CSE 509, SYSTEM SECURITY PROJECT

ANTIVIRUS FOR LINUX

TEAM MEMBERS:

Pratik Gaikar

Nehil Shah

Tushar Jain

Vishwajit Bhat

Table of Contents

Contents

Problem Statement:	4
Tools and Technologies:	4
Design:	4
Modes of Operation:	4
On demand mode:	4
On access mode:	5
Linux System Call Hijacking:	5
Whitelist:	5
Blacklist:	6
Netlink socket:	6
Implementation:	6
Hijacking:	6
main.c	6
Anti Virus Helper:	6
anti_virus_helper.c	6
file_operation.c	7
Whitelist:	7
whitelist.c	7
Blacklist:	7
blacklist.c	7
Netlink Socket	7
main.c	7
user.c	7
User Scan (On Demand Scan)	8
antivirus_check.c	8
Scripts:	8
Installation & Testing:	
Distinguishing Features:	
Netlink API:	
Page Size blacklist checking:	10

Whitelisted all the libc files:	10
Linux kernel version 4.2:	10
Block Diagram	11
Antivirus Module	11
On Demand Scan	12
On Access Scan	13
Snapshots:	14
Who Did What	21
Citations	22

Problem Statement:

Develop an antivirus scanner that supports updates over the Internet. The antivirus would be able to detect malware based on signatures of malicious programs. Each signature is a series of sequential bytes from the assembly instructions of an executable. The antivirus would run in two scanning modes, on-demand and on-access.

In On-demand scanning, the user can run a program to scan a file, or a directory containing files, e.g. antivirus-scan /home/jack/Downloads. If a file is found to be infected it removes all permissions (e.g. chmod 000 virus), append a ".virus" string in the file's name and alert the user.

In On-Access Scanning, when the user is trying to execute, or open a file, we scan the file before it is opened. If the file is found to be virus, then the kernel would refuse to open the file, removes all permissions, appends a ".virus" string in the file's name and alert the user.

Tools and Technologies:

VMware workstation pro. Kernel version: 4.2.0.27 Visual Paradigm 13.2

Design:

The detection logic of antivirus is implemented in the kernel-space. The logic is at only one location in the kernel and is called by both the on demand and on access modes. The system call table is hijacked to perform an additional operation in the **open, execve, openat and execveat** system calls. Before actually opening the file, it is checked whether the file is a virus file. If we find that the file is not a virus we go ahead and continue opening the file. That would return file's "fd" to the caller and thus successfully **open** the file. If the file is a virus file, we rename the file **appending ".virus",** remove the permissions on the file, and notify the user of the file being a virus file. Thus the file couldn't be opened.

Modes of Operation:

On demand mode:

- Input is a directory: When the input given is a directory, then the user program recursively goes through each of the files and subdirectories in the given directory, and opens each of the file. When a file is opened, it goes to the hook in the kernel space and checks for virus. It returns a negative value and a errno in case of a virus and "fd" if not a virus. If the file was successfully opened, we close that file and move to the next file. When the file was found to be virus, the kernel communicates the user through the use of netlink socket. A user program listens to a port and gets notified as soon as the kernel writes the virus file name. Send notify is used to show virus file name graphically to the user.
- Input is a file: Just open the file which hits the hook in the kernel and the operation is same as above.
- Input can be multiple files and directories, in which all the given files and directories are scanned for virus. For each directory we recursively go through all the files same as the above case 1.

On access mode:

- When a user opens a file using any editor or through some user program, the hook in the kernel is called which checks the file for virus and allows opening only if it's not a virus. In that case a valid "fd" (file descriptor) is returned. If the file is virus, negative value is returned which does not open the file.
- The use of netlink is important in this mode as we need a way to notify to user about virus from kernel, which can thereby provide a pop up to the user about the virus file.
 Without netlink, In case of file being virus, the file would not be opened and there would be no pop up, which might make the user feel clueless. If just printk statements were used then the user would have had to execute dmesg to know about the virus file.

