Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

Analyzing and Simplifying Log Files using Python

Yaser Mowlaiwzadah
Master Degree Student
ECE Department
REVA University
Bangalore, India

P. I. Basarkod
ECE Department
REVA University
Bangalore, India

R. B. Manjula
ECE Department
REVA University
Bangalore, India

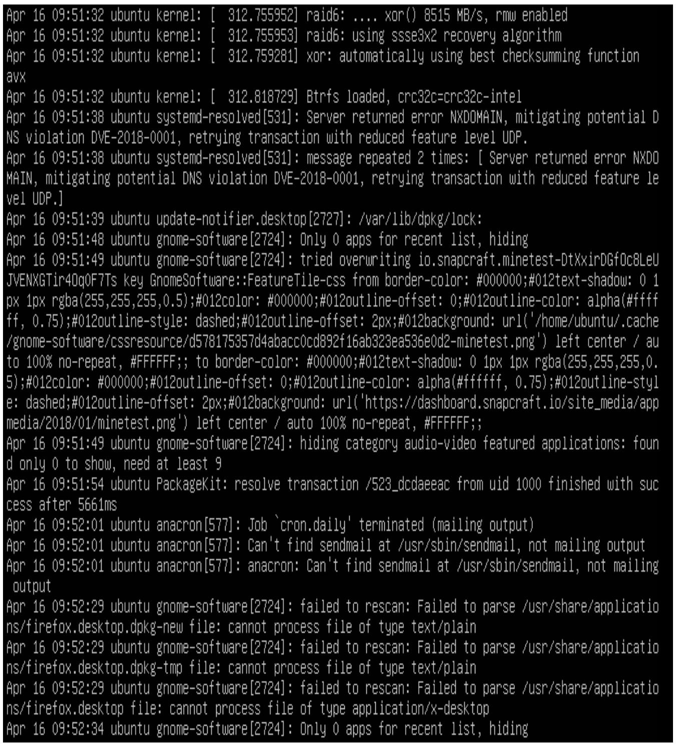
***Abstract*—Nowadays computer security has become an important subject that it discusses about detection and prevention of computer systems from unauthorized access and also human around the world whom have access to internet transmit their sensitive data through internet, all these activities of users during using computer systems and internet are logged into log files which log files have a key role to find information about attacks and unauthorized access to the systems and servers. In today's computer systems, a massive number of various logs is produced, which these logs can be security log or any other type of logs. Analyzing these logs can help an investigator to find useful information about system vulnerabilities and using techniques to prevent them. The purpose of this study is simplifying and analyzing log files by YM Log Analyzer tool, developed by python programming language, it's been more focused on server-based logs (Linux) like apace, Mail, DNS (Domain name System), DHCP (Dynamic Host Configuration Protocol), FTP (File Transfer Protocol), Authentication, Syslog, and History of commands logs. This program has two versions, Script version and Graphic version which the script version is used in servers with no GUI and the graphic version for Desktop user. Using this tool, the administrator is able to find what is happening in systems and realize the importance of log file in systems security.**

Keywords—Logs, Server-based, GUI, Security.

I. INTRODUCTION

First of all, what are logs? Mostly logs are providing a timeline of events for users about operating system, applications, and system and can be very useful for troubleshooting, and these logs are stored on files which are called log files, usually after encountering a problem the first thing an administrator should see through are log files.

