JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SECURITY ASSESSMENT FOR AN APPLICATION USING A SPAM FILTER AND A PATTEREN CLASSIFIER

[1]Dr. Soumya Patil., [2]Kavya N., [3]Aysiri Chidananda, [4]Kavita Basappa Melmari.,

[1]Associate Professor, [2]Student, [3]Student, [4]Student
[1]Computer Science and Engineering,
[1]Sir M Visvesvaraya Institute of Technology, Bangalore, India

Abstract: In hostile applications like biometric authentication, network intrusion detection, and spam filtering, where data may be deliberately altered by humans to impede function, pattern classification algorithms are frequently used. Because traditional design methods do not account for this adversarial scenario, pattern categorization systems may contain flaws that could be exploited to significantly lower their performance and, as a result, their usefulness in real- world applications. As a result, extending pattern classification theory and design methods to adversarial situations is a novel and highly pertinent research subject that hasn't been thoroughly explored before.

***Keywords*:** Pattern Classifiers, Data Mining, Spam Filter, Data Sets

I. INTRODUCTION

In this work we address issues above by developing a framework for the empirical evaluation of classifier security at design phase that extends the model selection and performance evaluation steps of the classical design cycle. We summarize previous work, and point out three main ideas that emerge from it. We then formalize and generalize them in our framework.

First, to pursue security in the context of an arms race it is not sufficient to react to observed attacks, but it is also necessary to proactively anticipate the adversary by predicting the most relevant, potential attacks through a what-if analysis; this allows one to develop suitable countermeasures before the attack actually occurs, according to the principle of security by design.

Second, to provide practical guidelines for simulating realistic attack scenarios, we define a general model of the adversary, in terms of her goal, knowledge and capability, which encompass and generalize models proposed in previous work.

Third, since the presence of carefully targeted attacks may affect the distribution of training and testing data separately, we propose a model of the data distribution that can formally characterize this behavior, and that allows us to take into account a large number of potential attacks; we also propose an algorithm for the generation of training and testing sets to be used for security evaluation, which can naturally accommodate application-specific and heuristic techniques for simulating attacks.

We make use of the concept of data mining to achieve these ideas. Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categories it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyses relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks.

II. RELATED WORK

Ribeiro, P.B., et al.,2015 [1], The authors of this paperemploy data mining methods to precisely pinpoint spamdetection anomalies. In this case, Weka is used to compare the effectiveness of several machine learning methods. The SPAMBASE dataset, which is a collection of emails dividedinto spam and non-spam categories, is used in this. Theresults of the experiment have been evaluated using theROC theory. The findings of this study demonstrated that tactics that emphasized trees and forests were more successful than others.

Gupta P, et al.,2019 [2], The writers of this essay want to address the issue of spam in emails and SMS. The dataset utilized, which is from the UCI Machine Learning library, represents SMS spams. The performance of the Naive Bayes(NB) and Support Vector Machine (SVM) algorithms isassessed after pre-processing the dataset. The outcomes demonstrated that the Naive Bayes method outperformedthe SVM strategy significantly.

Chaudhari N., et al.,2016 [3], This essay's authors talk onthe issue of SMS spam. Problem The strategies and methodsthat can be utilized to address the issue are thoroughly described using spam SMS filtering techniques. Naive Bayes, Bayesian classifiers, and support vector machines(SVM) were discovered to be superior than previous techniques for filtering spam SMS. By combining two or more algorithms, hybrid spam filtering solutions can increase the efficacy and accuracy of traditional spam SMS filtering procedures.

Rahman S.E, et al.,2020 [4], The authors of this articlediscuss the problem of email spam. To recognize spam emails, a novel algorithm based on sentiment analysis of thetextual data in the email body is suggested. Word embedding's and a bidirectional LSTM network are used to examine the emotive and sequential elements of texts. Convolution neural networks are used to quicken training and extract more intricate text features for the Bi-LSTM network. The two datasets that were used are the ling spam dataset and the spam text message categorization dataset. The proposed technique's performance is compared and evaluated using recall, precision, and f-score. The outcomes demonstrated that this model was more accurate than others,with a range of 98 to 99%.

Shrivastava S., et al.,2017 [5], The authors of this essay discuss the problem of spam mail. The goal is to develop a spam filter utilizing Bernoulli's formula and continuous probability distribution. Decision trees and Naive Bayes are the algorithms employed. In connection to overfitting, decision trees' usefulness and accuracy are assessed. How well a classifier model can distinguish between spam and non-spam emails will determine which one is more accurate. The experimental findings demonstrated that Bernoulli's probability distribution outperforms continuous probability distribution in terms of classifier model performance. Furthermore, it was discovered that neitherdecision trees nor naive baye's are unquestionably the best classifier models.

