Blockchain Leveraged Cyberbullying Preventing framework

Md Anawar Hossen Wadud
Mawlana Bhashani Science and Technology University
Md Ashraf Uddin (mdashrafuddin@students.federation.edu.au)
Jagannath University https://orcid.org/0000-0002-4316-4975
Shamima Parvez
Premier University
Mohammad Motiur Rahman
Mawlana Bhashani Science and Technology University
Ammar Alazab
Melbourne Institute of Technology
Omi Akter
University of Chittagong

Research

Keywords: Fog architecture, cyberbullying, Blockchain ,classification, feature selection, natural language processing, security

Posted Date: April 6th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-21075/v1

Blockchain Leveraged Cyberbullying Preventing framework

Md Anwar Hossen Wadud[1], Md Ashraf Uddin[2], Shamima Akter[3], Md Motiur Rahman[4], Ammar Alazab[5], and Omi Akter[6]

Mawalana Bhashani Science and Technology University[14] (mahwadud, mm73rahman)@gmail.com[14]
Federation University Australia[2] mdashrafuddin@students.federation.edu.au[2]
Premier University[3] shamima.kamal@yahoo.com[3]
Melbourne Institute of Technology[5] aalazab@mit.edu.au[5]
University of Chittagong[6] omiaktercu@gmail.com[6]

Abstract. The popularity of social media has exploded worldwide over the last few decades and becomes the most preferred mode of social interaction. The internet also provides a new platform through which adolescents are being bullied. Appropriate means of cyberbullying detection is still partial and in some cases very limited. Moreover, research on cyberbullying detection extensively focuses on surveys and its psychological impacts on victims. However, prevention has not been widely addressed. To bridge the gap, this paper aims to detect cyberbullying efficiently. This paper employs a standard machine learning method and natural language processing technique as a part of the detection process in decentralized Blockchain leveraged architecture. We provide a fog based architecture for cyberbullying detection, aiming at relieving the server's load by placing the detection and the prevention of cyberbullying processes at the fog layer. The proposal might offer a probable solution to save users, particularly adolescents from severe consequences of cyberbullying.

Keywords: Fog architecture, cyberbullying, Blockchain ,classification, feature selection, natural language processing, security

List of Abbreviations

DoS = Denial of Service
PKI = Public Key Infrstructure
NLP = Natural Language Processing
ML = Machine Learning
PoS = Proof of Stake
PoW = Proof of Work
BPD = Blockchain based Prevention and Detection

1 Introduction

The exponential growth of electronic and computer-based communication has dramatically changed an individual's form of communication and cyberbullying becoming a growing research issue for the researcher in today's hyper-connected society. Nowadays, more than half of young social media users are the victim of cyberbullying [3]. Adolescents are spending a noticeable amount of time in the social network to share enough personal information such as name, gender, age, e-mail id, mobile number, current city, religion, own interest, education, hobbies and so on. They are more vulnerable to cyberbullying negative consequences. Recent data indicates that bullying and suicide are becoming more common in adolescent. They commit suicide on a large scale after unmasked to cyberbullying [18]. Hinduja and Patchin [12] conclude that youth who experienced cyberbullying, as either perpetrator or victim, had more suicidal thoughts and were more likely to attempts suicide than those who had not experienced such type of peer aggregation. The adverse effects of cyberbullying also include depression, sad moods, school drop out, and frequent nightmares.

Cyberbullying is the method of repeatedly harming or harassing people in a wide range with the use of the internet, and electronic devices as mobile, computer [16]. Cyberbullying includes verbal harassment, uploading obscene pictures, unauthorized access to other's accounts on social media, impersonation, denigration [17]. The intangible nature of electronic devices made cyberbullying more popular. Twitter is one of the leading social media with 328 million monthly active users, and in Australia, the number is approximately 3 million [2] until July 2017. Due to et al. [9], showed that 10.9% of school children from 35 different countries aged 11 to 15 years were being bullied twice or thrice in a month. In Australia, the percentage is 28 for the children to be bullied in several weeks [5]. However, Australia does appear to have higher levels of internet use by children compared to other developed countries. Additionally, Angels Hope [10] shows that 1 in 5 children aged between 8 and 15 who have experienced cyberbullying in Australia, tentative to severe depressions and even suicide attempts. About 20% of Australian have experienced image-based abuse [15]. When someone shares intimate, nude or sexual image without the consent of a victim might feel embarrassed, humiliated, annoyed, angry, depressed or devastated.