Linux System Call Hijacking:

The system call table is made read write and the following four system calls are hijacked:

a. Open system call

Whenever any file is opened using either the on demand or the on access mode, the open system call is invoked and by hijacking it, we first scan the file to be opened. If virus is found, we return an error and do not open the file and if no virus is found then the file is opened regularly.

b. Openat system call

The openat system call operates in exactly the same way as open system call with the difference that openat opens a file relative to a directory file descriptor. Whenever any file is opened using either the on demand or the on access mode, the openat system call is invoked and by hijacking it, we first scan the file to be opened. If virus is found, we return an error and do not open the file and if no virus is found then the file is opened regularly.

c. Execve system call

Whenever any command or file is executed using the on demand or the on access mode, the execve system call is invoked and by hijacking it, we first scan the file to be executed. If virus is found, we return an error and do not execute the file and if no virus is found then the file is executed regularly.

d. Execveat system call

The execveat system call operates in exactly the same way as execve system call with the difference that execveat opens a file relative to a directory file descriptor. Whenever any command or file is executed using the on demand or the on access mode, the execveat system call is invoked and by hijacking it, we first scan the file to be executed. If virus is found, we return an error and do not execute the file and if no virus is found then the file is executed regularly.

Whitelist:

Whitelist file contains the SHA-1 hashes of programs and files that are definitely not malware. This whitelist is populated with the standard programs and files of a Linux

distribution to ensure that the antivirus will never break the system. If an input file's hash is on that list, then the file is not scanned against blacklist further.

Blacklist:

Blacklist consists of a set of sequences of bytes which are usually found in malicious programs and files. So, to detect if a program or a file has virus or not, we can do a byte by byte comparison of the virus pattern with the file content and check if the pattern mentioned in blacklist is present in them or not. If present it means that the it is a virus file. Otherwise, it is not a virus file. Since all the possible patterns of viruses are not known this method is not full proof but it can help detecting known viruses.

Netlink socket:

Netlink socket is a special IPC used for transferring information between kernel and user-space processes. It provides a full-duplex communication link between the two by way of standard socket APIs for user-space processes and a special kernel API for kernel modules. It is a nontrivial task to add system calls, ioctls or proc files for new features; we risk polluting the kernel and damaging the stability of the system. The code implementing a system call in the kernel is linked statically to the kernel in compilation time, thus, it is not appropriate to include system call code in a loadable module. With netlink socket, no compilation time dependency exists between the netlink core of Linux kernel and the netlink application living in loadable kernel modules.

We used netlink to communicate about the virus file from kernel space to user space. Once a virus is detected in the kernel space, it will notify the user through the user socket. The user program in turn uses "send-notify" to show about the virus graphically.

Implementation:

Below we have the details of the files in each module and their significance:

Hijacking:

main.c

Once the module is initialized, the system call address is retrieved from the kernel and it is changed from read only to read write in order to intercept it. Then, the original address of the open, openat, execve and execveat system calls are stored and replaced with our new intercepted system call implementations for the open, openat, execve and execveat system calls respectively. For each of these system calls, we first scan the file for virus and if a virus file is detected we return the -EBADF error, otherwise the original system call implementation is invoked to carry out the further process. Once the module is exiting, we again replace our new intercepted system calls with the original system call implementations so that the kernel returns to original state.

Anti Virus Helper:

anti_virus_helper.c

Whenever the new intercepted system call implementation is invoked, the antivirus helper function check_for_virus takes the appropriate parameters from the system call and processes

it. It opens the blacklist and whitelist files in read mode for virus scanning. It also opens the input file with the parameters provided and then the opened file proceeds for scanning. Before the file proceeds for whitelist and blacklist scanning, it is checked whether it's a character or a block device file, such type of files cannot be opened so they can't be scanned. After the file is checked with the whitelist and the blacklist pattern, if the file contains a virus then the rename function is called which renames the file with .virus and sets its permissions to 000.

file_operation.c

This helper file contains file operations methods for opening a file, renaming the file, changing the permissions of the file.

Whitelist:

whitelist.c

This file scans the input file and checks whether its SHA-1 hash is present in the whitelist patterns. The whitelist file is already populated with the SHA-1 of standard linux utilities, i.e. the programs and files of a Linux distribution to ensure that your antivirus will never "break" the system. The input file's contents are read and its sha1 is calculated from its contents. The calculated sha1 is then checked whether it matches with any of the whitelist patterns present in the whitelist file. If a match is found, then the input file is interpreted as a safe file and is not further checked for blacklist patterns. If no match occurs, then the input file is further checked for blacklist pattern.