Makkar, A, et al.,2020 [6], In this paper the authors address the security issue of IoT devices by detecting spam using machine learning. To solve the issue five machine learning models are evaluated using various metrics with a largecollection of inputs features sets. Each model computes a spam score by considering the refined input features. This score depicts the trustworthiness of IoT devices under various parameters. The dataset used is REFIT Smart Home.The results obtained prove the effectiveness of the proposed system in comparison to the other existing systems.

Baaqeel H., et al.,2020 [7], The study's authors discuss measures to prevent SMS spam. It is suggested to use a hybrid system that combines supervised and unsupervised machine learning techniques. Both spam filtering accuracy and F-measures accuracy are the objectives of the new hybrid approach. SVM was found to have the highest precision and no false-positive rates out of six different supervised models, while KNN did the least well. The best accuracy of 98.8% is achieved by K-means and SVM fusionfollowing a variety of hybrid model combinations.

Abdullahi M., et al.,2021 [8] The subject of image-based spam email is discussed by the writers in this work. Various machine learning methods pertinent to spam detection were investigated. The contributions made by researchers toward reducing spam using machine learning classifiers are also reviewed. Selected machine learning algorithms such Naive Bayes, Clustering methods, Random Forest, Decision Trees,and Support Vector Machines (SVM) were compared. It wasdiscovered that Naive Bayes is the most effective methodfor F measurements. Another finding was the impact of dataset variations on classifier performance.

Abinaya R, et al.,2020 [9] The authors of this research discuss spam on social media sites. The dataset is made upof comments on YouTube. The outcomes were based onfour entirely distinct machine learning algorithms: logistic regression, decision trees, random forest, ada boost classifier, and support vector machine. An accuracy of 95.40% was discovered to be feasible with Logistic Regression, outperforming the existing solution by around 18%.

Alsaleh M.,et al .,2015 [10] The authors of this article talk about comment spam. The goal is to create a technique for identifying and removing spam comments by analyzing the Document Object Model (DOM) of the target web pageusing a browser plugin. They manually labelled a fresh corpus of blog comments in order to assess the proposed classifier's accuracy. It was shown that classifiers that included all or parts of the target detection attributes, suchas neural networks, random forests, decision trees, and support vector machines, produced the best results. Using a subset of the tested features, a decision tree classifier serves as the foundation for the comment spam detection program.

III. METHODOLOGY

The system design mainly consists of:

A. Attack Scenario and Model of the Adversary

B. Pattern Classification

C. Adversarial classification

D. Security modules

A. Attack Scenario and Model of the Adversary

Although the definition of attack scenarios is ultimately an application-specific issue, it is possible to give general guidelines that can help the designer of a pattern recognitionsystem. Here we propose to specify the attack scenario in terms of a conceptual model of the adversary that encompasses, unifies, and extends different ideas from previous work. Our model is based on the assumption that the adversary acts rationally to attain a given goal,according to her knowledge of the classifier, and her capability of manipulating data. This allows one to derivethe corresponding optimal attack strategy.

B. Pattern Classification

Multimodal biometric systems for personal identity recognition have received great interest in the past few years. It has been shown that combining information coming from different biometric traits can overcome thelimits and the weaknesses inherent in every individual biometric, resulting in a higher accuracy. Moreover, it is commonly believed that multimodal systems also improve security against Spoofing attacks, which consist of claiming a false identity and submitting at least one fake biometric trait to the system (e.g., a “gummy” fingerprint or a photograph of a user’s face). The reason is that, to evade multimodal system, one expects that the adversary should spoof all the corresponding biometric traits. In this application example, we show how the designer of a multimodal system can verify if this hypothesis holds, before deploying the system, by simulating spoofing attacks against each of the matchers.

C. Adversarial classification:

Assume that a classifier has to discriminate between legitimate and spam emails on the basis of their textualcontent, and that the bag-of-words feature representation hasbeen chosen, with binary features denoting the occurrenceof a given set of words. Essentially all data mining algorithms assume that the data-generating process isindependent of the data miner's activities. However, in many domains, including spam detection, intrusion detection, fraud detection, surveillance and counter-terrorism, this isfar from the case: the data is actively manipulated by an adversary seeking to make the classifier produce false negatives. In these domains, the performance of a classifier can degrade rapidly after it is deployed, as the adversary learns to defeat it. Currently the only solution to this isrepeated, manual, ad hoc reconstruction of the classifier.

