การประดิษฐ์ซอฟต์แวร์รักษาความปลอดภัยโปรแกรมม้าโทรจันโดยเทคนิคการขุดข้อมูล

Invention Trojan Horse Security Software by Data Mining Technique

Santi Pattanavichai
Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi, RMUTT
39 Village No. 1 Rangsit-Nakornnayok Road, Klong 6 Thanyaburi, Pathumthani, 12110 Thailand
E-mail: pattanavichai@gmail.com , santi_p@rmutt.ac.th

บทคัดย่อ

ปัจจุบันระบบการจัดการความปลอดภัยของข้อมูลเป็นวิธีการหนึ่งที่มีความสำคัญในระบบจัดการทั้งหมดเพื่อความปลอดภัยของข้อมูล ซึ่งการจัดการด้านความปลอดภัยในแง่ของการจัดการการป้องกันเกี่ยวกับมัลแวร์สิ่งสำคัญคือต้องจัดการความปลอดภัย ซอฟต์แวร์รักษาความปลอดภัยม้าโทรจันจัดทำขึ้นโดยมีวัตถุประสงค์เพื่อสร้างโปรแกรมที่ใช้ในการสแกนไวรัสม้าโทรจันโดยการเขียนโปรแกรมเพื่อสร้างโปรแกรมสแกนไวรัสระบบจะค้นหาและตรวจสอบโครงสร้างของไฟล์โดยใช้ เทคนิคการขุดข้อมูล ที่มา เป็นไวรัสหรือไม่ เทคนิคนี้สามารถค้นหาคำทั้งหมดในไฟล์สำหรับการสแกนซึ่งถ้าไฟล์นั้นเป็นโปรแกรมเทคนิคนี้สามารถค้นหาคำในซอร์สโค้ดทั้งหมดได้ หลักการทำงานของโปรแกรมจะดูพฤติกรรมของไฟล์ที่เข้ามาตรวจสอบไฟล์ที่เราจะสแกนว่ามีความสุ่มเสี่ยงที่จะเป็นไวรัสหรือไม่และมีพฤติกรรมอย่างไรในการทำงานร่วมกับระบบคอมพิวเตอร์เช่น ข้อมูลถูกส่งไปยังเครือข่ายหรือไม่หรือมีการใช้ ไฟล์เข้าถึง windows คอมพิวเตอร์ของเราจะมีความสุ่มเสี่ยงที่อาจติดไวรัสหรือโทรจัน

Abstract

Nowadays, information security management systems are the one method importance in all manage system for the information security. In which safety management is in terms of managing protection about malware, it is important to manage security. Trojan Horse Security Software, prepared with the objective to create a program that is used to scan the Trojan horse virus by writing a program to create a virus scanner, the system will find and check the structure of the file by using Data Mining Technique that came as a virus or not. This technic can find all word in the file for scanning, which if the file is the program, this technic can find the word in all source code. The working principle of the program is will look at the behavior of the incoming files, checking for files that we will scan whether there is a random risk of being a virus or not and how to behave in collaboration with computer systems, such as whether the data is sent to the network or not or whether there is a use of the windows access file Our computers will have random risks that may be infected with viruses or Trojans.

Keywords: Trojan Horse, scanner, formatting, deleting, Quarantine, Data Mining Technique

1. Introduction

Nowadays, information security management systems are the one method importance in all manage system for the information security. In which safety management is in terms of managing protection about malware, it is important to manage security. As information becomes more important, information-security management becomes mandatory. Therefore, all of the organizational stakeholders at each level, e.g., staff, management, and board, must be aware of information security [1].

On the other hand, the quality and complexity of threats to information security are increasing daily. Some security threats include data theft, destruction, and computer. Computer-assisted fraud hacking affects organizations in many ways such as destroying the reputation of the organization compromise information loses the confidence of customers and interrupt business [2].

Another interruption to data security is when certainty occurs. The information is accessed by unauthorized persons with spyware. Viruses, malware, denial-of-service attacks, or ransomware [3]. Various motives can stimulate unauthorized access. Information including abuse of authority change information about data theft activities, or /a playful reason [4].

System environments that do not support information security measures increase the chances of unauthorized access.

User to access information Users who lack safety knowledge or awareness also increase their potential. For interrupting information security [5] information security. Interruptions can occur from an unlimited type of media, ranging from manual ones such as mobile flash drives. Network infrastructures such as LAN, WAN and the Internet and various bring your own equipment options [6]

Currently, there are various programs on the online world to facilitate many things, but some programs may have threats that you do not know or accidentally installed may affect our computer, so we should pay attention to the security of the machine. Our computer so we don't become a victim or a threat [7].

Computer virus, or simply in the industry as a virus, is a computer program that invades a computer without the user's consent. Most often they are malicious and damage the system of that computer. In terms of technology, the security of the computer system A virus is a computer program that can make copies of itself. To spread by inserting copies in the executable computer code or document information such as infection [8].

