A Collaborative Spam Detection Using Obfuscator, Parallel Classifier and Anomaly Detector

Mr. M. Dhamodaran ME[1], B. Indumathi[2], S Karthiga[2], K. Kowsalya[2], S. Samrin Banu[2]

[1]Assistant Professor, [2]B Tech

Department of Information Technology, Muthayammal College of Engineering, Rasipuram, Tamil Nadu, India

ABSTRACT

Spam has become the platform of choice used by cyber-criminals to spread malicious payloads such as viruses and trojans. In this project, we consider the problem of early detection of spam campaigns. Collaborative spam detection techniques can deal with large scale e-mail data contributed by multiple sources; however, they have the well- known problem of requiring disclosure of e-mail content. Spam has become the platform of choice used by cyber criminals to spread malicious payloads such as viruses and trojans. Distance- preserving hashes are one of the common solutions used for preserving the policy of e-mail content while enabling message classification for spam detection. However, distance- preserving hashes are not scalable, thus making large-scale collaborative solutions difficult to implement.

KEYWORDS: Obfuscator, parallel Classifier, Anomaly Detector.

I. INTRODUCTION

The Technology has developing day by day, at same time cyberattacks are also increased. The cyber -attacks caused by Cyber-criminals, they're attacking by various sources, they mainly using Spam. Spam has become the biggest platform and choices used by cyber-criminals. They spread various malicious payloads such as viruses and trojans. So, we consider some solutions to detect the spam campaigns early collaborative spam detection techniques can deal with large scale e-mail content. Spam detection techniques applications are BIG DATA. Big data is a field that treats ways to analyze, systematically extract information. Doug Laney articulated the now-mainstream definition of big data as three V's. Volume, Velocity, and Variety. Big data Applications are Health care, Internet of Things (IoT), Media and Entertainment, Auto driving car, Virtual personal Assistant tool, there are primary three types of Big- data. Structured. Un structured, semi- structured. They had four main components there are Ingestion, Storage, Analysis, Consumption. Then, Distance- preserving hashes are one of the common solutions used for, preserving the policy of e-mail content. However, distance- preserving hashes are not scalable, thus making large- scale collaborative solutions difficult to implement.

II. LITERATURE REVIEW

Recent advances in spam has turned into decision utilized by digital law breakers to spread malignant payloads, for example, infections and Trojans. The issue of early recognition of spam causes has been addressed. The security of email content while allowing message group of spam. A big data set is required for making predictions of the probability of each class. In this contribution, we present the state of the art of the most important similarity preserving hashing function s, analyzing the main features. The validity of hashing in email spam filters. Takes longer time to process. Investigate the issue arising in the design of a digest based spam detection mechanism. Which has to satisfy many conflicting requirements: protect message congeniality, be public, and proved difficult or expensive to fool by obfuscation techniques that automatically introduce difference into the same base spam message. Therefore, it has become necessary to recommend adaptive spam detection models.in this paper, an intelligent system for the detection and filtering of spam e mails is described. In this study, spam detection was carried out using machine learning methods. The classification achieved a success rate of 98.2% in spam detection. In existing system highly parallel encoding technique was used to detect the spam campaigns. Privacy preserving collaborative spam detection is used at the Receiver side. In the proposed system the modern spam filtering techniques are deployed at sender side. To increase the internet bandwidth and storage of the mail server to deploy World Net and Spambot. This thesis presents a method to prepend an efficient gradient boosting machine (GBM). Experimentally demonstrate effectiveness of proposed methods in various application scenarios over our dataset.

