A Methodology for Identification of the Guilt Agent based on IP Binding with MAC using Bivariate Gaussian Model

B. Raja Koti[1]
Research Scholar
Department of Information Technology,
Gitam Institute of Technology,
GITAM, Visakhapatnam, Andhra Pradesh, India

Dr. Y. Srinivas[3]
Professor
Department of Information Technology,
Gitam Institute of Technology,
GITAM, Visakhapatnam, Andhra Pradesh, India

Dr. G. V. S. Raj Kumar[2]
Associate Professor
Department of Information Technology,
Gitam Institute of Technology,
GITAM, Visakhapatnam, Andhra Pradesh, India

Dr. K. Naveen Kumar[4]
Assistant Professor
Department of Information Technology,
Gitam Institute of Technology,
GITAM, Visakhapatnam, Andhra Pradesh, India

***Abstract*—Enormous increase in data in the current world presents a major threat to the organization. Most of the organization maintains some sort of data that is sensitive and must be protected against the loss and leakage. In the IT field, the large amount of data will be exchanged between the multiple points at every moment. During this allocation of the data from the organization to the third party, there are enormous probabilities of data loss, leakages or alteration. Mostly an email is being utilized for correspondence in the working environment and from web-based like logins to ledgers; thereby an email is turning into a standard business application. An email can be abused to leave organization's elusive information open to trade off. Along these lines, it might be of little surprise that muggings on messages are normal and these issues need to be addressed. This paper completely focuses on the concept of data leakage, the technique to detect the data leakage and prevention.**

Keywords*—Data leakage; sensitive information; data leakage detection; bivariate normal distribution; probability density function

I. Introduction

In the current digital era, the usage of internet has increased rapidly, and every office and organizations were connected to the internet to simplify the works like saving files and sharing information from one to another in order to execute works very fast. In this process, of exchange of data due to the large storage systems, huge data collection is accumulated which is being stored in the server for getting access to a user and reuse the data. But in the large organizations, the data will be handled by the third party members, and sometimes it leads towards certain issues pertaining to data modification, either by technical men within the organization or unknowingly by the users and at times intentionally done by the few. So, in the current scenario, if the data is shared by an unauthorized user, it may create threat/ problem to the organization especially if it is a sensitive data and thereby incurs a huge loss to the organization. To overcome this type of issues, we need to identify the cause of sharing information, whether it was done purposefully or unintentionally, and also we need to provide security to data. If the data is already shared, then we have to protect the data which is already leaked. Protection of data is the very essential thing for any enterprise related to digital data, several methods are adapted to protect the data, based on the priority of information on the demand.

Among the recent challenges, network information expulsion is a vital concern for the business organizations. Unauthorized communication could have serious consequences for a corporation. To discontinue from the unwanted dealings of an organization, it is required to regulate the information flow within and outdoors of the organization and this process helps to secure the data. Recent news and reports indicate fifty percent of organizational data's are leaked within the sector either part or absolutely [1]. This can be terribly troublesome to spot the precise details of leaked information and therefore the informant. However, the information discharge has several channels to leak. Thus observing each channel is a challenging possible task, and also creates several serious problems. However, the principles will be profaned from completely different accessible channels like email, instant electronic communication, and via alternative social media attachments.

Threats to an e-Mail are that employees put organization's sensitive information and resources in danger in spite of arrangements that characterize adjust methodology. The accompanying illustrations indicate how representatives deliberately and unexpectedly spill delicate information. The some of the points referred to are (1) Usage of Unauthorized applications: In organizations, utilization of individual messages can put delicate information and individual data in danger. As indicated by a study report, 63 percent of workers concede that they utilize work PC for individual utilization. These applications don't take after corporate security gauges.

Thus, information spillage by a representative is high [14]. (2) Misuse of Corporate Computers: Employees purposefully utilize organization's PCs in numerous ways that undermine IT security arrangements which incorporates sharing of work gadgets and delicate data with non-workers. These practices can bring about spilling out the IP of an organization, which presents genuine dangers to an organization's security and productivity. (3) Misuse of Passwords and Login/Logout Procedures: When a worker leaves a framework logged in and with a secret phrase joined to it that welcomes an attacker to take the delicate information at their relaxation. On the off chance that worker utilized that PC for individual utilize which implies data is presently eagerly accessible to the attacker [15]. In this article, a model is presented to identify the appropriateness of the user and an error is generated as an unauthorized user if he/she is not authenticated.