A Trojan horse or Trojan is a type of malware that is often disguised as legitimate software. Trojans can be employed by cyber-thieves and hackers trying to gain access to users' systems. Users are typically tricked by some form of social engineering into loading and executing Trojans on their systems. Once activated, Trojans can enable cyber-criminals to spy on you, steal your sensitive data, and gain backdoor access to your system. These actions can include:

1.1 Deleting data

1.2 Blocking data

1.3 Modifying data

1.4 Copying data

Disrupting the performance of computers or computer networks.

This infected file is called a host. A virus is a type of program type. Malware or malicious programs in the commonly used meaning Viruses also include worms, which are another form of malware. Which sometimes causes computer users to become confused when the term virus is used in a specific meaning of the computer That virus generally does not cause direct damage to the hardware. But will damage the software While typical viruses cause damage (such as data destruction), there are many types that do not cause damage. Only causing annoyance Whether it is a type of virus that causes damage or not Will have a negative effect resulting from the uncontrolled spread of the virus [9].

This paper is organized as follows. Section I introduces the topic and elaborates on the background of the study. Section II describes previous studies related to Information security policy and detecting the malware about Trojan Horse. Section III discusses the proposed methodology about Data Mining Technique, and Section IV describes the application development and evaluation results. Finally, Section V, we discuss the conclusions and proposed future work.

2. Organization of the Text Information Security Policy and Detecting The Malware About Trojan Horse

Information security is one of the most essential elements in securing data because it is responsible for securing all data transmitted over a computer network [10].

Elements of Information Security. Confidentiality, Integrity, and Availability, known as the CIA, are three models designed to guide information security within the organization. Data confidentiality is an important component of data security. The main principles of confidentiality are maintaining the Integrity of the Specifications.

Confidentiality is roughly equivalent to privacy. Measures were undertaken to ensure confidentiality is designed to prevent sensitive information from reaching the wrong people, while making sure that the right people can in fact get it: Access must be restricted to those authorized to view the data in question. In the context of computer systems, allows authorized users to access sensitive and protected data. Specific mechanisms ensure confidentiality and safeguard data from harmful intruders [11].

Integrity involves maintaining the consistency, accuracy, and trustworthiness of data over its entire life cycle. Data must not be changed in transit, and steps must be taken to ensure that data cannot be altered by unauthorized people (for example, in a breach of confidentiality) [12].

Availability is best ensured by rigorously maintaining all hardware, performing hardware repairs immediately when needed and maintaining a correctly functioning operating system environment that is free of software conflicts. It’s also important to keep current with all necessary system upgrades [13].

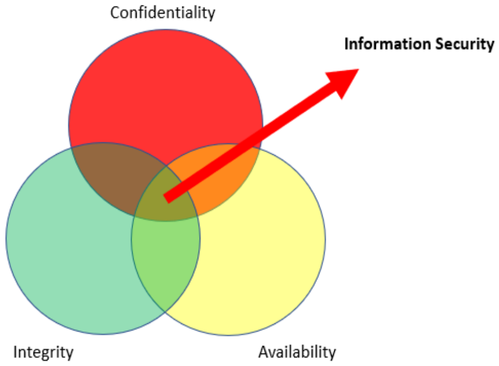


Fig. 1. Elements of Information security

In this context, confidentiality is a set of rules that restricts access to data, that is, ensures that data is reliable and accurate, and that data availability is guaranteed to show reliability. In Fig. 1 [14].

A Trojan horse. A Trojan horse, or Trojan, is a type of malicious code or software that looks legitimate but can take control of your computer. A Trojan is sometimes called a Trojan virus or a Trojan horse virus, but that’s a misnomer. Viruses can execute and replicate themselves. A Trojan cannot. A user has to execute Trojans. Even so, Trojan malware and Trojan virus are often used interchangeably. A Trojan is designed to damage, disrupt, steal, or in general inflict some other harmful action on your data or network. Trojan (Trojan) is a type of program that is designed to conceal and perform certain actions in our machine. From those who do not wish well for the name of this type of program. Comes from the legend of Troy's wooden horse Trojans are attached to e-cards, emails or programs that are available for download on the internet on underground websites. And can enter our machine. In which we are the recipients of it without knowing it Some users who use the internet to experience strange symptoms, such as suddenly the CD-ROMs drive turns on and off. Or the device will play music. But now the machine has an intruder into the machine and has taken control of the user's machine. The ability of this Trojan horse does not have just this. But can break the device and the There are many more Trojan horses. Which, if easily understood, is that an attacker can do anything to every user's device. As if an attacker had sat in front of the user [15].

The 7 Main Types of Trojan Horse.

1. Remote Access Trojan (RAT): Designed to provide the attacker full control of the infected machine. Trojan horse usually masqueraded as a utility.

2. Data Sending Trojan: Trojan horse that uses key logger technology to capture sensitive data like passwords, credit card and banking information, and IM messages, and sends them back to the attacker.

3. Destructive Trojan: Trojan horse designed to destroy data stored on the victim's computer.

4. Proxy Trojan: Trojan horse that uses the victim's computer as a proxy server, providing the attacker an opportunity to execute illicit acts from the infected computer, like banking fraud, and even malicious attacks over the internet.