There are some significant challenges due to the logical and fake or untrue existence of bully in online tons of unstructured data online to analysis for the detection of cyberbullying. The critical challenges in this area are to design an effective method that includes all aspects of offenses in the unreasonable amount of Twitter data to label, detect and prevent bullying behavior, figure out an efficient to label data effectively and discovering appropriate steps to mitigate cyberbullying against adolescent.

In real life, while most of the communications are neutral or positive, but some communications are offensive. Mishna et al. [15], proposed identified few properties such as physical structure, race and culture, sexuality, and level of intelligence to detect cyberbullying in school children and adolescents. The perpetrator is capable of harassing a victim with a high level of anonymity without

their won actual identity as the bullying is done online and does not require the physical existence of a victim. Although research [23] has been done to detect cyberbullying, an efficient decentralized architecture to run the bullying detection algorithm has not been addressed. In the context of the social network, we focus on cyberbullying detection and its solution to make the social network platform secure for young people. The proposed method aims to recognize the bully words from the posts or comments by training a classifier.

In this paper, our contribution includes a proposal of a Blockchain leveraged decentralized architecture to facilitate cyberbullying detection and prevention, and a lightweight decentralized consensus mechanism based on Proof of Stake for the Blockchain. We also propose a hybrid cyberbullying detection method to be executed at Fog devices.

Section 2 reviews related work. We describe our architecture and method in Section 3. Finally, we present high-level performance analysis, implementation of the model on a private Blockchain and accuracy of the detection method in Section 4 following with a conclusion.

2 Related Work

An online social network allows a user to stay connected with friends and relatives more easily. People enrich their relationships by sharing daily events and important moments. The way of sharing information is more open and details in adolescents. Some approaches have been proposed to tackle online bullying. The paper [17] shows that the aggressive behavior of anonymous users is more potential to lead cyberbullying. A different approach was suggested in 2012 by Chen et al. [4]. Specifically, to predict the probability of to post offensive content, they included user's writing style, structure and specific bully contents as a feature and introduced Lexical Syntactic Feature-based (LSF) model in the detection of online cyberbullying from person's post pattern. The paper [19] proposed a data mining methodologies are applied to detect cyberbullying time series modeling that identifies predator tactic in bullying which is determined by previous the state. This method also considers bag-of-words, emotions, slang and abbreviations, a dataset of real-world conversation, and a numeric label is used to indicate the severity of the predator's comments. Zhao and Mao [32] proposed a semantic-enhanced marginalized auto-encoder method to address hidden cyberbullying feature in Twitter and MySpace data. Many researchers proposed a machine learning approach to identify negative words and to learn language patterns between predators and victims with the ultimate goal of detecting the possible occurrence of cyberbullying [28]. Xu et al. [29] proposed NLP methods to investigate questions (which text underlying bullying, form, location, time, perpetrator and their evolution over time) of bullying episodes. The automated detection of cyberbullying is quite inefficient and straightforward, and regular expressions are the necessary tools to filter profane words. With careful feature extraction, different researchers show that it is possible to predict suspended users.

In some scenarios, the spammers use software to post malicious comments on their forums and blogs to increase the popularity of their sites in the search engine. However, the concept of spamming has extended to a broader concept. Yin et al. [30] proposed to combine BoW features, sentiment, and contextual features to train a support vector machine (SVM) to detect so-called online harassment, in which a user intentionally annoys other users in a web community. Kontostathis et al. [13] developed solutions for spam filtering to detect cyberbullying with precision and false positive. They collect comments from FormSpring.me website, where users are allowed to post questions randomly and invite openly to answer questions and proposed a "bag-of-words" model. They detect cyberbullying from online posts by using a supervised machine learning technique.