The rest of the paper is organized as follows; section 2 of the paper deals with the current scenarios, Data Leakages and Guilt Agents identification. Section 3 of the paper highlights about bivariate normal distribution, KL- Divergence and the proposed algorithm deals with the methodology developed in section 3. Results derived are presented in section 4. In section 5 the conclusions are summarized.

II. Related Work

Data Leakage is one of the necessary assets to several organizations having information, and for that matter, the protection of this information should take the primary priority [11]. Although several organizations have placed unavoidable security mechanisms and technical systems significantly firewalls, virtual personal networks (VPNs), and intrusion detection systems/intrusion bar systems (IDSs/IPSs; still information expulsion will occur [2]. Say again that if the information leak happens and once sensitive data is unconcealed to unauthorized users or parties either purposely or not. The information leakage will cause serious consequences or several threats to an organization. Let's say, the loss of the confidential or sensitive information will have an unembellished or confrontational impact on a company's name and quality, customers, worker confidence, competitive advantage, and in some cases, this will result in the closure of the organization. In additionally, information leak is a key concern for the business organizations during this progressively networked world today and for that matter, any unauthorized speech act of sensitive or confidential knowledge might have serious consequences for a corporation in each long and short-term [3].

In addition, consistent with the problem of information, a leak could be a growing concern among organizations and agents. Alneyadi et al [4] indicated that a lot of leakages occurred within the business sectors that they were within the government sector. Consistent with a report in 2014, the statistics stand at 500th within the sector and 200th within the government sector. They have declared that though in some cases the information leaks weren't harmful to organizations, however, others have caused many lots of dollars' value harm. More so, the quality of many businesses or organizations area unit comprised one such sensitive information as trade secrets, project documents, and client profiles area unit leaked to their competitors. Take it more than government sensitive information values political selections, enforcement, and national security data may also be leaked. A typical example of presidency sensitive information that was leaked was the United States diplomatic cables by Wiki Leaks. "The leak consisted of concerning 250,000 United States diplomatic cables and 400,000 military reports observed as 'war logs'. This revelation was administered by an enclosed entity exploitation an external disk drive and regarding 100,000 diplomatic cables were labeled confidential and 15,000 cables were classified as secret", this incident received high public criticisms from among civil rights organizations everywhere the global. In another development, hackers scarf 160 million credit and open-end credit numbers that targeted 800,000 bank accounts in the United States, that were thought of joined of the most important hacking incident that has occurred [5].

The need to deal with such serious problems culminated with the implementation of bound security management mechanisms similar to firewalls, VPNs, IDS, and IPSs by many organizations [6]. Consistent with these systems work satisfactorily once the information is well outlined, structured and constant. Alneyadi et al [7] Any such explicit attack; that once information is either changed, tag otherwise or compressed, these systems become naive and confidential information will still be leaked. As an example, a firewall will have rules to dam access to confidential information; however, constant information may be accessed through many means that compared to an email attachment and instant electronic messaging (IM). This suggests that the normal security mechanisms (firewall, VPNs, IDSs / IPSs) is incapacitated and lack the understanding of information linguistics [8]. To beat this deficiency in protecting sensitive information, a brand new paradigm shift known as information leak interference systems are introduced. Security and privacy problems have enhanced by the rate, volume, and style of bachelor's degree, comparable to large-scale cloud infrastructures, diversity of information sources and formats, streaming nature of information acquisition, and large amount information inter-cloud migration [9]. Maybe sensitive or non-sensitive, and irrespective of a leak of information may be expensive for any businesses or users. Parenthetically, a client MasterCard record that is leaked may be expensive to reach the bank and therefore the client. [10] Typically, information leak happens because of data sharing with users internally or outwardly to the organization, exchanging emails that contain sensitive data, publically cathartic information on the web or cloud, data that is taken with illicit motives or inadvertently. The guilt can be identified by using the MCA-IP address [12] bound to the particular log file and the timestamps allotted by the records and the leaked data is encrypted [13]. Information sensitivity varies comparable to banking data, MasterCard data, criminal and justice information, financial data, health records, etc., to feature to the current; the appearance of sensitive data has caused various information security challenges that need completely different mechanisms in managing things. Also, because of the voluminous of information that is generated and used recently by organizations, there should be subtle technologies and methodologies that may handle the voluminous of information firmly and with efficiency and to stop data leak. Finally, many DLP ways are designed, however,

there is very little done stop information leak in Sensitive data exploitation; the preventive approach which might facilitate organizations prevent the leak before they happen.