5. FTP Trojan: This type of Trojan horse uses the port 21 to enable the attackers to connect to the victim's computer using File Transfer Protocol.

6. Security software disabler Trojan: This Trojan horse is designed to disable security software like firewall and antivirus, enabling the attacker to use many invasion techniques to invade the victim's computer, and even to infect more than the computer.

7. Denial-of-Service attack Trojan: Trojan horse designed to give the attacker opportunity to realize Denial-of-Service attacks from victim's computer [16].

Trojans are generally spread by some form of social engineering, for example where a user is duped into executing an e-mail attachment disguised to appear not suspicious. Perform certain actions in our machine from those who do not wish well The name of this type of program. Comes from the legend of Troy's wooden horse Trojans are attached to e-cards, emails or programs that are available for download on the internet on underground websites. And it can enter our machine in which we are the recipients of it without knowing it Trojans are small programs that will be embedded in our machine and will benefit the owner of the Trojan that is sent to us

His benefits are like It may be a key lock program or it will lock the ID or password of some of the programs that they have specified, such as Ragnarok. Whether you go to change the pass 100 times, 1000 times if that Trojan program is still in the machine. When you open the file and enter the id and pass, the Trojan program will start to lock the id and pass. When you connect to the internet, Trojan owners will be able to hack the device, easily get the ID and pass [16].

Trojan horses are programs that are loaded into the computer to perform "secret" operations such as passwords, user names, and personal information about system logins that are typed through the keyboard by users. In most cases, the hacker will send a Trojan horse program. This program enters the computer to capture the information and then used to penetrate the system or to attack the server [17]. By installing effective anti-malware software, you can defend your devices by this program which we create artificial Trojan horse security software by data mining technique.

3. Data Mining Technique

Data mining technique as identified is part of the knowledge discovery process. They can be defined as a set of techniques that help analyze and explore data to find important rules or models hidden within large archives. Using fully or partially automated procedures.

Data mining, which is the process of extracting meaningful knowledge from large quantity of data, has been applied to a wide range of applications in its relatively short existence [18].

Textual Data Mining is "roughly equivalent to text analysis, referring to the acquisition process." High-quality text-based data.” [19] It is also considered the path of the data mining field. When algorithms and data mining methods are used for finding knowledge in textual data. Bulk text mining techniques form can be applied individually or combined with others.

Techniques for discovering reasonable and valuable knowledge Although text data mining has been a lot of success. Some of the more successful applications in the field of in-depth text analysis remain challenging and not dealt The real difference of technology endeavors versus mining text data from general concepts. Of data mining is the possibility of dealing with the effects of implicit data in the form of text. Data that appears to be essentially unstructured However, the general practice of such technological possibilities.

It is still in the early stages to disclose the necessary knowledge drawn from this unstructured text. Information [20] Therefore, the key point that motivates this article is that textual information can be classified as multiple points of view. Very complex implicit data structures in this context, the main problem in dealing with this type of information is design a hybrid technique of data mining algorithms that can manipulate implicit and data.

Further concerted efforts of the Knowledge Search application may reveal the association. Many viewpoints or forms of common data formats are currently available in textual data formats. Methods of identifying data mining methods and less machine learning for these semi-structured learning, or unstructured data types in text data formats [21].

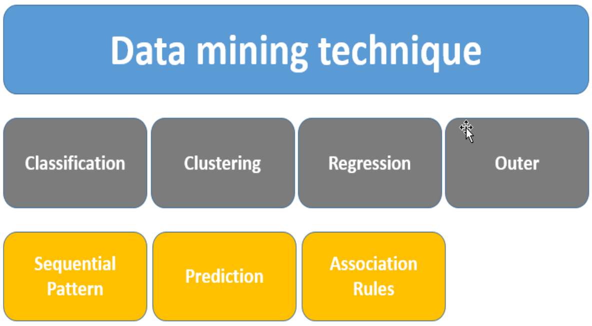


Fig. 2. Data mining technique process

1.Classification:

This analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes.

2. Clustering:

Clustering analysis is a data mining technique to identify data that are like each other. This process helps to understand the differences and similarities between the data.

3. Regression:

Regression analysis is the data mining method of identifying and analyzing the relationship between variables. It is used to identify the likelihood of a specific variable, given the presence of other variables.

4. Association Rules:

This data mining technique helps to find the association between two or more Items. It discovers a hidden pattern in the data set.

5. Outer detection:

This type of data mining technique refers to observation of data items in the dataset which do not match an expected pattern or expected behavior. This technique can be used in a variety of domains, such as intrusion, detection, fraud or fault detection, etc. Outer detection is also called Outlier Analysis or Outlier mining.

6. Sequential Patterns:

This data mining technique helps to discover or identify similar patterns or trends in transaction data for certain period.

7. Prediction:

Prediction has used a combination of the other data mining techniques like trends, sequential patterns, clustering, classification, etc.

It analyzes past events or instances in a right sequence for predicting a future event [23].