Logs located in files are so hard to be comprehended for someone new to Linux Administration and networking, below is a screenshot shows logs located in /var/etc/syslog file which is responsible for storing system logs, what if someone new to the Linux is searching for a keyword among log file or they want logs within a known period of time, even if someone is professional in the field it is good to have some ready program to search for specific keyword or time periods, it can avoid timewasting.



```
Apr 16 09:51:32 ubuntu kernel: [  312.755952] raid6: .... xor() 8515 MB/s, rmw enabled
Apr 16 09:51:32 ubuntu kernel: [  312.755953] raid6: using ssse3x2 recovery algorithm
Apr 16 09:51:32 ubuntu kernel: [  312.759281] xor: automatically using best checksumming function
avx
Apr 16 09:51:32 ubuntu kernel: [  312.818729] Btrfs loaded, crc32c=crc32c-intel
Apr 16 09:51:38 ubuntu systemd-resolved[531]: Server returned error NXDOMAIN, mitigating potential D
NS violation DVE-2018-0001, retrying transaction with reduced feature level UDP.
Apr 16 09:51:38 ubuntu systemd-resolved[531]: message repeated 2 times: [ Server returned error NXDO
MAIN, mitigating potential DNS violation DVE-2018-0001, retrying transaction with reduced feature le
vel UDP.]
Apr 16 09:51:39 ubuntu update-notifier.desktop[2727]: /var/lib/dpkg/lock:
Apr 16 09:51:48 ubuntu gnome-software[2724]: Only 0 apps for recent list, hiding
Apr 16 09:51:49 ubuntu gnome-software[2724]: tried overwriting io.snapcraft.minetest-DtXxirDGfOc8LeU
JVENXGTir4OqOF7Ts key GnomeSoftware::FeatureTile-css from border-color: #000000;#012text-shadow: 0 1
px 1px rgba(255,255,255,0.5);#012color: #000000;#012outline-offset: 0;#012outline-color: alpha(#ffff
ff, 0.75);#012outline-style: dashed;#012outline-offset: 2px;#012background: url('/home/ubuntu/.cache
/gnome-software/cssresource/d578175357d4abacc0cd892f16ab323ea536e0d2-minetest.png') left center / au
to 100% no-repeat, #FFFFFF;; to border-color: #000000;#012text-shadow: 0 1px 1px rgba(255,255,255,0.
5);#012color: #000000;#012outline-offset: 0;#012outline-color: alpha(#ffffff, 0.75);#012outline-styl
e: dashed;#012outline-offset: 2px;#012background: url('https://dashboard.snapcraft.io/site_media/app
media/2018/01/minetest.png') left center / auto 100% no-repeat, #FFFFFF;;
Apr 16 09:51:49 ubuntu gnome-software[2724]: hiding category audio-video featured applications: foun
d only 0 to show, need at least 9
Apr 16 09:51:54 ubuntu PackageKit: resolve transaction /523_dcdaeeac from uid 1000 finished with suc
cess after 5661ms
Apr 16 09:52:01 ubuntu anacron[577]: Job `cron.daily' terminated (mailing output)
Apr 16 09:52:01 ubuntu anacron[577]: Can't find sendmail at /usr/sbin/sendmail, not mailing output
Apr 16 09:52:01 ubuntu anacron[577]: anacron: Can't find sendmail at /usr/sbin/sendmail, not mailing
 output
Apr 16 09:52:29 ubuntu gnome-software[2724]: failed to rescan: Failed to parse /usr/share/applicatio
ns/firefox.desktop.dpkg-new file: cannot process file of type text/plain
Apr 16 09:52:29 ubuntu gnome-software[2724]: failed to rescan: Failed to parse /usr/share/applicatio
ns/firefox.desktop.dpkg-tmp file: cannot process file of type text/plain
Apr 16 09:52:29 ubuntu gnome-software[2724]: failed to rescan: Failed to parse /usr/share/applicatio
ns/firefox.desktop file: cannot process file of type application/x-desktop
Apr 16 09:52:34 ubuntu gnome-software[2724]: Only 0 apps for recent list, hiding
```

Fig. 1 Syslog File Logs

A. Prior Work

As it is said oil is no more the most important and valuable asset it is data that is more valuable than oil, and logs are one type of data that are having a big share of these data, anyone who is having access to this data would be able to do many things with it, so many researches are done on logs in all fields of computer systems not only networks but other parts also, as mentioned before after a system is facing a problem the first thing to be analyzed are logs so the system admin would be able to answer who or what was the problem cause, when it started, why did it happen, all sort of these things, and admin make sure that in future same thing never happens. For all above reasons logs are making a big part in systems troubleshooting, maintenance and security issues, so there is need for research in this field which is widely done already and is being done every day at the present time, these researches on log processing are done on every field and type of subject, some of the prior works that are done and are worthy to be looked upon are:

- Vehicular networks: however vehicular networks are a new technology but log analyzing has already gone through them, analyzing logs on these networks would give good information about the driver, trip, vehicle and road traffic conditions so they are going to be analyzed in real-time and give useful data to avoid any predictable type of problem by monitoring the services and conditions.
- Logs are of wide usage in oil industry also, researches are done and implemented in oil industry which uses log files to provide real time monitoring and reports for better managing of FPSO (Floating Production and Offloading System) which is responsible for production, hydrocarbons processing and oil storing.
- Super computer are another field where log files are used widely, many algorithms and applications are developed to extract and analyze log files on super computers and used them for better management of the systems and also most of the time for providing real time monitoring and reports.
- Logs are also used widely in digital communication where data transaction is in need of high technologies and good management, log are used to provide more optimized and better solutions for data transaction in this technology and are also used for many other tasks that are usual and are used also in other platforms and technologies.