The authors [17]presented machine learning classification methodology more accurately to detect user's cyberbullying behavior on Twitter, and they employ their Random Forest classifier with the machine learning method. Dinakar et al. [8] discovered cyberbullying by decomposing the problem according to topics on YouTube by the support vector machine classifier. Instead of features and characteristics of both perpetrator and victim, written comments are the primary focus in cyberbullying detection. Apart from these proposals, Dadvar et al. [6] used gender information with a supervised learning method. In 2013, Dadvar et al. [7], proposed to emphasize on user's activity history to improve cyberbullying detection accuracy. But a significant limitation of this is the dependency in the Bag-of-Words assumption, which may not be robust. However, Sanchez and Kumar [21] use Twitter Streaming API and text processor to speed up the process of labeling data and uses the Ling Pipe Naive Bayes machine learning classifier to detect cyberbullying. To address different emotions in identifying bullying, some researchers worked in sentiment analysis. Sentiment analysis of text also contributes extensively in the field of bullying detection by studying natural language processing in social media. Many algorithms are developed to learn whether words and phrases are used to express positive or negative sentiments. To identify the risky user, Xu et al. [4] depend on sentiment on Twitter data. But, there are no specific models to detect cyberbullying. Moreover, Go et al. [11], introduced a machine learning method to classify sentiments according to emotion from Twitter messages automatically. Wyman et al. [27] proposed Sources of Strength suicide prevention program to enhance protective factors related to the reduction of suicidal attempts at the school level where peer leaders play an essential role.

Although solutions to prevent bullying include parents' effort, educators and law enforcement, but a technological advancement in filtering is not incorporated with current research. Nowadays, the attempt to stop bullying is now an emerging issue to save adolescent from the severe consequences of cyberbullying. The existing research has not focused on the development of a decentralized secure architecture for detecting and preventing cyberbullying. Blockchain technology has been successfully introduced in cryptocurrency to ensure high security and privacy. Further, Blockchain has been explored to use in healthcare, supply chain,

and the Internet of Things. Blockchain technology in detecting and preventing cyberbully has not still explored. In this article, we propose a Blockchain oriented cyberbullying framework to enhance the security of the detection method.

3 The CyberBullying Prevention Architecture

We design a Fog based architecture for detecting and preventing cyberbullying. The proposed architecture depicted in figure 1 includes the user's device, Fog devices, and servers, often remote Cloud storage. The architecture is described below.

Most of the conventional cyberbullying architecture comprises server and client. In General, bullying prevention or detection method is installed on a server or a client. The disadvantages of such a paradigm are that users might experience poor QoS(Quality of Service). The server or client requires to keep the bullying prevention module active all the time which demands higher power consumption in the server or client. Further, the bullying prevention method installed on the server or client is often vulnerable to cyber-attacks including Ransomware, Denial of Service(DoS), and malware. To overcome these shortcomings, we proposed a tier-based secure cyberbullying preventing architecture. The architecture has three tiers where the bottom tier includes user's devices including smartphone, laptop, and computer, the middle tier consists of Edge or Fog devices and the upper-tier includes traditional servers or Cloud servers. Fog [14] has recently brought computing resources closest to user's devices. In Fog technology, computing capacities such as storage, processing are provided to a conventional router, switch, and Gateway that is deployed close to IoT(Internet of Things) devices. User's devices send data to Fog node which performs preprocessing before transferring the data or results to the Cloud server for further processing. Cloud services are often migrated to Fog devices to provide users with QoS such as low latency. In our proposal, the bullying prevention applications reside on the middle tier's devices to have faster processing of the Edge or Fog computing.

But, the bullying detection and prevention applications on a particular Fog device can be attacked and compromised by malware or other cyber rogues or malicious users. Ensuring appropriate security and privacy at Fog devices is challenging because diverse stakeholders belong to Fog devices and often remain unattended. Besides, heterogeneous Fog devices are attributed to diverse security protocols and standards. Therefore, there needs a common mechanism to prevent malicious attackers from modifying or stopping bullying prevention modules on the Edge network. Blockchain [24] has recently emerged as a promising technology to withstand major security attacks. The Blockchain runs on a peer to peer network and maintains a tamper-proof ledger replicated among each node that participates in mining. A chunk of data also called Block is added to the Blockchain after a majority of Blockchain's nodes reach an agreement that the new Block is formatted following a suite of Blockchain's rules. This process is called a consensus protocol. Once a Block is validated and then included in the