III. METHODOLOGY

Data leakages are considered to be one of the most targeting problems for each and every organization. With the latest technology evolution, most of the industries are migrating towards the development of a framework that suits their organization. This professional difference towards the latest technology changes has given to fold both job seeker and job provider. The number of technical jobs in this area research has been increased enormously and thereby creating a lot of ventures to the software industry and also given a scope for the budding entrepreneurs to start their own entrepreneurs by mean of establishing start-up companies. On the other hand, due to professional opportunities available, a lot of challenges also arose because of the security constraints. The security issues may be either with respect to the attacks on the organization with the only purpose of hacking the valid information is to surplus the information available at an organization with the objective of professional rivalry. These attacks are turned as outside attacks were apart from hacking and attacking the information other means of attacks such as spreading malicious virus, worms etc., with the purpose of destroying the meaningful information other types of attacks are also victims in the current day challenges and these types of attacks are called as the internal attacks were the objective is to leak the content from the source organization working with and supply the information to the third party for financial and other sort of benefits. Many articles are being proposed and developed in the literature to identify the guilt agent. However, this issue remains to be a still in challenge state and leads directions to make the work progress. Hence we propose a Bivariate Gaussian distribution model and KL divergence which helps to identify the leakage more appropriate.

A. Bivariate Normal Density

The bivariate normal distribution can be defined as the probability density functions (pdf) of two variables X and Y that is linear functions of the same independent normal random variables of the function is

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}Q(x,y)\right) \tag{1}$$

with the quadratic form

$$Q(x,y)=\frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\frac{(x-\mu_x)(x-\mu_y)}{\sigma_x\sigma_y}\right] \tag{2}$$

Given the joint density function of a bivariate normal distribution, with the parameters σ_x^2, σ_y^2 and ρ, the 2-by-2 symmetric matrix is given by

$$\sum=\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix},\quad x=\begin{bmatrix}x\\ y\end{bmatrix}\quad\text{and}\quad \mu=\begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix} \tag{3}$$

The quadratic form can be expressed as

$$Q(x,y)=(x-\mu)^T\sum(x-\mu)=\begin{bmatrix}x-\mu_x & y-\mu_y\end{bmatrix}\begin{bmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}x-\mu_x\\ y-\mu_y\end{bmatrix} \tag{4}$$

Suppose that two random variables X and Y has the bivariate normal distribution (1). The following properties are to find out the bivariate normal distribution.

1) Their marginal distributions

$f_X(x)$ and $f_Y(x)$ become

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma_x}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\} \tag{5}$$

and

$$f_Y(y)=\frac{1}{\sqrt{2\pi\sigma_y}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\} \tag{6}$$

Thus, X and Y are normally distributed with respective parameters (μ_x,σ_x^2) and (μ_y,σ_y^2).

2) If $\rho=0$, then the quadratic form (2) becomes and consequently, we have $f(x,y)=f_X(x)f_Y(y)$. Thus, X and Y are independent when the $\rho=0$.

$$Q(x,y)=\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{x-\mu_y}{\sigma_y}\right)^2 \tag{7}$$

3) If *X* and *Y* are not independent (that is, $\rho \neq 0$), we can compute the conditional density function $f_{Y|X}(y|x)$ given $X=x$ as

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi\sigma_y}\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right\} \tag{8}$$

Which is the normal density function with a parameter $(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(y-\mu_x),\ \sigma_y^2(1-\rho^2))$ similarly, the conditional density function $f_{X|Y}(y|x)$ is given $Y=y$ becomes this is the normal density function with parameter $(\mu_x+\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y),\sigma_x^2(1-\rho^2))$

$$f_{X|Y}(x|y)=\frac{1}{\sqrt{2\pi\sigma_x}\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)}{\sigma_x\sqrt{1-\rho^2}}\right)^2\right\} \tag{9}$$

B. Covariance

Let (X,Y) be a bivariate normal random variables with a parameters $(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2\rho)$. The covariance is given by

$$f_{X|Y}(x|y)=\frac{1}{\sqrt{2\pi\sigma_x}\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)}{\sigma_x\sqrt{1-\rho^2}}\right)^2\right\}\sigma_y^2(1-\rho^2) \quad (10)$$

The correlation coefficient of X and Y is defined by

$$\rho=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \quad (11)$$

It implies that the parameter ρ of the bivariate normal distribution represents the correlation coefficient of X and Y.