There are also many developed tools that are available for log analyzing and they are advanced tools also, they provide many useful functionalities, can be used on many platforms all of them are having GUI interfaces but too less of them are having non-graphical interface, still they can work with the servers because they are going to connect to the servers and access logs on the servers and provide useful reports some of the tools are:

- Solarwinds Manager: designed for windows only, is a centralized log analyzing tool, data in transaction is encrypted so no unauthorized access, not free
- PRTG: have tools for both windows logs detecting and syslogs, admins can also set alarms for some types of predefined activities, high customizable notification system which also gives the ability for Email or SMS messaging, not free
- Datadog: provides logs analyzing in form of graphs so network performance can be seen in real time also, provides centralized management so all the logs are collected to central storage, not free
- Papetrail: filters can be applied on the output information, free trail only allows up to 100MBs per month,

Beside all above tools there are also many other tools maybe hundred tools that are all doing the same thing and giving the same service, having a look on all these tools and functionalities that are provided by them it would be more clear that how important logs and log files are in today technological world

B. *Why Linux*

As mentioned in abstract we are analyzing logs on Linux system, as answer to why log files are specifically studied on Linux operating system in this paper? Wide usage of Linux operating system should be considered: top 500 supercomputers are ran by Linux, 96.3 percent of the top 1 million web servers are ran by Linux, by 2019 1.99% of personal computers around the globe were running Linux, 90% of cloud infrastructure running Linux, 85% of all smartphones are running Linux, Android is a distribution of Linux in other words, out of 5 smartphones 4 are Linux based, Linux is ran by every major space program such as SpaceX running vehicle Falcon 9, even in Hollywood 90% of special effects are made using Linux, certain countries announced it as national OS, many military establishments prefer using Linux rather than any other operating System and more are migrating to Linux, these are all cases which Linux is being used and many more fields that are not mentioned above, and these are reasons which this paper is focused on Linux.

Logs in Linux are a valuable troubleshooting tool whenever a system admin encounters an issue, everything in Linux has logs for example: system, packet manager, kernel, Apache, boot processes, and etc.

Also as a paper on IEEE named “A comparative study of network based system log management tools” in 2015 which is a comparison between five log analyzing tools only one of them supports Linux, while other four don’t, so need for more log analyzers with different capabilities to support Linux is felt while Linux is leading operating system in many fields and it is being used and growing widely.

Logs in Linux are stored in /var/log directory mostly, log files in the directory can be explored and checked easily by a text editor or any command line command capable of doing so, such as cat, less, tail, head or etc.

C. *Tools Involved*

Tools that are involved and used in this research paper are:

Python: the fundamental tool and language used to develop this tool is python programming language

TkInter: under python the GUI package of TkInter is used to develop the graphical interface for systems with graphical interfaces

OS package: inside the program some Linux system commands are also involved which are based on bash or default shell for Linux so OS package is also imported and used.

D. *Why Python*

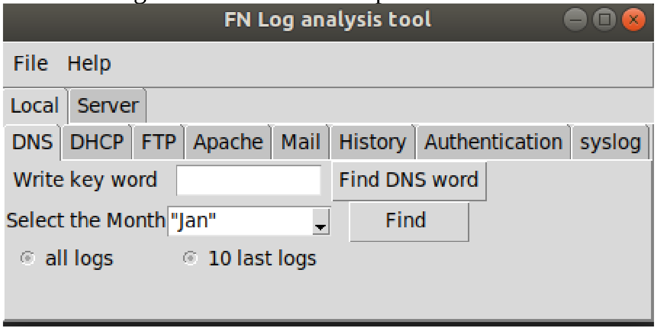
Python is a powerful, high-level, easy to use and general-purpose programming language, python can be used in different fields like: artificial intelligent, data analyzing, web designing, software development, networking software development, and many other fields, and below are some key points that why python is used here:

Automatic compile to byte code, high level data types and operations, wide range of supported extensions, readable code with a distinct C-like quality supports maintenance, large library of contributed applications and tools, object-oriented model, portability across architectures, excellent documentation.

II. METHODOLOGY

The program is only developed for Linux operating system and cannot be used on any other platform, the program is consisted of two separate files which are:

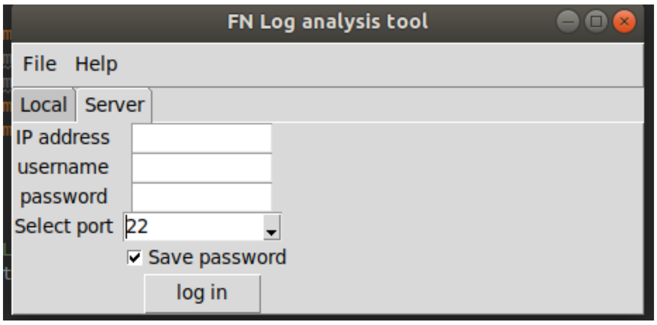
GUI file: this file is the main file for local systems with graphical user interface access. When this file is running the following windows would be opened.