Blockchain network, the Block cannot be modified or changed further. A smart contract that represents a set of rules encoded by any programming language is stored in every Blockchain node. The smart contract is triggered when a particular event related to that contract occurs. For example, the smart contract to transfer transactions from a buyer is automatically executed upon receiving a car from a seller and it does not require any third parties as an intermediary. In this architecture, the bullying identification and prevention module will be stored on the Blockchain as a smart contract. The Edge network facilitates peer to peer network to operate the Blockchain. The Fog device runs the Blockchain algorithm to enforce high security and availability of the cyberbullying prevention module. Further, the user does not need to trust third parties providing services such as Facebook to run the cyberbullying detection module.

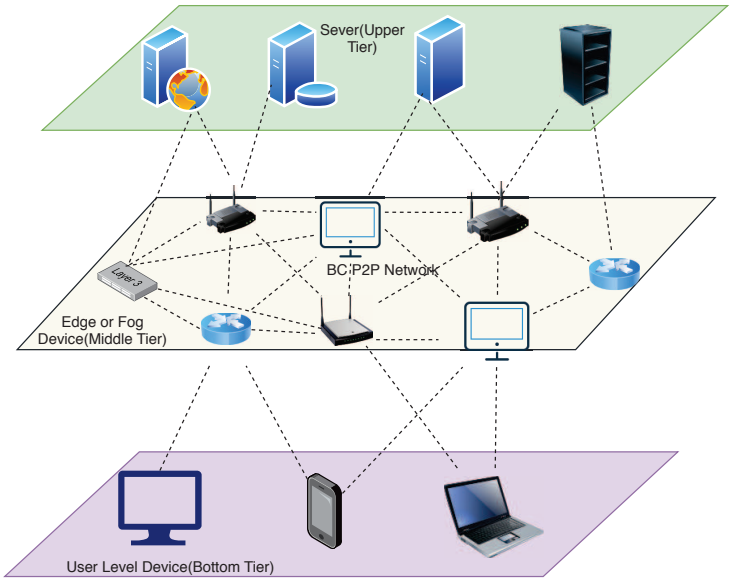


Fig. 1: A tier based Blockchain leveraged architecture

A typical Blockchain is depicted in figure 2. Logical components of a typical Blockchain include transaction, Block, and consensus protocol. The logical components of a Blockchain are described below.

Transaction: Basic data unit in the Blockchain is called transaction. A transaction is generated by the Blockchain wallet when health data or money is required to transfer from one account to another. A transaction typical contains sender, receiver address, digital signature, and health data. Digital signature and sender or receiver address are formed using PKI(Public Key Infrastructure) that features anonymous properties of the transaction. A data transaction format is illustrated in Table 1.

Table 1: The General Format of Transaction

Transaction Identifier			
Source Address		Receiver Address	
Digital Signature	Sender pubKey	Script	Receiver pubKey
Data			

Block: Transactions created on the Blockchain peer to peer network is stored in a pool. The miner organizes a certain number of transactions into a Block. The transactions are packed into the Merkle tree in the Block to preserve the integrity of transactions. A typical Block contains two parts- header and contents. The header has nonce, timestamp, and previous hash code field. The content part holds the root of the Merkle tree depicted in figure 2. The previous hash field contains the hash of the latest Block of the current Blockchain. The nonce field is incremented to come up with a target hash from the Block.