C. KL-Divergence

Kullback - Leibler Divergence (KL Divergence) is considered to identify the divergence between two probability density functions and the formula for computing the same is given by

$$=\int_{-\infty}^{\infty}(x-\mu_X)^2 f(x)dx=\rho\frac{1}{2} \quad (12)$$

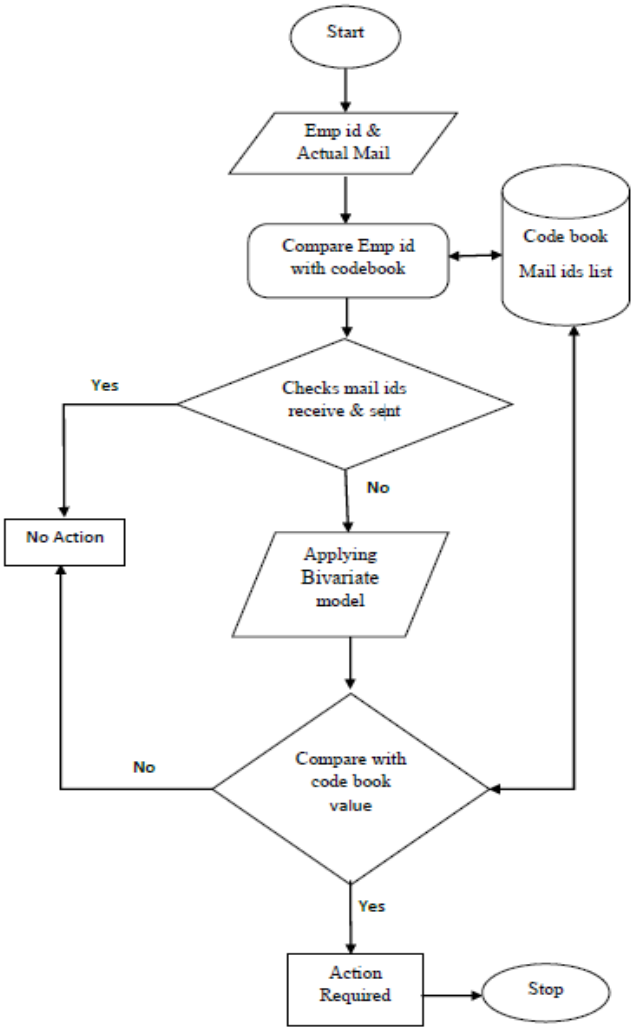


Fig. 1. Data Flow Diagram.

D. General Architecture

In this paper, we have considered a Bivariate Gaussian model, with the assumption that as it considers two attributes into consideration, the identification of the guilt agent becomes easier. In this method, we have considered the unique id of an employee within the organization, together with the MAC address of the system being used and the corresponding probability density function are mapped which is shown in fig 1. Every data that is transmitted from the source are identified as the super user or administrator and against each of the probability density function, the corresponding MAC id and system IP are notified. During the transmission process whenever a data is transferred from source to destination, three parameters are considered and checked against the e-mail ids. Every employee performing for an organization is entitled to transmit data are receiving data from a set of valid users. Every e-mail i.e. received or transmitted is cross-checked against the validity of the e-mails to which they are sent or receive in order to identify the guilt agent.