D. Security modules:

Intrusion detection systems analyze network traffic to prevent and detect malicious activities like intrusion attempts, ROC curves of the considered multimodal biometric system under a simulated spoof attack against the fingerprint or the face matcher. Port scans, and denial-of- service attacks. When suspected malicious traffic is detected, an alarm is raised by the IDS and subsequently handled by the system administrator. Two main kinds of IDSs exist: misuse detectors and anomaly-based ones. Misuse detectors match the analyzed network traffic against a database of signatures of known malicious activities. The main drawback is that they are not able to detect never- before-seen malicious activities, or even variants of known ones. To overcome this issue, anomaly-based detectors have been proposed. They build a statistical model of the normal traffic using machine learning techniques, usually one-class classifiers, and raise an alarm when anomalous traffic is detected. Their training set is constructed, and periodically updated to follow the changes of normal traffic, bycollecting unsupervised network traffic during operation,assuming that it is normal (it can be filtered by a misuse detector, and should)

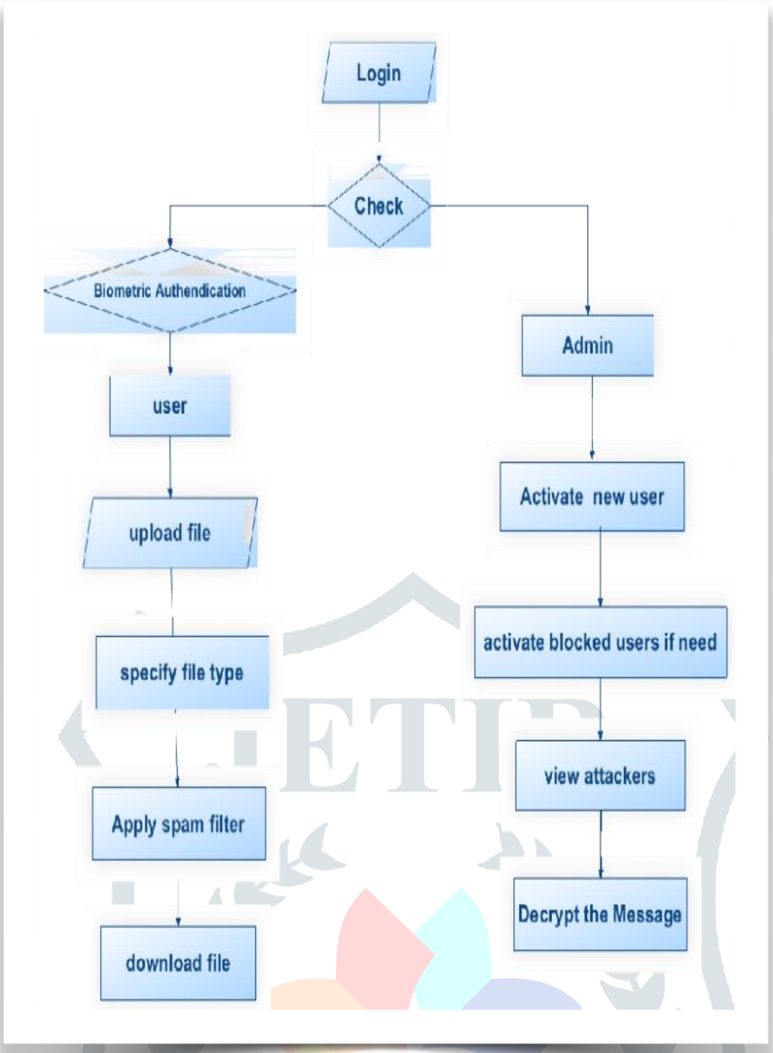

Fig.01: Dataflow Diagram

To implement the above modules we execute the following steps:

1. **User Registration:**

Users have to register themselves with the following details: name, username, password, image, mail id and phone number. These details will be used for verification at the time of user login.

2. **User Activation by Admin:**

Once the user is registered a notification with the message "registered! You can login after the intimidation of admin" will appear. The admin will login on admin side and activate the new user. Now this user will be able to login.

3. **User Login with Biometric authentication:**

User can login on user side by entering the correct username, password and image chosen at the time of registration. Providing image at the time of login is a type of biometric authentication. If any of these credentials is incorrect the user will not be able to login and a relevant error message will be displayed. User will be given 3 attempts for login after which his account will be blocked.

4. **User uploads file:**

User uploads a text file of choice by browsing through the files in their system.

5. **User specifies the file type:**

User has to specify the file type: public or private. If the file is made public it can be viewed by other users who login. If the file is made private it can be viewed only by the ownerof the file.

6. **User chooses to apply spam filter on the file:**

Once the file is uploaded the user chooses to apply spam filter on the file. By doing this all the spam words in that filewill be starred.