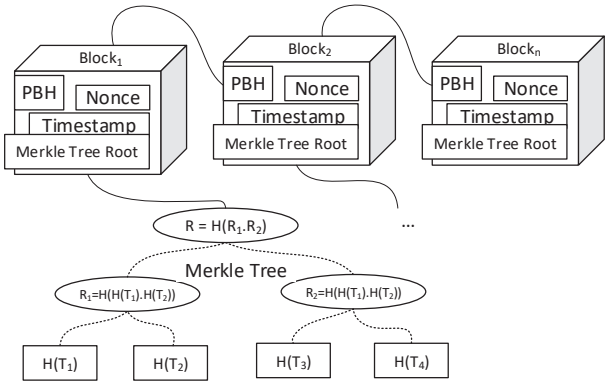


Fig. 2: A typical Blockchain

3.1 Modified Proof of Stake mechasnism for Fog network

The consensus protocol is a mechanism executed by a particular group of network nodes called Miners to reach the agreement regarding the validity of a new Block to be inserted into the Blockchain. Bitcoin and Ethereum Blockchain apply Proof of Work(PoW) where Miners require to solve a mathematical puzzle that represents a hash code having certain numbers of leading zeroes. The miner first performs PoW is rewarded. The Blockchain nodes verify the previous hash code, the target hash code, digital signature of transactions, and others of the Block's header before inserting the Block into the Blockchain. The node at Fog

computing varies in the level of processing, and memory capacity. Typically, Fog devices are attributed to limited resources. Therefore, Edge-based Blockchain can not accommodate Proof of Work(PoW) that involves high computational cost. In contrast, Proof of Stake(PoS) which is power and time-efficient is suitable for Fog devices. With PoS, a node with a higher stake has the highest probability to mine the next Block. But, PoS is less decentralized than that of PoW. The rich node always mines the Block and becomes richer. We propose to divide Fog based Blockchain network into a different zone as depicted in figure 3. The miner from a zone is nominated based on the following rules.

- **The first round**, let one node with the highest stake from a zone mine the next Block.
- **The second round**, let one node with the highest availability from a zone mine the next Block.
- **The third round**, let one node with higher processing capabilities from a zone mine the next Block.
- **The fourth round**, let one node with the smallest stake from a zone mine the next Block.
- **The fifth round**, let one node with a medium stake from a zone mine the next Block.
- **The sixth round** lets randomly one node from a zone mine the next Block.

The miner will be selected from each zone repeatedly according to the above-mentioned rule. More rules can be set to make the selection more decentralized.

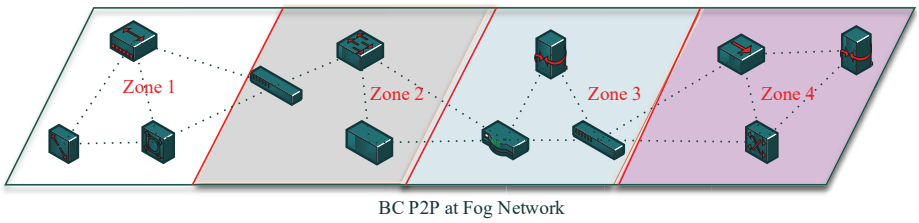


Fig. 3: Different zone of a Fog based Blockchain network

3.2 The Cyberbullying Detection Method on the Blockchain

We utilized the Fog Gateway to prevent and detect the cyberbullying in a social network as a mediator. In a server-client paradigm, a user has to directly experience the negative impact of cyberbullying if the bullying can not be prevented on time. Detection and then prevention at server sides might cause a delay in responding to user's queries. Fog devices executing the detection and prevention method on behalf of the server or client can make processing faster. The smart contract on Bullying Detection and Prevention(BDP) activates whenever user

logins to his or her social media profile like Facebook, twitter. All the requests to be performed on the user's profile go through the BPD module at the Fog device. If a stranger wants to comment on the social profile of a user, the BDP decides to allow him or her to leave comments or simply block the access.

In this paper, we propose a hybrid bullying detection and prevention method that includes natural language processing(NLP), and machine learning(ML). The hybrid method deployed at Fog device is illustrated in figure 4. The method combines natural language processing and machine learning to efficiently identify and stop bullying. When a user leaves any comments on a user's profile, a score is generated by analyzing the texts through NLP(Natural Language Processing) before committing the comments on the profile. The machine learning approach identifies a stranger as bullies or nonbullies based on several features such as previous history, age, profession, login pattern, friend list, etc.

3.3 Blockchain Based Cyberbullying Prevention

In our proposed hybrid bullying detection and prevention method, deep learning model used to detect bullying messages from different attacker and malicious entities. The Blockchain network has been used to track valid personal messages and attackers' information. Bullying messages are added to the existing chain of Blocks. As a result, this results in the prevention of bullying from the same attacker. The Blockchain network can keep track of all personal information including name, gender, age, mobile number, religion, education, hobbies, and personal messages of all users in the network and monitor all kinds of incoming messages or other social communications such as fake email, fraud lottery.

