological perspective and review," J. Big Data, vol. 3, no. 1, 2016.

[2] J. Shu, Y. Zhang, J. Li, B. Li, and D. Gu, "Why data deletion fails? A study on deletion flaws and data remanence in Android systems," ACM Trans. Embed. Comput. Syst., vol. 16, no. 2, pp. 1–22, 2017.

[3] J. P. Van Belle, R. De Beer, and A. Stander, “Anti-forensics: A practitioner perspective,” Int. J. Cyber-Security Digit. Forensics, vol. 4, no. 2, pp. 390–403, 2015.

[4] P. Gutmann, “Secure deletion of data from magnetic and solid-state memory,” in Proceedings of the Sixth USENIX Security Symposium, San Jose, California, July 22-25, 1996, pp. 77–89.

[5] M. Wei, L. Grupp, F. Spada, and S. Swanson, “Reliably erasing data from flash-based solid state drives,” Proc. 9th USENIX Conf. File Storage Technol., pp. 8–8, 2011.

[6] N. A. Yusof, S. Abdullah, M. F. Senan, N. Z. Abidin, and M. B. Sahri, “Data sanitization framework for computer hard disk drive: A case study in Malaysia,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 11, pp. 398–406, 2019.

[7] M. Ölvecký and D. Gabriška, “Wiping techniques and anti-forensics methods,” in SISY 2018: IEEE 16th International Symposium on Intelligent Systems and Informatics, 2018, pp. 127–131, doi: 10.1109/SISY.2018.8524756.

[8] D. Chauhan and P. Bansal, “Study on need of data sanitization and sanitization techniques for memory devices,” Open Access Int. J. Sci. Engineeering, vol. 2, no. 11, 2017.

[9] P. F. Bennison and P. J. Lasher, “Data security issues relating to end of life equipment,” IEEE Int. Symp. Electron. Environ., pp. 317–320, 2004.

[10] B. Schneier, Applied Cryptography, Second Edition: Protocols, Algorithms, and Source Code in C (cloth). John Wiley & Sons, Inc., 1996.

[11] A. Jones and I. Afrifa, “An evaluation of data erasing tools,” J. Digit. Forensics, Secur. Law, vol. 15, no. 1, 2020.

[12] F. Xia, J. Liu, H. Nie, Y. Fu, L. Wan, and X. Kong, “Random walks: A review of algorithms and applications,” IEEE Trans. Emerg. Top. Comput. Intell., vol. 4, no. 2, pp. 95–107, Apr. 2020.

[13] A. Gaeini, A. Mirghadri, G. Jandaghi, and B. Keshavarzi, “Comparing some pseudo-random number generators and cryptography algorithms using a general evaluation pattern,” Int. J. Inf. Technol. Comput. Sci., vol. 8, no. 9, pp. 25–31, 2016.

[14] M. S. Stipčevi´c, Ç. K. Koç, “True random number generators,” in Koç Ç. (ed.) Open Problems in Mathematics and Computational Science, 2014, pp. 275–315.

[15] CBL, “CBL Data Shredder,” CBL Data Recovery, 2019. [Online]. Available: CBL Data Recovery, https://www.cbldatarecovery.com/data-shredder/. [Accessed: 03-Nov-2019].

[16] Eraser, “Eraser – Erase files from hard drives,” Eraser, 2016. [Online]. Available: Eraser, https://eraser.heidi.ie/. [Accessed: 03-Nov-2019].

[17] DBAN, “Data removal: Darik’s Boot and Nuke,” Blancco Ltd., 2019. [Online]. Available: DBAN, https://dban.org/. [Accessed: 03-Nov-2019].

[18] CCleaner, “Recuva,” Piriform Software Ltd., 2016. [Online]. Available: CCleaner, https://www.ccleaner.com/recuva. [Accessed: 09-Jul-2021].

[19] M. Huda, Y. D. S. Arya, and M. H. Khan, “Testability quantification framework of object oriented software: A new perspective,” Int. J. Adv. Res. Comput. Commun. Eng., vol. 4, no. 1, pp. 298–301, 2015.