Blacklist:

blacklist.c

Once the file is checked for whitelist pattern and no match is found then the input file is checked whether its contents matches with one of the blacklist patterns. The blacklist file has all the virus patterns and each pattern is a series of sequential bytes. The input file is scanned and if its contents match with one of the virus patterns then a virus would be detected and that file would be treated as a virus file. If the input file does not match with any of the virus patterns then no action is executed on the file. On detecting a file as a virus, that file is renamed to .virus, its permissions are set to 000 and if we try to open or execute this file an error "Bad File Descriptor" (-EBADF) would occur indicating that the file could not be opened.

Netlink Socket

main.c

The communication between kernel and user space is achieved through socket netlink. Once the antivirus module is initialized, the socket is created between the kernel side and the user side. Whenever, we want to notify the user regarding a virus file, we'll send this message out from kernel space to the user space on this socket connection. Once, the module is uninstalled, the socket connection between the kernel and user is terminated.

user.c

This file contains the user space code of the socket connection between the kernel and the user. It listens for any messages coming from the kernel and if any message arrives, it notifies the user

by creating a pop up on the screen so that the user is correctly aware if any virus file is found while on demand or on access scanning.

User Scan (On Demand Scan)

antivirus_check.c

This file is the user space program for the functionality of on demand scanning. The user executes this program in order to scan certain files or directories. The user can provide any number of files or directories while executing this program and all these inputs provided would be scanned for any virus in them. We used nftw library for recursively iterating within the directories so that we could get all the files within the directory provided as input by the user. Each file is scanned and if virus occurs in any of them then the user is appropriately notified about this. Also, a summary is provided to the user with the number of files scanned and the number of virus files detected.

Scripts:

antivirus update

This script updates the existing blacklist and whitelist files stored at a server. By executing this script, the updated blacklist and whitelist files at a server would replace the current files. Also, the SHA-1 hash of the blacklist and whitelist files are added to the whitelist file.

antivirus.config

This file contains the server path where the blacklist and whitelist files are stored.

antivirus_install

This script performs the following functions:

- **a.** It compiles all the user program files and copies its executables to /usr/bin folder by which the files can be executed from anywhere in the kernel.
- **b.** It creates a directory /etc/antivirusfiles which would contain the blacklist and the whitelist files.
- **c.** It copies the antivirus_update script to /usr/bin and executes it by which the blacklist and whitelist kept at a server are copied into the /etc/antivirusfiles location It builds and starts the execution of the kernel module.
- **d.** It calculates the SHA-1 hash of the whitelist and blacklist files and adds it to the whitelist patterns.

sha1sum.sh

This script calculates the SHA-1 hashes of the linux internal files and utilities.

antivirus_uninstall

This script is used to uninstall the antivirus module.

testprogram

This file is used for programmatically testing all the open and execve related system calls by calling all of them within a C program (testprogram.c)

Installation & Testing:

The antivirus software can be installed only if the user has root privileges. Antivirus scan can be executed by normal user.

Here are the steps to get the antivirus code and install antivirus:

1. The module is installed on the VMWARE virtual machine with linux kernel version 4.2.0-27-generic.

Run the following git clone command in the /usr/src/linux-headers-4.2.0-27-generic folder to clone the code present in the git repository LinuxAntivirus:

git clone https://github.com/pratikgaikar/LinuxAntivirus.git

2. To update the whitelist and blacklist file, the following script is to be executed when the antivirus module is not inserted:

antivirus_update

3. To install antivirus run the following script from the directory /usr/src/linux-headers-4.2.0-27-generic/LinuxAntivirus/scripts/:

antivirus_install

- 4. After the script is executed, the antivirus module is inserted into the kernel and scanning starts
- 5. The on demand mode of antivirus scan can be executed using the following command:

antivirus scan testfiles/

This would scan the whole of testfiles directory.

antivirus scan testfiles/testfile1

This would scan only testfile1 of testfiles directory.

antivirus_scan testfiles/ testfiles2/

This would first scan testfiles directory, then scan testfiles2 directory.