In this process, a user can send any type of messages to a receiver. The deep learning model detects if message is cyberbullying using the training dataset stored in the distributed ledger of the Blockchain. If the deep learning model detects a new message as cyberbullying, it stores this information into a separate chain of Blocks so that the receiver do not see it. The new message will be added to the existing training dataset for future use. In contrast, if the new message has been identified as a valid one, the model stores this message in another chain of Block and the message is transmitted to the receiver.

The Blockchain helps to store the history of communication messages from every user in an immutable chain. Consequently, the deep learning model can form a big training dataset by collecting communication history of each user from the Blockchain. Blockchain facilitates the secure sharing of user's messages and hosts the deep learning model in secured way. Further, a sender generates the hash code of his or her message and stores it in the Blockchain before sending the message to the receiver. If an attacker changes a message, first fog node and the receiver can re-generate the hash code of the message will check whether the hash code stored in the Blockchain is similar to re-generated hash code.

The part of natural language processing includes data collection and text categorization, pre-processing and semantic analysis.

1. **Data collection and text categorization:** Words are corrected(right spelling) to distinguish bullying words from other normal words. Bullying

words are identified by using an enriched dataset that consists of bullying and nonbullying words. The bullying word identification varies from country to country and culture to culture. For example, "Australia says yes to same-sex marriage", the sentence contains the word sex which is not identified as bullying word in Australia but the same word is considered censored word in Bangladesh.
2. **Pre-processing:** It involves removing stop-words and white space, tokenizing words, tagging part-of-speech, and lemmatizing.
3. **Semantic Analysis:** This module analyses words of the post or comments and outputs a score or rating based on the number of bullying words. Correlation-based semantic analysis [31] can be used to generate the score.

The part of **machine learning approach** includes Feature Selection, Classification, Ensemble, Classifier Selection.

1. **Feature Selection:** Feature selection indicates determining an important set of features from a good number of features. Not all attributes or features carry out the same importance to identify a person as bullies or bullied. For instance, the age of a person is not as important attribute as a previous bullying history of that person. Other kinds of features might be education, profession, number of a friend, likeness, emotional symbol, nature of sharing a document, etc. Decision tree [22] can be used to select a subset of important features. Some important features that are most responsible for bullying can be found out using a decision tree.
2. **Classification:** Some classifiers including Multilayer Neural Network, Bayesian Naive, SVM, and Random Forest are trained with the training datasets.
3. **Ensemble:** Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model to decrease variance (bagging), bias (boosting) or improve predictions (stacking). We also use an ensemble method to improve the accuracy of the classifier.
4. **Classifier Selection:** The accuracy of a classifier varies depending on the dataset. Several classifiers are trained using the same dataset. Finally, this module selects a classifier with the highest accuracy.

Finally, two outcomes produced by NLP and ML(Machine Learning) determines if the request will be committed or blocked. If the NLP score is above a threshold value and ML generates "Yes", then the system blocks the stranger. If the NLP score is below the threshold and ML generates "Yes", the comments of the stranger are committed for the first time but the stranger is listed in the suspected list for fuhrer examining. If the NLP score is above the threshold but ML produces "No", the comment of the stranger is sent to the user for the approval. If the NLP score is below threshold and ML produces "No", the comment is committed without generating any warnings.