Therefore, this research focuses on the difficulty of improving the accuracy of this information for classification and clustering. The classification of sentimental textual information through the file is the program, the first step scans the information in source code of the program and the Trojan horse virus scanning checks the file extension that it is these files about .txt, .pdf, and .docx. If yes, these files will not scan these files. The clustering groups the dataset for scanning the big data which contains a lot of information in Fig. 2.

4. Process of Creating Trojan Horse Security Software by Data Mining Technique

Data Mining is the process of dealing with large amounts of data in order to find the patterns and relationships hidden in that data set. Data mining, also known as Knowledge Discovery in Databases (KDD) is a technique for automatically finding patterns from large amounts of data. By using algorithms from statistics Machine learning and pattern recognition or in another definition Data mining is the process of dealing with data (most often in large numbers) to find patterns, approaches, and relationships hidden in that data set. Based on statistical recognition, machine learning and mathematical principles [24].

Establishing a Trojan horse virus scan program by using Visual Studio 2017 and C# in program development to create Trojan Horse Security Software.

Flowchart of processing Trojan Horse Security Software. Flowchart is an image or symbol. That is used to write in place of text descriptions or words used in algorithms Because to understand the simple and consistent sideburns The use of words or text. May be more difficult than using images or symbols. Describe in Fig. 3.

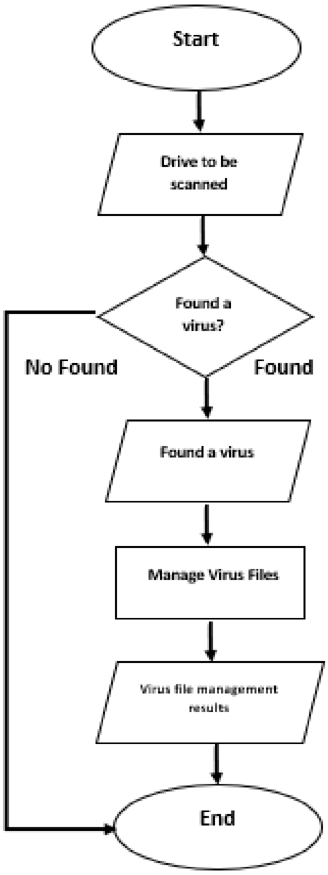


Fig. 3. Flowchart operation of the Trojan horse virus scanner program

The work is designed using Flowchart how each step or method is carried out and the design of the program page by creating the user interface of the Trojan horse virus scanner program. Describe in Fig. 4.

A technique that uses a scanner to search for files that are identified as being hidden by a virus in memory and scan the word about Virus, Trojan, and Hack in all areas of the file and include source code of the file which is the program.

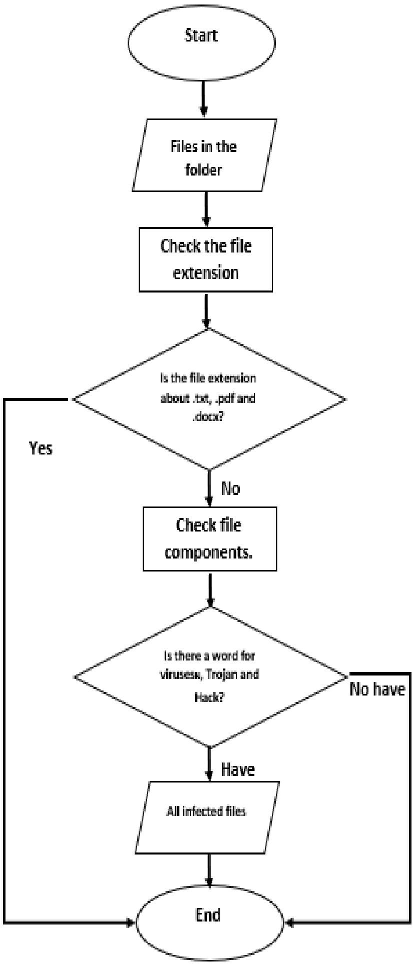


Fig. 4. Flowchart operation of the Trojan horse virus scan

The first step for operation of the Trojan horse virus scanning checks the file extension that it is these files about .txt, .pdf, and .docx. If yes, these files will not scan these files.

The second step for operation of the Trojan horse virus scanning to manage the files virus two choices to select to manage foe delete Virus files and Quarantine Virus files. Describe in Fig. 5.

