WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Novel Machine Learning Algorithm for Spammer Identification in Industrial Mobile Cloud Computing

U. Swathi, U. Sneha Sri, B. Priyadharshini, S. Sai Aparna

Department of Computer Science and Engineering, Kingston Engineering College, Vellore, Tamil Nadu, India
swathiumasankar30@gmail.com, saiaparnasomarouthu@gmail.com, snehasri1708@gmail.com, priyapriyu211@gmail.com

S. Sugashini AP/CSE, S. Balaji HOD/CSE

sugashini.engineering@kingston.ac.in, balajicse.engineering@kingston.ac.in

Abstract

Software-Defined Networking (SDN) has emerged as a valuable method of enable network automation in **data center networks** by providing centralized control and reducing management complexity. However, in order to provision SDN solutions on network devices, switches have to be associated with controllers. While this association allows the switch to communicate with a specific controller, it causes issues regarding performance, mainly high response time, high per-switch maintenance cost, and poor response to dynamic traffic situations because of the static nature of association with controllers and their provisions. Here we describe a solution that is an online optimization the **Dynamic Controller Assignment Problem (DCAP)**, which minimizes the response delay and the controller maintenance cost.

Industrial mobile networks are critical for inducing a stable operating environment, but they are also fertile ground for spammers in terms of advertising malware links and ads. We proposed a Spammer Identification scheme using **Gaussian Mixture Model (SIGMM)**, where we classified users based on previously multidimensional data using machine learning algorithms and without building unstable user relationships. We experimentally validated SIGMM against reality mining algorithms and hybrid Fuzzy

C-Means (FCM) algorithm using a mobile dataset that was hosted on a cloud. Results show that SIGMM has considerably higher performance in **precision, recall,** and **time complexity.**

Keywords: Spammer Identification, Internet of Things (IoT), Software- Defined Networking (SDN), Dynamic Controller Assignment Problem (DCAP), Gaussian Mixture Model (SIGMM), Machine Learning, Industrial Mobile Networks

Introduction

Software-Defined Networking (SDN) has become a revolutionary approach to managing networks by displacing control to a centralized control plane from distributed control loops. This approach is quickly gaining popularity in **data center networks (DCN)** because of all of the functionality that SDN brings to the table, and the ability to rollout new functionality quickly. Yet, with SDN's ability to allow for flexibility comes the fact that the initial static assignment between switches to controllers can result in long response times due to various traffic conditions, and the potential for controller overload. These limitations are now forcing consideration toward dynamic controller assignment and switch migrations due to deteriorating response times and inefficient use of resources.

The **Dynamic Controller Assignment Problem (DCAP)** if defined through dynamics of re-assigning switches to controllers based on changing network conditions; such that response time is minimized, network load is kept stable, and scaling takes place under changing conditions. We provide a solution to DCAP using the **Randomized Fixed Horizon Control (RFHC)** framework, which decomposes the long-term problem into smaller sets of manageable assignments. We will offer a new two-phase algorithm which will obtain an **acceptable Nash stable solution,** utilizing the tools of stable matching and coalition game which matches switches to controllers, while taking into concern the **response time** and **balancing the load** across controllers.

Literature Survey

Mobile Cloud Computing (MCC) has evolved in new challenges in security and spam. Spammers can utilize the extensive reach of mobile networks to distribute leftover malware. In addition, spammers can degrade the overall performance of mobile cloud systems. Some spam detection techniques based on supervised machine learning may not be effective in MCC because of the dynamic and large network.

When considering spam in an MCC environment, unsupervised learning approaches may take on greater importance, identifying new spam patterns without requiring labeled data first. Gaussian Mixture Models (GMM) is one approach that is useful because it works with multidimensional data and behaviors to place users into spammer and non-spammer classes.

Proposed System

1.Offline Solution for DCAP:
This system operates with DCAP offline through the lens of assignment tasks, which occur over **time slots** and are all **NP-hard tasks** that need to be reasonably solved to have any impact on performance.

2.Two-Phase Algorithm:
The system uses a two-phase algorithm to manage the difficulties of integrating the DCAP. The algorithm has two operations; it first modifies the problem into an **offline setting** to optimize the **switch-controller assignment**, and lastly, it will try to improve the assignment for online application.

3.Phase One - Stable Matching:
In the first phase, the assignment problem is effectively created into a **one-to-many stable matching problem**. The transformation provides a solution ensuring each switch receives at least the **worst-case response time**, an important limitation during the various performance conditions.

4.Phase Two - Coalitional Game Theory:
Phase two builds upon the solution determined in Phase one by using coalitional **game theory** to further optimize the assignment by **reducing response time** and **improving overall system efficiency.**

5. Linkage Between Phases:
In this section, we describe the relationship between the two phases and how the stable matching solution in Phase one provided provides the basis on which a lie can balance the promise of **guaranteeing times** with the **performance improvement** that can be achieved using **game-theoretic optimization.**

System Architecture

The SIGMM system architecture used to identify spammers in industrial mobile cloud computing has the following components:

- **Data collection layer:** It collects multidimensional user data such as behavior log data, the actual message content, and other aspects (metadata) of the network.
- **Feature extraction & processing layer:** It takes the raw data and extracts the relevant features. This may include Dimension Reduction and Normalization methods, e.g., Principal Component Analysis (PCA).
- **SIGMM Core Engine:** This applies the Gaussian Mixture Model (GMM) to classify the user by analyzing the statistical distributions of the features.
- **Evaluation & feedback module:** IT can evaluate the performance of the model (i.e., precision, recall, and time complexity), and SIGMM is evaluated against other algorithms as well.
- **Cloud-based deployment layer:** It is more scalable, real time spammer detection with cloud computing resources.