4 Implementation and Performance Analysis

The paper focuses on detecting cyberbully efficiently thus to give timely prevention. To achieve that, we utilize Blockchain leveraged Fog based architecture

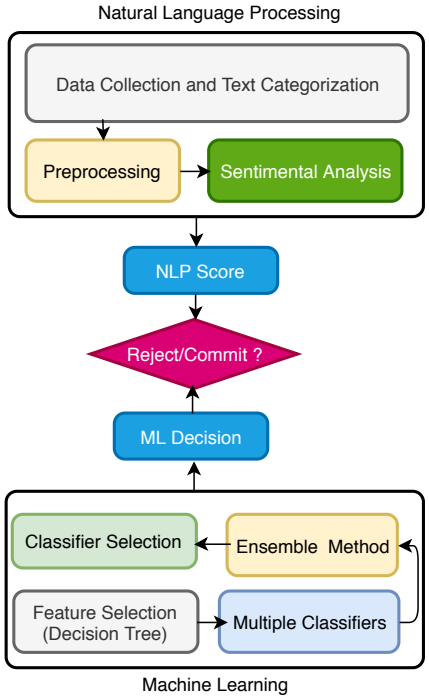


Fig. 4: Cyberbullying detection module at fog layer

with standard machine learning method and natural language processing technique. The comparison of the proposed bullying detection framework with other existing bullying detection methods is presented in Table 2.

Table 2: The comparison of the proposed approach with existing approaches

Parameters	Existing Method-1 [20]	Existing Method-2 [8]	Existing Method-3 [1]	Proposed Method
Anonymous	No	No	No	Blockchain uses public/private key pair as address of a user. Address is not identifiable.
CIA(Confidentiality, Integrity and Availability)	Low availability	low availability	Medium availability	Blockchain guarantees integrity and high availability.
Withstand Cyber Attacks	No	No	No	Blockchain can successfully withstand DoS and Ransomeware attack. Further, malware can not modify all versions of the application stored in multiple nodes
Fault Tolerance	No	No	No	If one Fog device is down, another one can be invoked to run the applications
Throughput	low	low	Medium	High as more than one Blockchain nodes can be activated to serve user's requests
Reliability	Medium	Medium	Medium	High
Power consumption	Medium	Medium	Medium	High as Blockhain mining consumes a bit high power
Latency	High	High	High	Medium as Fog devices executes the application

Figure 5 shows visualization of proposed Blockchain network where miners are labeled with uppercase letter. The network properties such as graph density, connected components, diameter, and average degree are presented in Table 3 and figure 6.

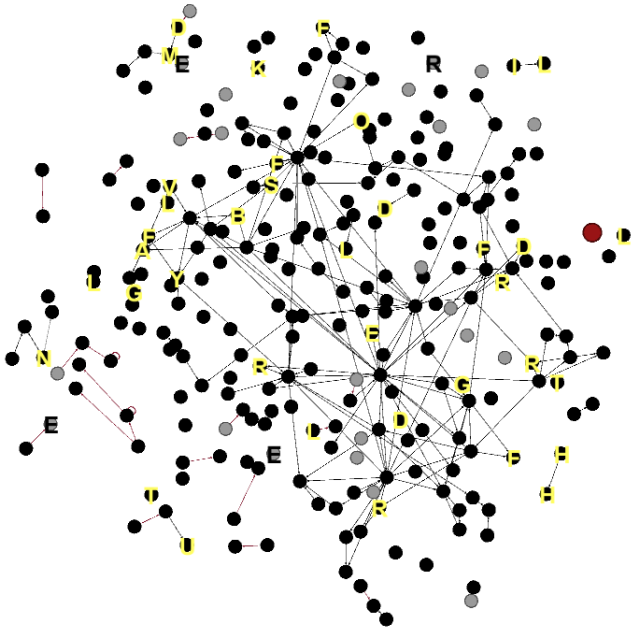
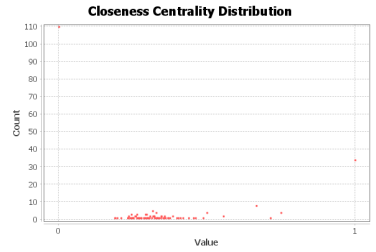


Fig. 5: Visualization of proposed Blockchain network

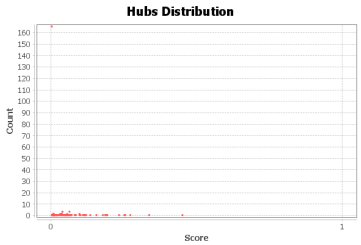
Closeness centrality distribution graph presented in figure 6(a) shows how close a node is to all other nodes in the network. Hubs distribution graph presented in figure 6(b) shows connections between different nodes. Size distribution graph depicted in figure 6(c) shows node allocation with different cluster or different subnetwork and clustering coefficient distribution graph depicted in figure 6(d) shows the number of neighbor nodes connected in a certain range.