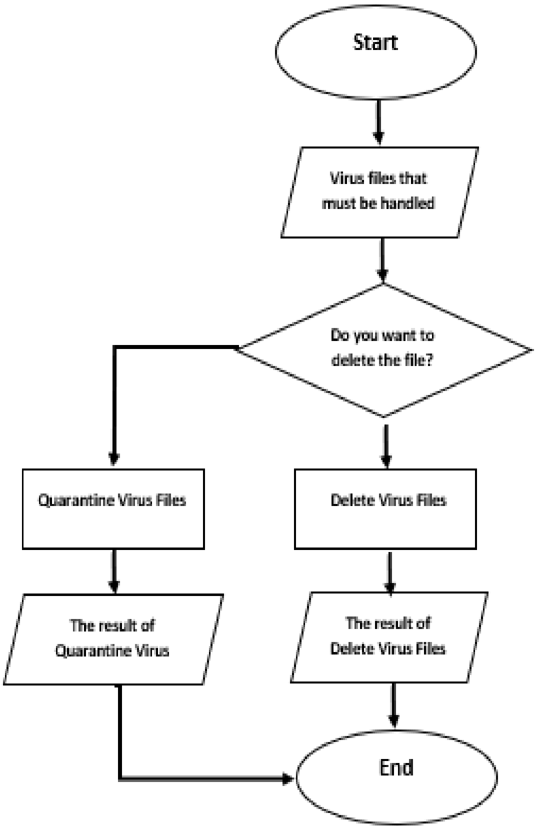

Start
Virus files that must be handled
Do you want to delete the file?
Quarantine Virus Files
Delete Virus Files
The result of Quarantine Virus
The result of Delete Virus Files
End

Fig. 5. Flowchart operation of virus management

Steps for using Trojan Horse Security Software

4.1 The main screen of the program

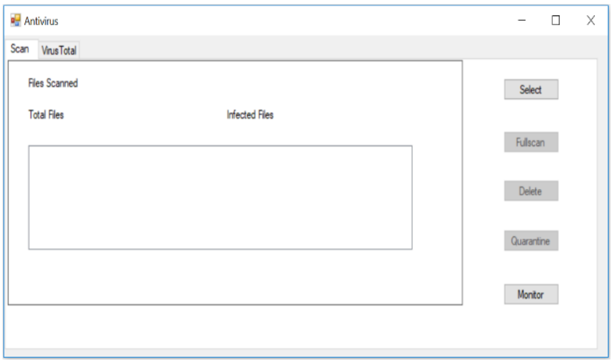


Fig. 6. The main screen of the program

4.2 Click the Select button to select the drive or folder that you want to scan.

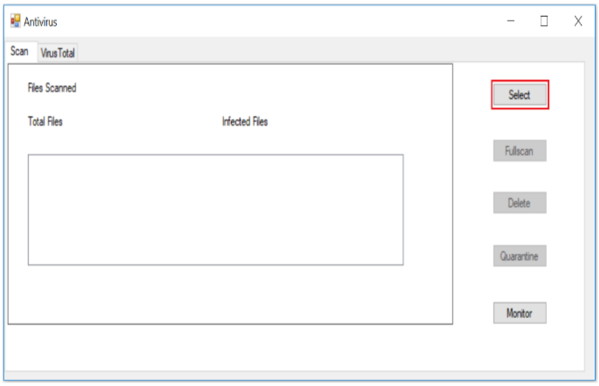


Fig. 7. Click the Select button to select the drive or folder that you want to scan

4.3 When you have finished selecting the drive or folder click the OK button to start the scan.

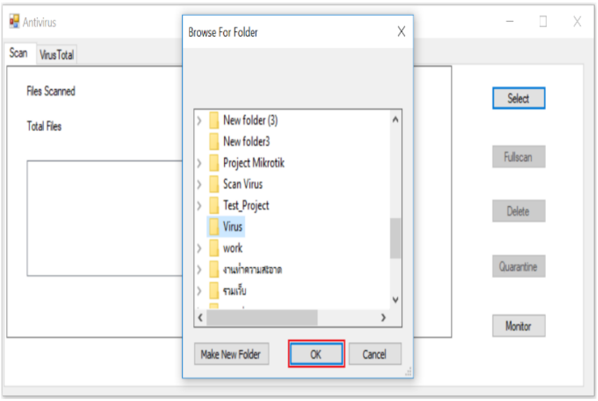
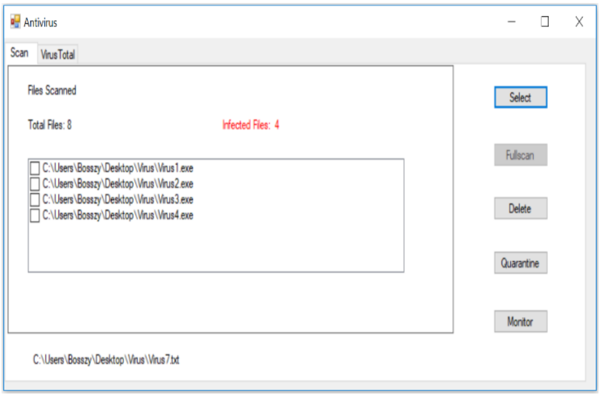


Fig. 8. Click the Select button to select the drive or folder that you want to scan