7. **User downloads new file with spam filter applied:**

The user will now be able to download a file on which spamfilter has been applied.

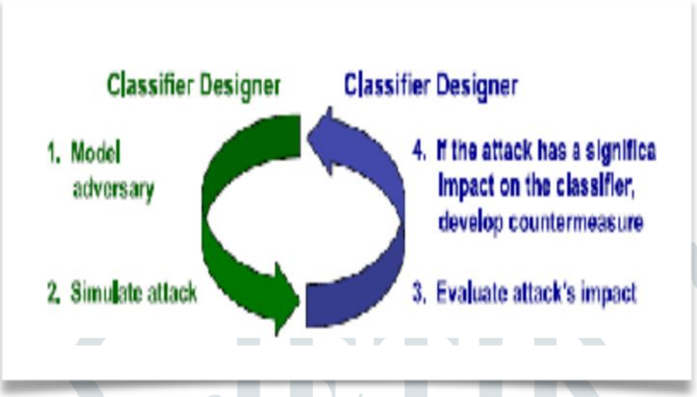
Classifier Designer
Classifier Designer
1. Model adversary
2. Simulate attack
3. Evaluate attack's impact
4. If the attack has a significa Impact on the classifier, develop countermeasure

Fig.02: System Architecture

In an attack scenario:

1. **User gets blocked :**

User will be given 3 attempts at login. If he fails to login with the correct credentials at the 3rd attempt a notification with the message “Oops ! Your account blocked due toinvalid authentication. Mail to administrator to recover account…” will appear.

2. **Activation of blocked account:**

This has to be done by admin on admin login side. The admin will go to the blocked users section and then activate the user account if needed.

3. **Details of attacker:**

The details of attacker can be viewed by admin user the attacker section. The day and time of attack along with theIP address of the attacker is available. This information can then be used to track down the attacker.

4. **Admin is also capable of monitoring user activity:**

Admin is also cable of seeing the files uploaded, if spam is applied on them or not, files downloaded along with the details of the owner of the file, time of upload or download and name of file.

.

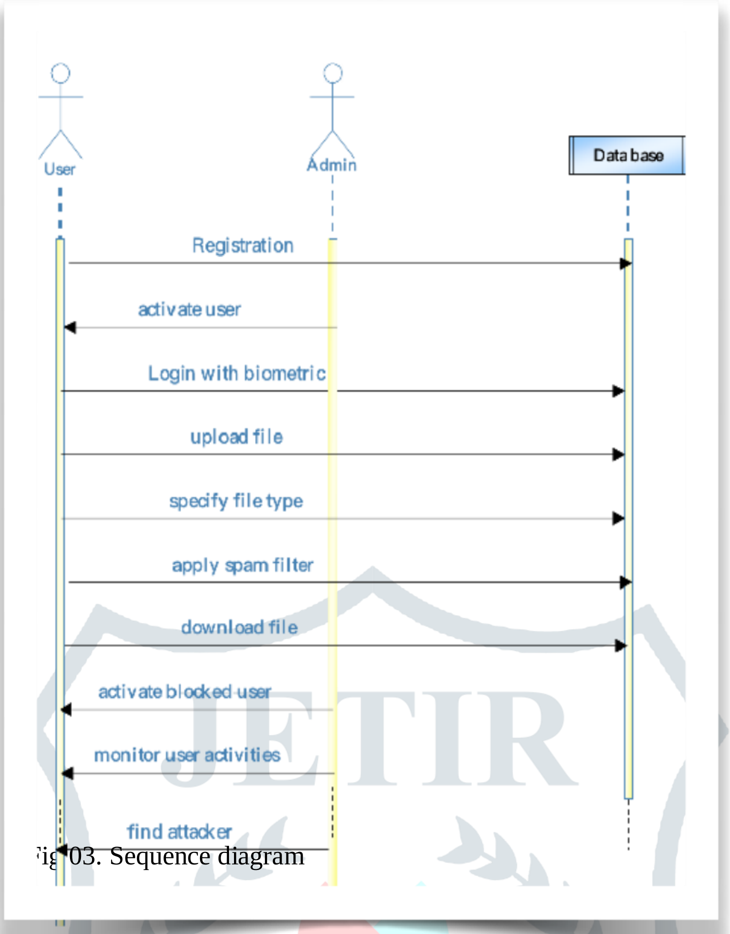

Fig 03. Sequence diagram

III. RESULTS

The proposed system provides a safe environment for users to upload and download files on which spam filter is applied. This ensures that these files do not contain any harmful information or spam words. This word can be improved by adding different biometric authentication that are more secure for example fingerprint recognition or retinal eye scan. This improves the security of the whole system and also ensures the safety and authenticity of its users.