E. Proposed Algorithm

Algorithm: Checking Codebook Algorithm
Input: T1, T2 (codebook), Dept, and Mail id
//T2**:** Original codebook data
//Dept**:** department of an organization
Output: L, display miss match data of T1 and T2

1. **begin**
2. L = 0;
3. **If** match (Mail. To, organization) AND match (Mail. Dept) **then**
4. Return L;
5. **end if**
6. **If** match ((Mail. To, organization) AND not match (Mail, Dept)) OR not match (Mail. To, organization emp id) **then**
7. **If** match (receive, mail) AND match (send, mail) then
8. Return L;
9. **end if**
10. **end if**
11. **If** not match (receive, mail) not AND match (send, mail) then
12. **for all** t1 ∋ Mail **do**
13. **for all** t2 ∋ T2**do**
14. **If** match (t1, t2) **then**
15. L = L compare (t1 ↔ t2);
16. **end if**
17. **end for**
18. **If** L ≠ 0 **then**
19. Apply Action ();
20. **end if**
21. Return L;
22. **end for**
23. **end if**
24. **end**

IV. Experimental Results

To evaluate the performance of the proposed model that considers the sensitive data leakage problem. We implemented the proposed framework by using Python environment. In our experimental the developed model is presented in two phases, in the first phase we have considered a model that is developed by considering the MAC and along with employ id and output derived are given as the input to the model based on Bivariate Gaussian mixer model in the second phase instead of MAC address we assumed that IP is static and linked to the employ id to get a unique value based on the consider model. The two scenarios have experimented with different typed of employ ids and their MAC addresses, IP addresses.

For this implementation we use the data set which has unique employ ids, for each employ id we have created a codebook which contains emails ids list for that particular employee has to receive and sent, this experiment was done with 150 emails ids which are unique, from those values we generated mean, variance, standard deviation, correlation and to calculate Bivariate values all those sample values are given in below from table I to table III.

TABLE I. Mean, Variance and Standard Deviation for the Dataset

	EMP ID	MAC	IP
Count	150	150	150
Mean	75.5	649.2267	849.88
Variance	1887.5	1578.499	1483.999
Standard Deviation	43.44537	39.73032	38.52271

TABLE II. Correlation and Covariance for EMP, MAC and IP

	EMP, MAC	EMP, IP	MAC, IP
Correlation	0.565266	0.571449	0.3615
Covariance	975.7047	956.396	553.2824

TABLE III. Bivariate Gaussian Model Values for EMP and MAC, EMP and IP

EMP IDs	Bivariate Gaussian model for EMP and MAC	Bivariate Gaussian model for EMP and IP
5	0.000299749	0.000501601
10	0.000511379	0.000194190
19	0.000628811	0.000644221
27	0.000729978	0.000177475

From the fig 2 Bivariate Gaussian model values which are one employ id, for that, we generated values for the taken data set to substitute in the equation 1. Here we consider x and y values are for the first phase employ id is x and MAC is y, in the same way in the second phase we consider employ id is x and IP is y. After getting two phases pdf values are stored in the codebook which is on the server side to compare with the test data.

We performed our experiments in 2 cases; based on IP and EMP id, MAC and EMP id categories. The results in Table IV show that the accuracy of MAC and EMP id category is the best case when compared to the other.

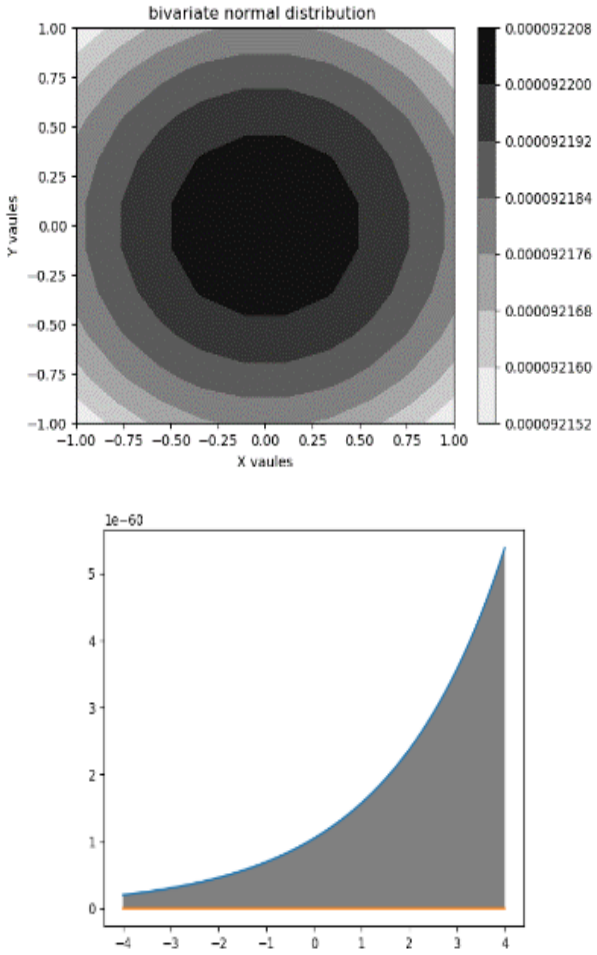


Fig. 2. Bivariate Gaussian Model Values.