4.4 When the scan is complete, the results will be displayed as the image.



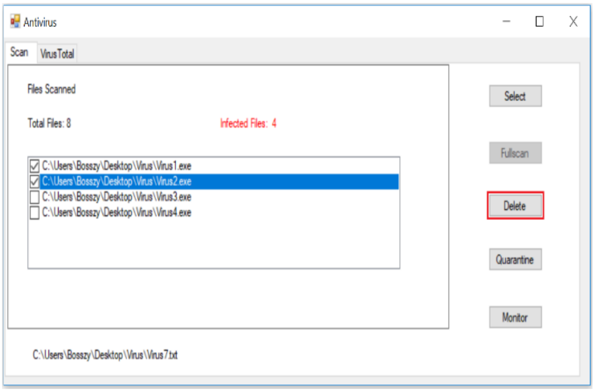
```
Total Files: 8          Infected Files:  4

□ C:\Users\Bosszy\Desktop\Virus\Virus1.exe
□ C:\Users\Bosszy\Desktop\Virus\Virus2.exe
□ C:\Users\Bosszy\Desktop\Virus\Virus3.exe
□ C:\Users\Bosszy\Desktop\Virus\Virus4.exe

C:\Users\Bosszy\Desktop\Virus\Virus7.txt
```

Fig. 9. Show scan results will identify which files are viruses

4.5 Then can click the Delete button to delete the selected infected file.



Files Scanned

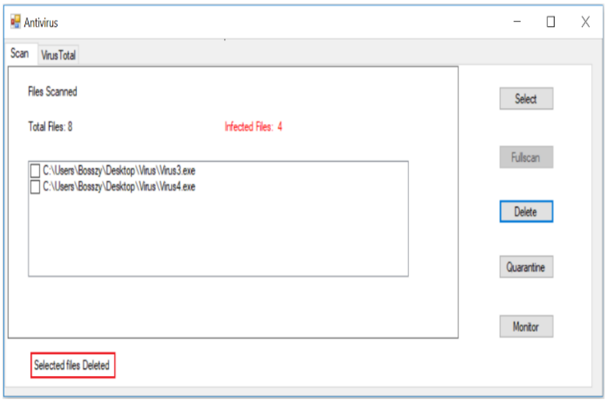
Total Files: 8 Infected Files: 4

☑ C:\Users\Bosszy\Desktop\Virus\Virus1.exe
☑ C:\Users\Bosszy\Desktop\Virus\Virus2.exe
☐ C:\Users\Bosszy\Desktop\Virus\Virus3.exe
☐ C:\Users\Bosszy\Desktop\Virus\Virus4.exe

C:\Users\Bosszy\Desktop\Virus\Virus7.txt

Fig. 10. Click the Delete button for the virus file that we want to delete

4.6 When pressing the Delete button to delete the selected infected file already Will display the message "Selected file Deleted".



Files Scanned

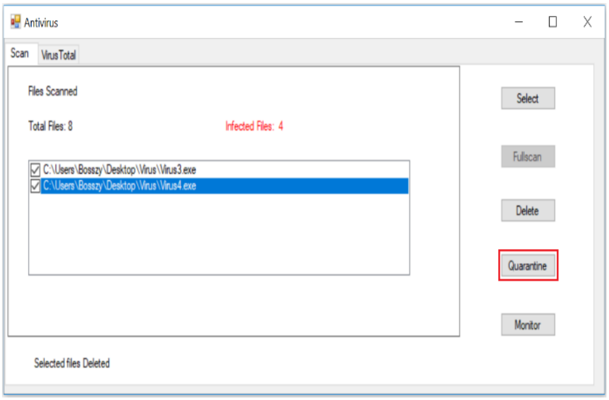
Total Files: 8 Infected Files: 4

☐ C:\Users\Bosszy\Desktop\Virus\Virus3.exe
☐ C:\Users\Bosszy\Desktop\Virus\Virus4.exe

Selected files Deleted

Fig. 11. Show "Selected file Deleted" message is to delete the selected virus file just a moment ago

4.7 Then can click the Delete button to delete the selected infected file.



Files Scanned

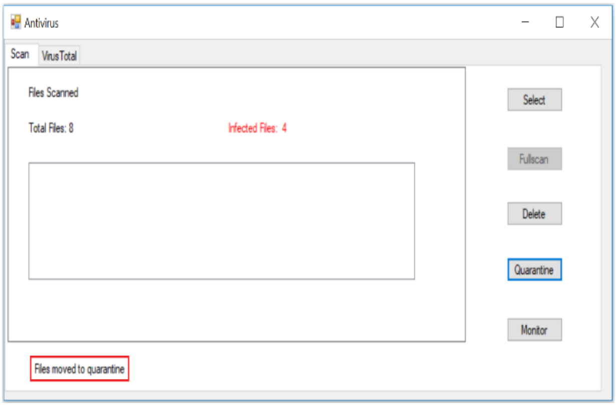
Total Files: 8 Infected Files: 4

☑ C:\Users\Bosszy\Desktop\Virus\Virus3.exe
☑ C:\Users\Bosszy\Desktop\Virus\Virus4.exe

Selected files Deleted

Fig. 12. Click the Quarantine button to Disabled the selected infected file.To move to the quarantined folder

4.8 When the Quarantine button is pressed to quarantine the selected infected file, the message "Files moved to quarantine" appears.".