IV. CONCLUSION

The article provides a paradigm for empirical security evaluation that formalizes and generalizes principles from past work and canbe used with a variety of classifiers, learning algorithms, and classification tasks. It is built on a formal model of the adversary and a model of data distribution that can reflect all of the assaultsconsidered in past work. It also provides a methodical procedure for developing training and test sets that enables security assessment and can take into account application-specific attack simulation methodologies. This is unquestionably an advanceover earlier work because most of the suggested solutions (whichare typically created especially for a particular classifier model, attack, and application) could not be applied to other problems without a universal framework.

V. REFERENCES

1. Ribeiro, P.B., da Silva, L.A. and da Costa, K.A.P., 2015, May. Spam intrusion detection in computer networks using intelligent techniques. In 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM) (pp. 1357-1360). IEEE. doi: 10.1109/INM.2015.7140495.

2. Junnarkar, A., Adhikari, S., Fagania, J., Chimurkar, P. and Karia, D., 2021, February. E-mail spam classification via machine learning and natural language processing. In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 693-699). IEEE. doi: 10.1109/ICICV 50876.2021.9388530

3. Gupta, P., Dubey, R.K. and Mishra, D.S., 2019. Detecting spam emails/SMS using Naive Bayes and support vectormachine. International Journal of Scientific & Technology Research, 8(11).

4. Chaudhari, N., Jayvala, P. and Vinitashah, P., 2016. Survey on Spam SMS filtering using Data mining Techniques. International Journal of Advanced Research in Computer and Communication Engineering, 5(11).

5. Rahman, S.E. and Ullah, S., 2020, June. Email spam detection using bidirectional long short-term memory with a convolutional neural network. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 1307- 1311). IEEE. doi: 10.1109/TENSYMP50017.2020.9230769.

6. Shrivastava, S. and Anju, R., 2017, December. Spam mail detection through data mining techniques. In 2017 International conference on intelligent communication and computational techniques (ICCT) (pp. 61- 64). IEEE.doi: 10.1109/INTELCCT.2017.8324021.

7. Makkar, A., Garg, S., Kumar, N., Hossain, M.S., Ghoneim,
A. and Alrashoud, M., 2020. An efficient spam detection technique for IoT devices using machine learning. IEEETransactions on Industrial Informatics, 17(2), pp.903-912. doi: 10.1109/TII. 2020.2968927.

8. Baaqeel, H. and Zagrouba, R., 2020, November. Hybrid SMS Spam Filtering System Using Machine Learning Techniques. In 2020 21st International Arab Conference on Information Technology (ACIT) (pp. 1-8). IEEE. doi: 10.1109/ACIT 50332.2020.9300071.

9. Abdullahi, M., Mohammed, A.D., Bashir, S.A. and Abisoye,O.O., 2021, February. A Review on Machine Learning Techniques for Image Based Spam Emails Detection. In2020 IEEE 2nd International Conference on Cyberspace (CYBER NIGERIA) (pp. 59-65). IEEE. doi: 10.1109/ CYBERNIGERIA51635.2021.9428826.

10. Abinaya, R. and Naveen, P., 2020, July. Spam Detection On Social Media Platforms. In 2020 7th InternationalConference on Smart Structures and Systems (ICSSS) (pp. 1-3). IEEE. doi: 10.1109/ICSSS49621.2020.9201948.

11. Alsaleh, M., Alarifi, A., Al-Quayed, F. and Al-Salman, A., 2015, December. Combating comment spam with machine learning approaches. In 2015 IEEE 14th InternationalConference on Machine Learning and Applications(ICMLA) (pp. 295-300). IEEE. doi: 10. 1109 /ICMLA.2015.192

12. Sethi, P., Bhandari, V. and Kohli, B., 2017, October. SMS spam detection and comparison of various machine learning algorithms. In the 2017 international conference on computing and communication technologies for a smart nation (IC3TSN) (pp. 28-31). IEEE. doi: 10.1109/IC3TSN
.2017.8284445

13. Kumar, S., Gao, X., Welch, I. and Mansoori, M., 2016,March. A machine learning-based web spam filtering approach. In 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA) (pp. 973-980). IEEE. doi: 10.1109/AINA.2016.177

14. Gupta, M., Bakliwal, A., Agarwal, S. and Mehndiratta, P., 2018, August. A comparative study of spam SMS detection using machine learning classifiers. In 2018 Eleventh International Conference on Contemporary Computing (IC3)(pp. 1-7). IEEE. doi: 10.1109/IC3.2018.8530469