- 6. Every file which is opened or executed by the user is scanned for virus and appropriate action is taken.
- 7. Run cat command to open a virus file On Access mode cat testfiles/testfile1
- 8. The module could also be tested by executing the following command within the /usr/src/linux-headers-4.2.0-27-generic/LinuxAntivirus/scripts/ folder by which all the open and execve related system calls are invoked

./testprogram

Distinguishing Features:

Netlink API:

We have implemented the virus scan in kernel space. This is useful as a particular user might not have access to some directories and files, and thus makes sense to the scan in kernel space. Since we come to know whether a file is a virus or not in the kernel space, we need a way to communicate this to the user space. Various ways to do it are:

Shared Memory: We can keep a file in which the kernel writes and a user program continuously runs reading the file. This involves IO operation and also a performance hit as the User program has to continuously run reading the file.

Notify via socket: We used server and client sockets to communicate from kernel to user space. Once a virus is detected in the kernel space, it writes it to the server socket. The user socket listens to the client port, and gets notified when something is written to its port.

In this way, we avoid IO operation and also avoid user continuously checking if virus is found. This optimizes the performance of the antivirus.

Page Size blacklist checking:

Since kernel uses physical memory, a huge file would not get completely loaded into memory due to limited main memory. So, we read a file into a buffer of PAGE SIZE and compare it with the blacklist bytes. We reuse the same buffer to continue the following bytes of the file. This makes sure that we use only PAGE SIZE of memory to compare the whole file.

Whitelisted all the libc files:

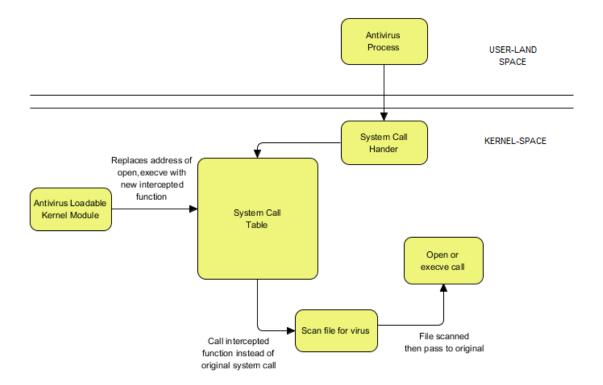
Since we have a hook in 4 system calls, various applications come across this hook and a huge number of system files will encounter this hook.

Linux kernel version 4.2:

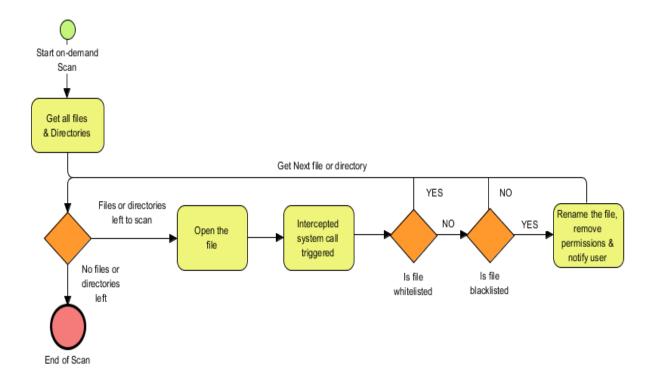
Implemented antivirus on closer to the latest official release version of kernel.