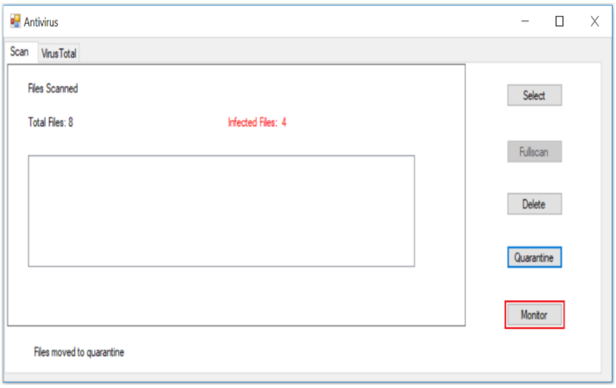
Files Scanned

Total Files: 8 Infected Files: 4

Files moved to quarantine

Fig. 13. Will display the "Files moved to quarantine" message. Meaning that the selected virus files were moved to the quarantined folder

4.9 After that, we can press the Monitor button to view the file quarantine information.



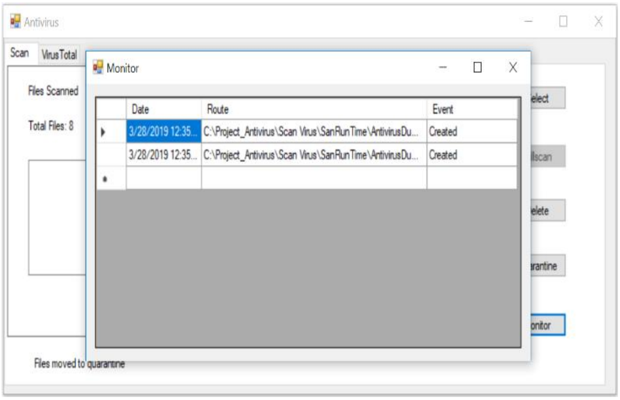
Total Files: 8

Infected Files: 4

Files moved to quarantine

Fig. 14. Click the Monitor button to see the files that we have quarantined.

4.10 Click on the Monitor button. The Monitor window will appear.



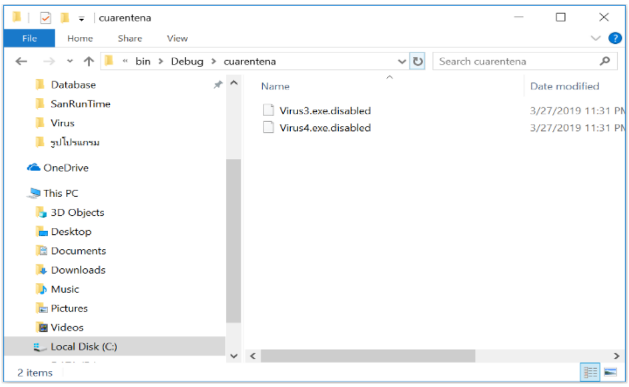
Total Files: 8

	Date	Route	Event
▶	3/28/2019 12:35...	C:\Project_Antivirus\Scan Virus\SanRunTime\AntivirusDu...	Created
	3/28/2019 12:35...	C:\Project_Antivirus\Scan Virus\SanRunTime\AntivirusDu...	Created
*			

Files moved to quarantine

Fig. 15. Show the Monitor window and the virus file that we have quarantined.

4.11 Quarantined infected files will be moved to the cuarentena folder in the disabled state



Name	Date modified
Virus3.exe.disabled	3/27/2019 11:31 P
Virus4.exe.disabled	3/27/2019 11:31 P

Fig. 16. The infected files that are quarantined will be moved to the cuarentena folder.

This program can use for scanning all files although the file is the program or not but that it is these files about .txt, .pdf, and .docx. If yes, these files will not scan these files.

In the scanning, will find and check the structure of the file by using Data Mining Technique that came as a virus or not. This technic can find all word in the file for scanning, which if the file is the program, this technic can find the word in all source code.

5. Conclusions

This paper presents an outline of a program in Systems and Software to scan the malware which specific for Trojan Horse by using Data Mining Technique in this program. This program can scan to find any word in all source code which that file is the program.

The working principle of the program is will look at the behavior of the incoming files, checking for files that we will scan whether there is a random risk of being a virus or not and how to behave in collaboration with computer systems, such as whether the data is sent to the network or not or whether there is a use of the windows access file Our computers will have random risks that may be infected with viruses or Trojans. A technique that uses a scanner to search for files that are identified as being hidden by a virus in memory and scan the word about Virus, Trojan, and Hack in all areas of the file and include source code of the file which is the program.

Summary of trial results of the Trojan horse virus scanner program from the trial of the Trojan horse virus scan program inside the computer equipment Can detect files that are at risk of being infected by viruses. Once the virus is found, the program can quarantine or delete the virus. and the Trojan horse virus scanner can also choose which folder or drive we want to scan for viruses The Trojan horse virus scanner can also be used easily and conveniently as well.

Command speed Depending on the equipment and specifications of the computer that is operating at how fast. If there is work that is too slow May cause the computer equipment to freeze.