In order to evaluate the procedure, we have considered the evaluation metrics based on Precision and Recall and based on these values the accuracy measure is evaluated, and the formula for estimating the accuracy is given by

$$\text{Accuracy} = (TP + TN)/N \quad (13)$$

$$\text{Precision} = TP/(TP+FP) \quad (14)$$

$$\text{Recall} = TP/(TP+FN) \quad (15)$$

TABLE IV. Precision , Recall and Accuracy

	MAC and EMP ID	IP and EMP ID
TP	120	110
TN	245	255
FP	30	40
FN	10	15
Precision	0.80	0.73
Recall	0.92	0.88
Accuracy	0.90	0.87

where N = TP + FP + TN + FN, and where TP is True positive, FP False positive, TN True negative, FN False negative.

In order to identify the output, KL Divergence is considered, KL Divergence is used to find relevance in between two generated probability density function values. Here the KL Divergence is to compare and relevance of the training set and test data, if the divergent between the test and train is more than this, says there is a diversion in the data and they may be a possible scope for data leakage.

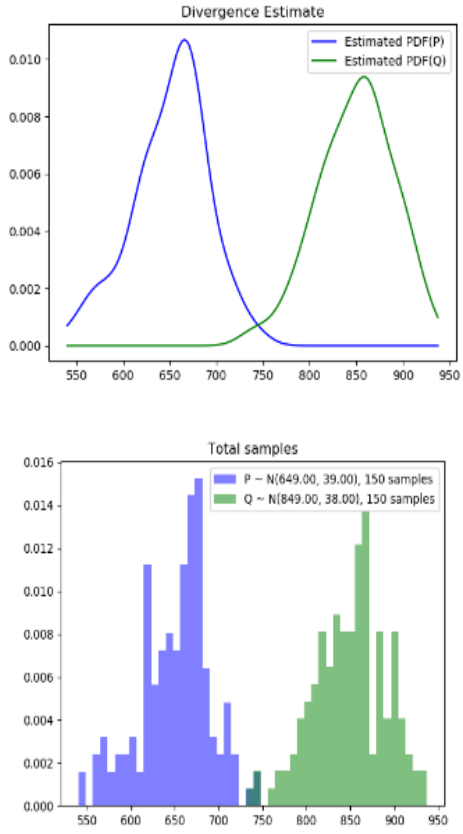
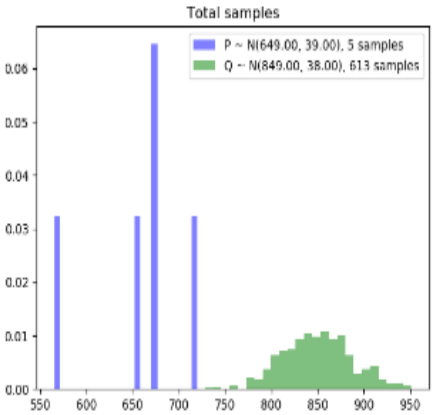


Fig. 3. Total Samples that we are taken for P, Q and Estimated PDF's for P and Q.



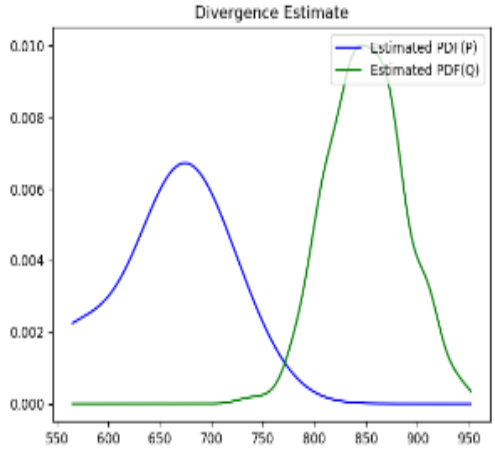


Fig. 4. PSamples that we are taken for test with mean 649.00 and Estimated PDF's for P and Q.