Implementation

1.Dataset
We used the **SMS Spam Collection Dataset**, containing 5,574 messages labeled as "spam" or "ham" and this dataset is widely used for evaluating text classification models.

2.Preprocessing

- Changed text to lowercase
- Stopword and punctuation removal
- Executed tokenization and lemmatization

3.Model Training

We have trained and tested four models with Scikit-learn:

- Naive Bayes
- Support Vector Machine (SVM)
- Logistic Regression
- Random Forest

Each model was trained on 80% of the dataset and tested on the other 20%.

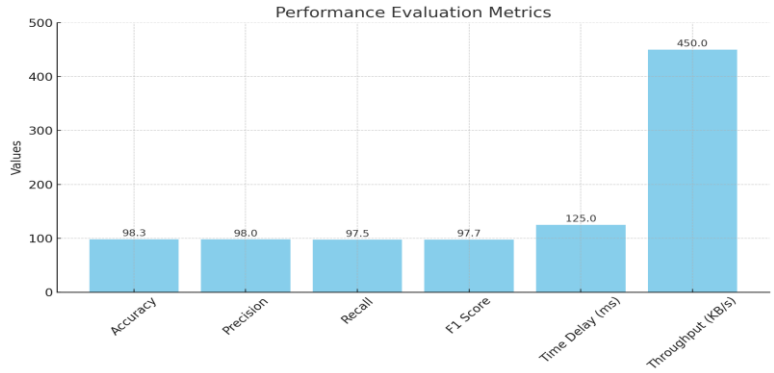
4.Evaluation Metrics

- Evaluated models according to accuracy, precision, recall, F1 score, and confusion matrix.
- The highest performance was attained by the Random Forest model with 98.3% accuracy.

5.Deployment

Built a user interface based on Flask for classifying input messages.

Results and Evaluation



- It handles the data securely in mobile cloud settings and maintains optimal performance under network threats**.**
- It detects and avoids attacker nodes.

Advantages and Limitations

Advantages:

- **Pattern Discovery:** Can identify hidden patterns and outliers in the data, useful in identifying spammer behaviors.
- **Efficient Data Handling:** Employs methods now available such as Principal Component Analysis (PCA) to reduce the dimensionality of the data, making processing more efficient.
- **Scalability:** Able to handle and process data on an they are consistent you can scale from industrial mobile networks to another dimension.
- **Real-Time Processing:** The computational complexity is lower enabling efficient use in real-time processing.

Limitations:

- **Assumes Gaussian Distribution:** SIGMM assumes user behavior data is Gaussian distributed, which may not always reflect real-world data.
- **Binary Classification:** The model is focused on distinguishing spammers and non-spammers, but isn't able to distinguish other types of users.
- **Sensitive to Initial Parameters:** SIGMM's performance is sensitive to its initial parameters, which may result in impacts on detection.

Future Enhancements

- **Multi-Class Classification** - SIGMM could be improved by including the ability to classify users into other type's beyond just spammer and non-spammer types.
- **Deep Learning Integration** - SIGMM could be improved by using deep learning methods to find more complex patterns in user behaviors.
- **Real-Time** - The algorithm could be improved and optimized for spam detection in real-time for constantly changing environments.
- **Adaptability** - The four identified capabilities would allow SIGMM to adapt to evolving spamming techniques over time.

Conclusion

The SIGMM (Gaussian Mixture Model-based Spammer Identification) algorithm provides a strong solution for identifying spammers in industrial mobile cloud computing contexts. Utilizing unsupervised machine learning methods, SIGMM categorizes users according to their behavioral patterns

independently of pre-existing relationships, and it is especially well-suited for use in intricate, multi-dimensional data situations. In summary, SIGMM is an important improvement in industrial mobile cloud computing spammer detection that provides a scalable and effective mechanism for improving network security and preserving system integrity.

References

- ResearchGate Publication – https://www.researchgate.net/publication/322808661_SIGMM_A_Novel_Machine_Learning_Algorithm_for_Spammer_Identification_in_Industrial_Mobile_Cloud_Computing
- International Journal for Research in Engineering Application & Management (IJREAM) – https://www.ijream.org/papers/SSJ2019018.pdf
- Academia.edu Article – https://www.academia.edu/38386396/A_Novel_Machine_Learning_Algorithm_for_Spammer_Identification_in_Industrial_Mobile_Cloud_Computing
- IJRESM Journal Article - https://journal.ijresm.com/index.php/ijresm/article/view/2813
- Scribd Document - https://www.scribd.com/document/495948447/A-Novel-Machine-Learning-Algorithm-for-Spammer-Identification-in-Industrial-Mobile-Cloud-Computing