III. PROPOSED SYSTEM

E mail system is one of the most effective and commonly used sources of communication. Unfortunately; the E mail system is getting threatened by spam e mails. Email spam, also known as junk email or unsolicited bulk email (UBE), is a category of electronic spam which includes almost identical messages sent by email to multiple recipients. In existing system highly parallel encoding technique was used to detect the spam campaigns. Preserving collaborative spam detection is used at the receiver side. In the proposed system the modern spam filtering techniques are deployed at the sender side. To increase the internet bandwidth and storage of the mail server to deploy World Net and Spambot. The large body of knowledge on spam detection, they not try to provide a complete survey of the domain. The idea of using message digesting or hashing for preserving e-mail confidentiality while filtering spam has been put forward since long. However, cryptographic hashing techniques are easily defeated by simply interesting minor changes into the e mails. In fact, "hash busters", which are random strings that are interested into the are inserted into the end of an e mail, are made specifically for altering the hash value and are used by spammers regularly. Technological developments are making individuals and organizations ever more dependent on e-mail t communicate and share information. Increasing use of e-mail as an important and popular method of communication poses potentially serious threats to the internet and society.
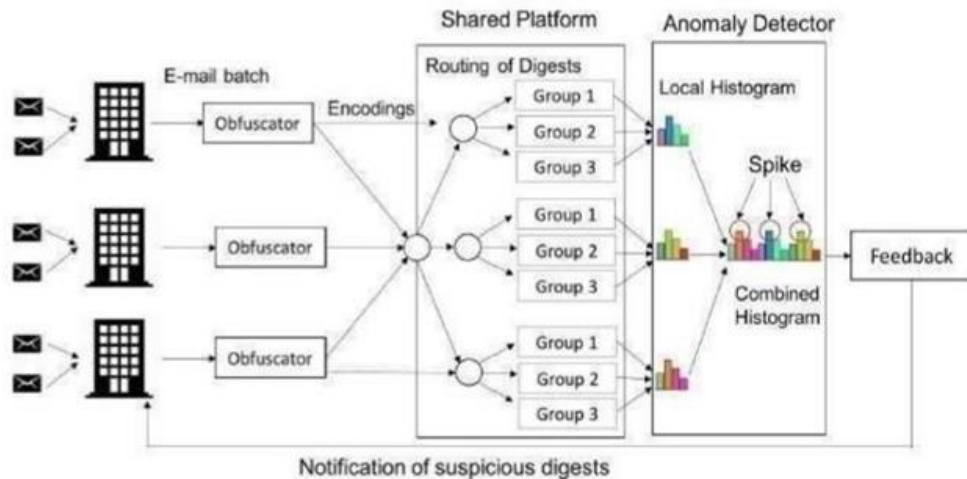
The PHP Hypertext pre processor (PHP) is a programming language that allows web developer to create dynamic content that interacts with databases. PHP is basically used for developing web based software application. This tutorial helps you to build you to build your base with PHP. PHP is a must for students and working professionals to become a great software engineer specially when they are working in web Development Domain. I will list down some of the key advantages of learning PHP. PHP is a server side scripting language that is embedded in HTML. It is used to manage dynamic content, database, session tracking, even build entire e_ commerce sites. It is integrated with a number of popular data base, including MySQL, PostgreSQL, oracle, Sybase, Informix, and Microsoft SQL server.

2.6 UML DIAGRAM
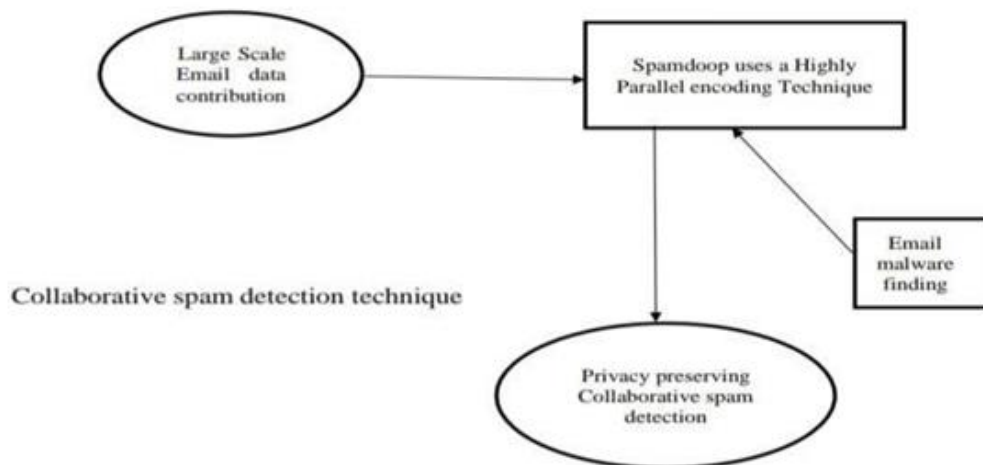
PHP is pleasingly zippy in its execution, especially when compiled as an Apache module on the Unix side. The MySQL server, once started, executes even very complex queries with huge result sets in record-setting time. PHP supports a large number of major protocols such as POP3, IMAP, and LDAP. PHP4 added support for Java and distributed object architectures (COM and CORBA), making n- tier development a possibility for the first time. PHP is forgiving PHP language tries to be as forgiving as possible. PHP syntax is C-Like.

IV. HARDWARE DETAILS

Intel Core2 Duo Processor

Intel core2 duo is the processor family encompassing a range of intel's consumer 64-bit x86-64 single- dual- and quad- core microprocessors based on the core microarchitecture. The intel core 2 duo is an entry level CPU. It can handle basic tasks like web browsing and document editing well. Other than that, it's not very good for anything. For it's time yes The single and dual- core models are single- die, whereas the quad- core models comprise two dies, each containing two cores, packaged in a multi- chip module. The core 2 range was the last flagship range of intel desktop processor to use a front-side bus. The introduction of core relegated the Pentium brand to the mid- range market, and reunified laptop and desktop CPU lines for marketing purpose under the same product name which were formerly divided into Pentium 4, Pentium D, and Pentium M brands. The core 2 processor line was introduced on July 27, 2006, comprising the duo (Dual -core) and Extreme (dual- or quad- core CPU's for Enthusiasts), and in 2007, the Quad (quad- core) and solo (single- core) sub brands. Intel core 2 processors with vPro technology (designed for businesses) include the dual- core and quad- core branches.