Select the Month "Jan"

Fig. 2 Local Section Interface

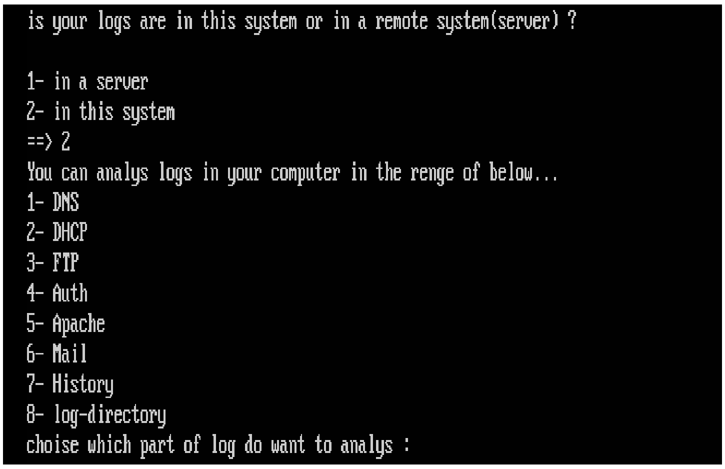
As illustrated the program is consisted of two main parts which are local and server parts in local part logs located on your own local system can be searched and analyzed, while server part is only a predesigned page for future work if anyone is willing to work and develop it a only a primilinary login page is developed.



Select port 22
☑ Save password

Fig. 3 Server Section Interface

Server-based file: this file contains the code for non-GUI environment and server environment where there is no access to GUI interfaces, this file also works same as the before file except that it is not having GUI interface



```
is your logs are in this system or in a remote system(server) ?

1- in a server
2- in this system
==> 2
You can analys logs in your computer in the renge of below...
1- DNS
2- DHCP
3- FTP
4- Auth
5- Apache
6- Mail
7- History
8- log-directory
choise which part of log do want to analys :
```

Fig. 4 Non GUI Interface

There are differences between the GUI and non-GUI version in workflow and results such that in GUI you can search for a specific word or date but in non-GUI it is not possible to search with this much of specificity. Since the program is an open source program so files' directories can be changed inside the program if the user wishes so.

III. RESULT AND DISCUSSION

As result the program that is developed is capable of picking up the requested logs from log files with thousand lines of logs and this is good for avoiding time wasting and simplifying this enormous amount of data that is provided and stored daily on systems, the coded files are available online and can be altered as convinces the user.

IV. CONCLUSION AND FUTURE SCOPE

As for future and further development on the file the server log analyzing and simplifying part on the program is ready to be developed and be provided with more features. Both GUI and script files are uploaded online so that be available for public and anyone wishes to work on them, for downloading the files click the link, after opening the link both files are available for download and they can be recognized by their names.

REFERENCES

[1] Amit , A., & Shyam Tukadiya. (2015). A Comparative Study of Network Based System Log Management Tools. IEEE, 6.

[2] Chen , R., Ji, W., Duan, S., Ling, Q., & Li, F. (2017). A Novel Method to Analyze Logs Generated by Wireless Telecommunication Systems. IEEE, 4.

[3] Hacker, T. (2016). A Markov Random Field Based Approach for Analyzing Supercomputer System Logs. IEEE, 14.

[4] Hongli, W. (2019). A Flow Real-time Data Analyzer for Log of FPSO Central Control System. IEEE, 3.

[5] Jeffrey , S., & Purtilo , J. (2104). Mining Security Vulnerabilities from Linux Distribution Metadata. IEEE, 6.

[6] Mastsumoto, S., Sato, A., Shinjo, Y., Nakai, H., Itano , K., Shomura, Y., & Yoshida, K. (2010). A Method for Analyzing network Traffic Using Cardinality Information in Firewall Logs . IEEE, 6.

[7] Shaout, A., Mysuru, D., & Raghupathy, K. (2018). CAN Sniffing for Vehicle Condition, Driver Behavior Analysis and Data Logging. IEEE, 6.

[8] Stackify. (2017, June 23). What are Linux Logs? How to View Them, Most Important Directories, and More. Retrieved from Stackify : https://stackify.com/linux-logs/

[9] Stackify. (2017, June 23). What are Linux Logs? How to View Them, Most Important Directories, and More. Retrieved from Stackify : https://stackify.com/linux-logs/

[10] Swati , C., Hitendra , C., Tomar, S., & Anil , R. (2014). User and Device Tracking in Private Networks by Correlating Logs: A system for Responsive Forensic Analysis . IEEE, 6.

[11] Team, D. (2019, December 10). Advantages and Disadvantages of Python – How it is dominating programming world. Retrieved from Data-Flair Training: https://data-flair.training/blogs/advantages-and-disadvantages-of-python/

[12] Vaughan-Nichols, S. J. (2015, October 15). Can the Internet exist without Linux? Retrieved from ZDNet: *https://www.zdnet.com/article/can-the-internet-exist-without-linux/*

[13] Keary, T. (2019 , April 26). 11 Best Log Analysis Tools . Retrieved from Comparitech : https://www.comparitech.com/net-admin/best-log-analysis-tools/