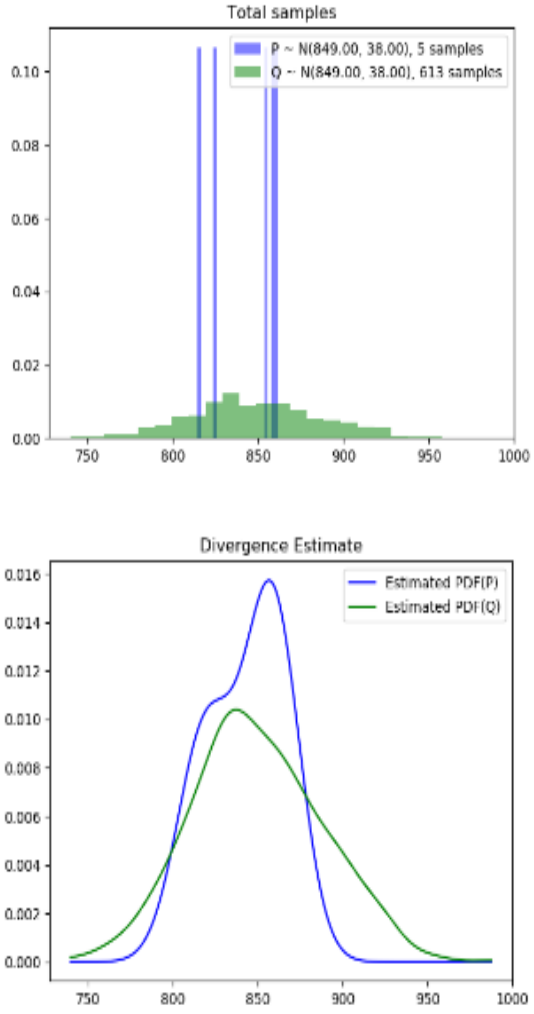


Fig. 5. PSamples that we are taken for test with mean 849.00 and Estimated PDF's for P and Q.

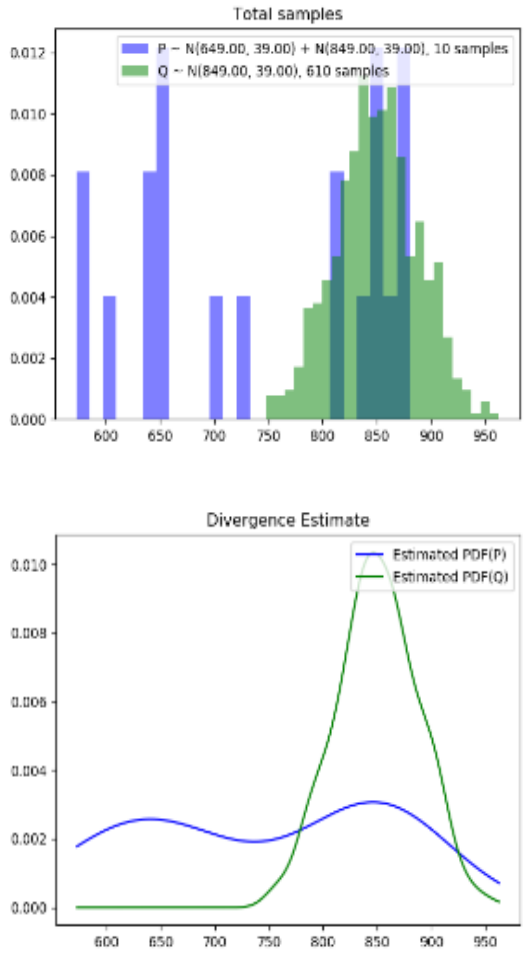


Fig. 6. KL Divergence graphs for the P, Q and Estimated PDF's for P and Q.

The above fig -3 to fig - 6 the pdf values of the two data sets, viz trained data pdf values that stored in server side and the original values of the acquired are considered for each employ id separately and then another figure denotes the test data pdf values which we have collected after the employ communicated with mail id. Then the KL Divergence is used to find relevance in between two generated probability density functions values and find out odd one out in the organization.