Although wood crest processor are also based on the core architecture, they are available under the xenon brand. From December 2006, all core 2 duo processor were manufactured from 300 milli meter plates at Feb 12 factory in Arizona and at Feb 24- 2in country Kildare, Ireland.

The core 2- branded CPUs include: Conroe/ Allendale (dual- core for desktops), Merom (dual- core laptops, Merom- L (single- core for laptops), Kents field (quad- core for desktops), and the updated variants named Wolf Dale (dual-core for desktops), Penryn (dual-core for laptops) and York field (quad-core for desktops) .

The Core 2-branded processors feature Virtualization Technology without EPT (with some exceptions), the NX bit and SSE3. The Core microarchitecture introduced SSSE3, Trusted Execution Technology, Enhanced Speed Step and Active Management Technology (iAMT2). The Penryn microarchitecture, a shrink of the former, introduced SSE4.1. With a maximum thermal design power (TDP) of 65 W, the Core 2 Duo Conroe dissipates half the power of the less capable contemporary Pentium D-branded desktop chips that have a max TDP of 130 W.

V. RESULT

The output of a Collaborative spam detection using obfuscator, parallel classifier and anamoly detector is

Figure 2.7.2.1: User Registration

Figure 2.7.2.2. User Login

Figure2.7.2.3 Uploading Data Files

Figure   2.7.2.4 Notification for Uploading Data Files
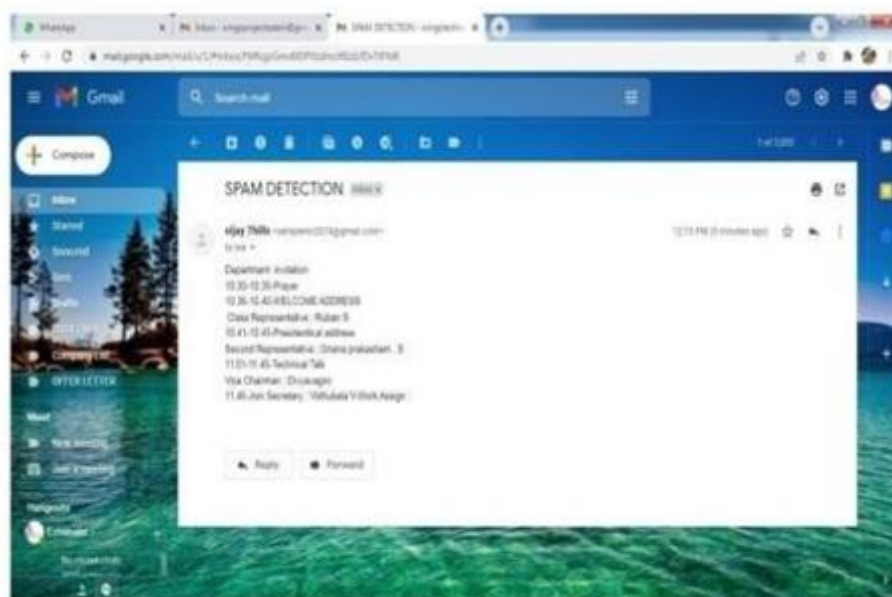

Figure:2.7.2.5   Detection of Spam Message Via Email Notification

VI. CONCLUSION

The collaborative spam detection platform being referred to in this paper offers multiple benefits in term of safeguarding the privacy of all the stakeholders involved in the amount of data being used.

The encoding technique employed is effectively scalable on MapReduce platforms outdoing various distance-preserving hashing techniques.

The techniques used for bucketing simplified the process of offering easy classification and grouping of object along – with anomaly detection.

VII. ACKNOWLEDGEMENTS

VIII. REFERENCES

[1]. Al Mahmoud, Abdelrahman, et al. "Spamdoop A privacy-preserving Big Data platform for collaborative spam detection, "IEEE Transaction on Big data, 2017.

[2]. Kaspersky lab, "spam and phishing statics for 2016". https://www.kaspersky.com/about/press-release/2016kaspersky- labreports -significant-increase-in-malicious-spam-emails-in-ql-2016.

[3]. Chen, long, and Guoyin wang. "An efficient piecewise hashing methods for computer forensics." knowledge Discovery and data mining, 2014. WKDD 2014. First international workshop on. IEEE 2014.

[4]. Wang, Min, et al. "A General framework for Linear Distance Preserving Hashing." IEEE Transaction on image Processing, vol. 27, no. 2, pp. 907-922, 2013.

[5]. J. Francois, S. Wang, W. Bronzi, R. State, and T Engel. "Bot-cloud: detecting botness using MapReduce," in IEEE International Workshop on Information Forensics and Security (WIFS).

[6]. Nagiwale, Amin Nazir, and Manish R. Umale. "Design of self- adjusting algorithm for data-intensive MapReduce application, 2007 International conference on. IEEE 2007.

[7]. Zhong, Zhenyu, Lakshmi Ramaswamy, and Kang Li. "ALPACAS: A large-scale privacy- aware collaborative anti-spam system." INFOCOM 2006.