Block Diagram

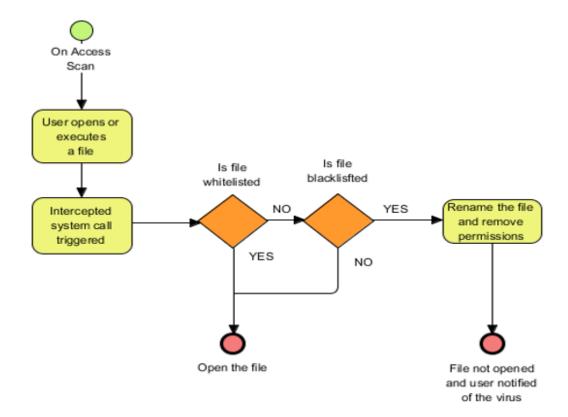
Antivirus Module



On Demand Scan

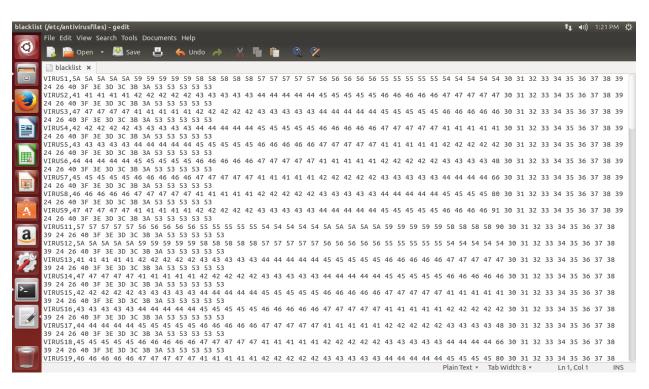


On Access Scan

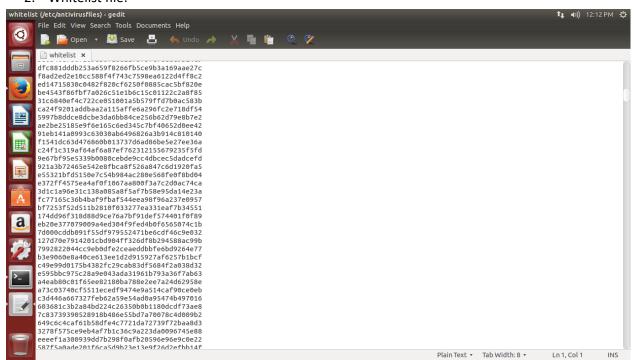


Snapshots:

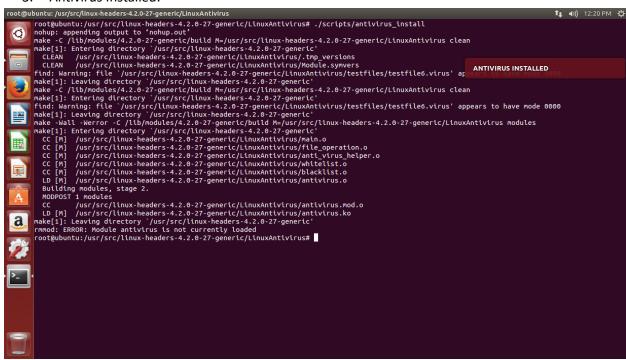
1. Blacklist file:



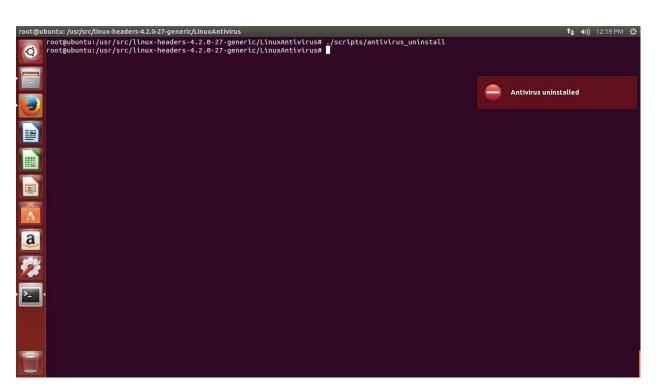
2. Whitelist file:



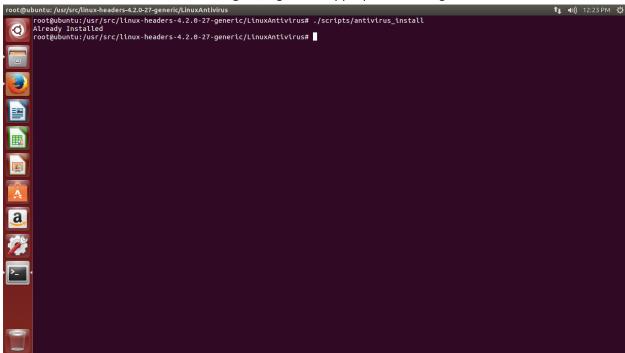
3. Antivirus Installed:



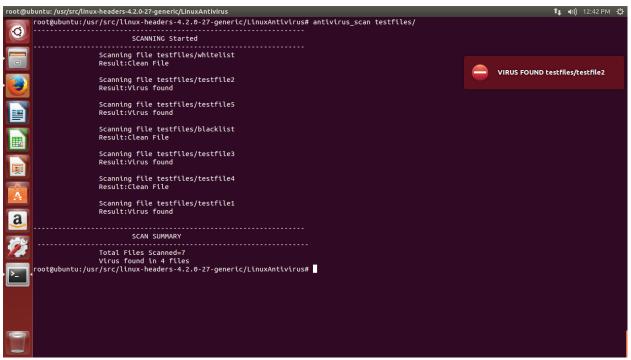
4. Antivirus Uninstalled:

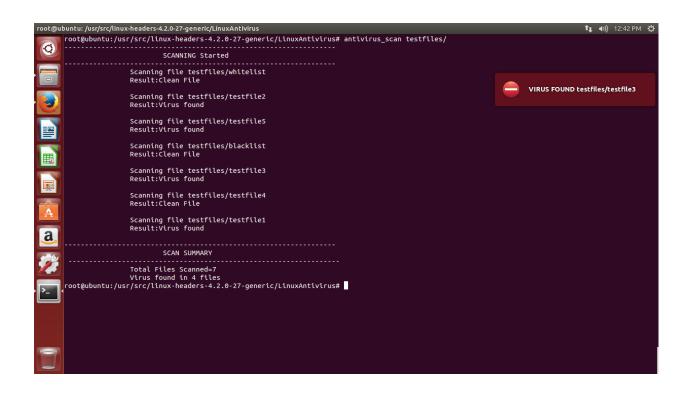


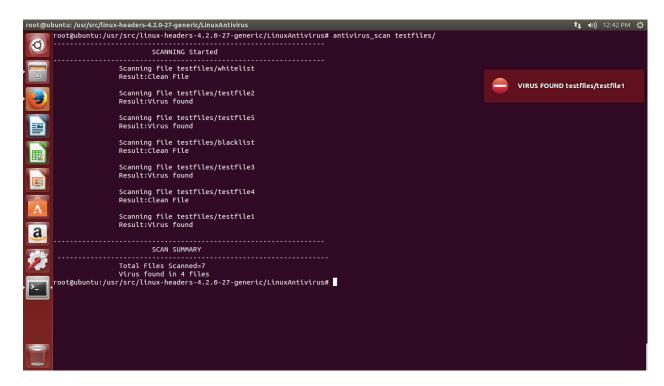
5. When tried to antivirus install again, it gives an appropriate message:



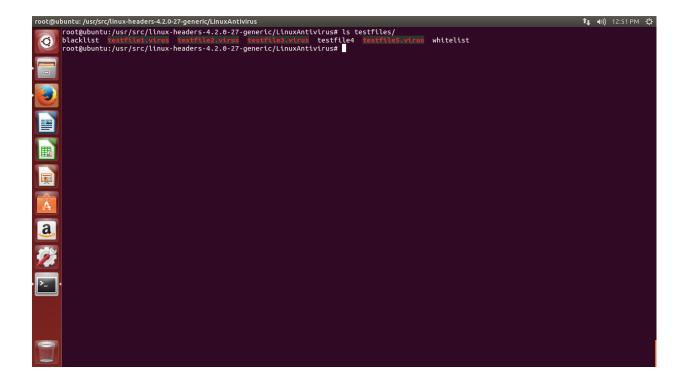
6. On Demand scan: User notified of virus files:



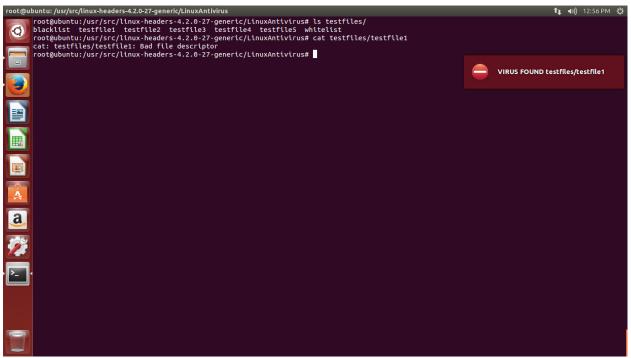




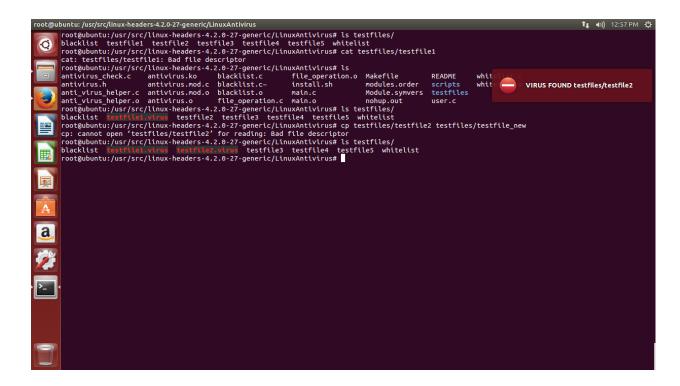
7. All virus files are renamed, permissions removed and thus colored red:



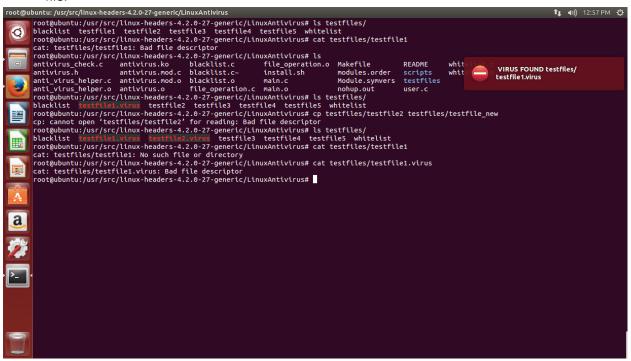
8. On Access mode: When tried to cat a virus file:



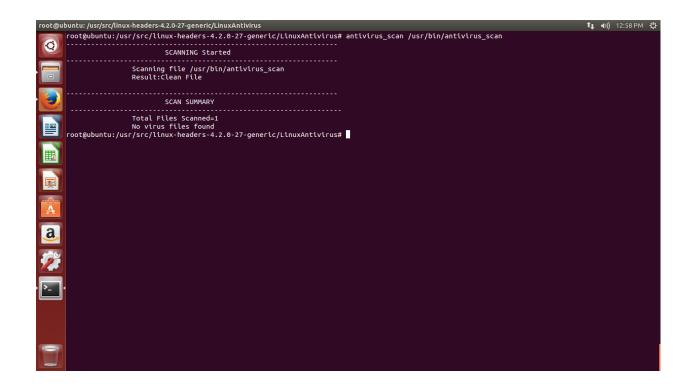
9. When tried to copy a virus file, it notifies the user of virus and does not allow to copy:



10. When tried to cat a virus file which existed before running the virus program and renamed virus file:



11. When tried to scan an executable:



Who Did What

Tasks	Done by
Antivirus Module, Overall design and Brainstorming	All
Netlink Socket: Kernel and User communication.	Pratik
User program for on demand access	Vishwajit
Check whether the file contains blacklist content	Nehil
Hijacking of 4 different system calls : open, openAt, execve,	Tushar
exceveAt	
Check whether the file contains whitelist content	Vishwajit and Pratik
Send Notify to show a bubble to the user informing about	Nehil and Tushar
the virus	
Scripts for installation and environment setup	Vishwajit
File rename and permission handling	Pratik and Tushar
Flowchart creation in Visual Paradigm	Nehil
PPT and Report	All
Testing	All
Check for memory leakage, code commenting	All

Citations

http://lxr.free-electrons.com/

http://man7.org/linux/man-pages/man3/nftw.3p.html

http://man7.org/linux/man-pages/man7/netlink.7.html

http://man7.org/linux/man-pages/man2/open.2.html

https://linux.die.net/man/2/openat

http://man7.org/linux/man-pages/man2/execve.2.html

http://man7.org/linux/man-pages/man2/execveat.2.html

http://man7.org/linux/man-pages/man3/exec.3.html

https://tnichols.org/2015/10/19/Hooking-the-Linux-System-Call-Table/

http://stackoverflow.com/questions/3221170/how-to-turn-a-hex-string-into-an-unsigned-char-array