V. Conclusion

Data leakage is a current drawback within the field of IT mostly in the data security. There are various researches from numerous domains who are endlessly working towards developing and safeguarding the sensitive information from the data leak detection and finding ways to reduce this drawback. Protective confidential and sensitive data is a lot of vital importance. The aim was to develop the model with a known the scope of leakage and apply a technique to detect the data leakage and to identify the guilty agent(s) in a particular organization. We can't predict an organization that does not utilize an email for correspondence and similarly it is hard to consider an organization that would not get any advantage from an email security. So, we proposed a Bivariate Gaussian distribution model that helps to identify the guilty agent and leakage more appropriately. Thus, it provides the necessary security to the sensitive information and finding the guilt agent also, which is very helpful in various areas that hold sensitive data, especially where data is shared through emails. The proposed system can disconnect inbound or outbound emails. Future endeavors can be made in actualizing this technique that can handle the present reality situation thereby helping to identify the guilt agents and also protect the leaked data with more accuracy.

References

[1] Data loss db. Data loss statistics. Retrieved from ⟨http://datalossdb.org/⟩; 2015.

[2] Tahboub, R & Saleh, Y. (2014), Data Leakage / Loss Prevention Systems (DLP), NNGT Journal: International Journal of Information Systems, vol. 1, pp. 13-18.

[3] Soumya, S. R. &Smitha, E. S. (2014), Data Leakage Prevention System by Context based Keyword Matching and Encrypted Data Detection, International Journal of Advanced Research in Computer Science Engineering and Information Technology, vol. 3, issue 1, pp. 375-384.

[4] Alneyadi, S., Sithirasenan, E. &Muthukkumarasamy, V. (2016), A survey on data leakage prevention systems, Journal of Network and Computer Applications, vol. 62, issue C, pp. 137-152.

[5] Vadsola, R., Desai, D., Brahmbhatt, M. &Patanwadia, A. (2014), Data Leakage Prevention by Using Word Gram Based Classification and Clustering, International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, issue 9, pp. 8040-8041.

[6] Kale, A. V., Bajpayee, V. & Dubey, S. P. (2015), Analysis of Data Leakage Prevention Solutions, International Journal for Engineering Applications and Technology (IJFEAT), vol. 1, issue, 12, pp. 54-57.

[7] Alneyadi, S., Sithirasenan, E. and Muthukkumarasamy, V. (2015), Detecting Data Semantic: A Data Leakage Prevention Approach, In the Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA, August 20 - 22, IEEE Computer Society Washington DC, USA, vol. 1, pp. 910-917.

[8] Alneyadi S., Sithirasenan E. &Muthukkumarasamy V. (2014), A Semantics-Aware Classification Approach for Data Leakage Prevention, In: Susilo W., Mu Y. (eds) Information Security and Privacy, ACISP 2014, Lecture Notes in Computer Science, vol. 8544, pp.413-421, Springer, Cham.

[9] Shirudkar, K. &Motwani, D. (2015), Big-Data Security. International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, issue 3, pp. 1100-1109.

[10] Tidke, P., Wagh, A., Bharade, D. &Dongre, A. G. (2015), Data Leakage Prevention with E-Mail Filtering, International Journal of Advance Foundation and Research in Computer (IJAFRC), vol. 2, issue 2, pp. 28-32.

[11] B. Raja Koti, Dr. G.V.S. Raj Kumar, and Dr. Y. Srinivas, "A Comprehensive Study and Comparison of Various Methods On Data Leakages", International Journal of Advanced Research in Computer Science, Volume 8, No.7, July – August 2017, pp-627-631, ISSN No: 0976-5697

[12] B Raja Koti, GVS Raj Kumar, Y Srinivas, "Identification of Guilt Agent and Leaked Data by Using MAC-IP", International Journal of Applied Engineering Research, 2017, Volume 12, Issue 22, pp 12237-12245.

[13] B Raja Koti, G V S Raj Kumar, "Information leakage detection and protection of leaked information by using the MAC-IP binding technique", International Journal of Engineering &Technology, Volume 7, Issue 1.7, 2018, pp230-235.

[14] P. Zilberman, S. Dolev, G. Katz, Y. Elovici and A. Shabtai, "Analyzing group communication for preventing data leakage via email," Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, Beijing, 2011, pp. 37-41.

[15] R. Tahboub and Y. Saleh, "Data Leakage/Loss Prevention Systems (DLP)," 2014 World Congress on Computer Applications and Information Systems (WCCAIS), Hammamet, 2014, pp. 1-6.