The future work, we can apply this program to find any word in the file which you want to find any word in the file for scanning by using Data Mining Technique. Data mining, which is the process of extracting meaningful knowledge from large quantity of data, has been applied to a wide range of applications in its relatively short existence.

Acknowledgements

This research was financially supported by Rajamangala University of Technology Thanyaburi, RMUTT.

References

[1] Candra J. W., Briliyant O. C., and. Tamba S. R, “ISMS planning based on ISO/IEC 27001:2013 using analytical hierarchy process at gap analysis phase (Case study : XYZ institute),” in Proceedings of the 2017 11th International Conference on Telecommunication Systems Services and Applications, TSSA 2017, 2018, vol. 2018-January, no. 4, pp. 1-6. DOI: 10.1109/IWBIS.2018.8471700

[2] Achmadi D., Suryanto Y., and Raml K. “On Developing Information Security Management System (ISMS) Framework for ISO 27001- based Data Center,” in 2018 International Workshop on Big Data and

Information Security, IWBIS 2018, 2018, pp. 149-157.
DOI: 10.1109/IWBIS.2018.8471700.

[3] Department for Digital Culture Media and Sport, "Cyber Security Breaches Survey 2016," 2018.
DOI: 10.13140/RG.2.1.4332.6324.

[4] F. G. I. T. U.S. Congress, Office of Technology Assessment, "Electronic Record Systems and Individual Privacy," no. June, 1986. DOI: 10.1016/0167-4048(86)90061-1.

[5] Glaspie H. W. and Karwowski W. "Human Factors in Information Security Culture: A Literature Review," in Proceedings of the AHFE 2017 International Conference on Human Factors in Cybersecurity, 269-280, June 2017. DOI: 10.1007/978-3-319-60585-2_25.

[6] Retnowardhan A., Diputra R. H., and Triana Y. S., "Security risk analysis of bring your own device (BYOD) system in manufacturing company at Tangerang," TELKOMNIKA (Telecommunication Comput. Electron. Control.), Vol. 17, No. 2, 753, Apr. 2019. DOI: 10.12928/ telkomnika. v17i2.10165

[7] Bedi P., Gandotra V., Singhal A., Vats V., and Mishra N., (2009) "Avoiding Threats Using Multi Agent System Planning for Web Based Systems". 1st International conference on Computational Collective Intelligence – Semantic Web, Social Networks and Multiagent Systems, Wroclaw, Poland, October 2009, LNAI, Springer-Verlag Berlin Heidelberg, 709-719,

[8] Information on https://www.secure-cyber.net/integrity-checker/

[9] Information on http://parum.forumotion.net/t34-topic,

[10] Chen S., Iyer R., and Whismant K., (2002) "Evaluating the Security Threat of Firewall Data Corruption Caused by Instruction Transient Errors," In Proceedings of the 2002 International Conference on Dependable Systems & In Proceedings of the 2002 International Conference on Dependable Systems & Network, Washington, D.C.,

[11] Kim H., (2004.) "Design and Implementation of a Private and Public Key Crypto Processor and Its Application to a Security System," IEEE Transactions on Consumer Electronics, Vol. 50, No. 1.

[12] Information on https://www.techopedia.com/definition/10254/confidentiality

[13] Information on http://whatis.techtarget.com/definition/Confidentiality-integrity-and-availability-CIA

[14] Information on http://whatis.techtarget.com/definition/Confidentiality-integrity-and-availability-CIA

[15] Pattanavichai S., Prasartkaew C. (2017) “Design Network Model for Good Performance of Network Security,” ICT: Big Data, Cloud and Security (ICT-BDCS 2017), Singapore, 114-118.

[16] Information on https://owasp.org/www.community/attacks/Trojan_Horse

[17] Shusen W., and Ping C. (2012) “The Trojan horse attack principle and control strategy of [J]”. software guide, 2012 (6): 146-148.

[18] Lei D., (2008) “Analysis and prevention of computer viruses [J]”, Silicon Valley, (22): 23-24.

[19] Hormanzi A. and Giles S. (2004) “Data mining: a competitive weapon for banking and retail industries”. Information Systems Management, Vol 21. No 2, 62–71.

[20] Solka J. L. (2008) “Text data mining: Theory and methods,” Statistics Surveys, Vol. 2, 94-112.

[21] Amado A., Paulo Cortez, Paulo R., Sergio M. (2018) “Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis,” European Research on Management and Business Economics, Vol. 24, No. 1, 1-7.

[22] Hashimi H., et al. (2015) “Selection criteria for text mining approaches,” Computers in Human Behavior, Vol. 51, 729-733.

[23] Zablith F. and IOsman. H. (2019) “Review Modus: Text classification and sentiment prediction of unstructured reviews using a hybrid combination of machine learning and evaluation models,” Applied Mathematical Modelling, Vol. 71, 569-583.

[24] Information on https://www.guru99.com/data-mining-tutorial.html

[25] Usama F.; Piatetsky-Shapiro G., and Smyth P. (2018) "From Data Mining to Knowledge Discovery in Databases".
http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf
Retrieved,12-17.