We implemented our proposed architecture and deep learning model [1] using Python Keras, TensorFlow. The Blockchain network is visualized using Gephi software. We consider 120 nodes as fog nodes. The fog nodes randomly establish network connection between them. The MySQL database is simulated as the Cloud database to store the chain of Blocks. The main procedure of our implementation is presented in figure 7

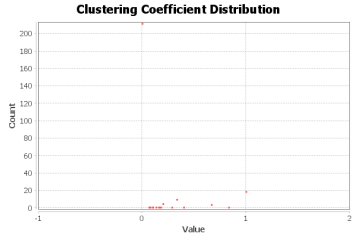
For measuring the accuracy of the model, three real-world datasets have been used: Formspring(almost 12000 posts) [20], Twitter (almost 16000 posts) [25]



(a) Closeness centrality distribution



(b) Hubs distribution



(c) Size distribution

(d) Clustering coefficient distribution

Fig. 6: Distribution graph of the proposed network

Table 3: The network overview

Network Properties	Values
Average Degree	0.76
Network Diameter	8
Average Path length	3.37
Graph Density	0.003
Modularity	0.672
Connected Components	133
Average Clustering Coefficient	0.054

and Wikipedia(almost 100000 posts) [26] and around 500 cyberbullying words (such as "fellate", "fellatio", "fingerfuck", "sex" ,"shag", "shagging", "shemale", "shit", "shitdick", "shite", "shithead shits shitted") have been collected. The number of bullying words gradually increases when the model execute. The pseudo code of the Blockchain leveraged cyberbullying detection and prevention is illustrated in figure 8.

The classifiers trained include Naive Bayesian(BN), J48, Random Forest(RF) and Multilayer Perceptron(MLP) and K-nearest neighbor(IBK). 10 fold cross-validation is used to generate the accuracy of these classifiers. Each classifier is

```
def mine():

    if len(blockchain.transactions):
        msg = blockchain.transactions[0]['value']
        badWord = blockchain.chain
        isCyberbullying = detectCyberbulling(badWord, msg)
        if isCyberbullying:
            sender.setWorningMessage("Cyberbulling Detected")
            last_block = blacklist.chain[-1]
            nonce = blacklist.proof_of_work()
            previous_hash = blacklist.hash(last_block)
             blockchain.create_block(nonce, previous_hash)
      else:
            last_block = blockchain.chain[-1]
            nonce = blockchain.proof_of_work()
            previous_hash = blockchain.hash(last_block)
            blockchain.submit_transaction(sender_address=MINING_SENDER,
            recipient_address=blockchain.node_id, value=MINING_REWARD,
            signature="")
            previous_hash = blockchain.hash(last_block)
            block = blockchain.create_block(nonce, previous_hash)
```

Fig. 7: A sample code from the implementation

Input:
Sender & Recipient Address, sender private key, encoded message.

Output:
Set Waring to sender Address & discard to send message if there have any cyberbullying related information. Else send message to recipient address.

1. Select `random_node` from n fog nodes as miners
2. `decoded_message = message_decoder (private_key, encoded_message)`
3. **Go to** step 4 to 6 for cyberbullying detection
4. Get all cyberbullying related keyword as `black_list` from cloud database
5. Apply machine learning model on `decoded_message` to detect cyberbullying based on `black_list`
6. **If** `(isCyberbullying)` then execute line 7 to 9:
7. Store new `cyberbullying_word` to `cloud_database` for further processing
8. `setWaring(sender address, "waring_message")`
9. **Exit**
10. **Else** execute following 11 to 16 line

```
nonce = blockchain. proof_of_work()
blockchain.submit_transaction (sender_address, recipient
_address, message, signature)
previous_hash = blockchain.hash(last_block)
block = blockchain.create_block(nonce, previous_hash)
```

Set `block` to recipient address

17. **Exit**

Fig. 8: Pseudocode of Cyberbullying Detection and Prevention Model

again trained with the same dataset using an ensemble method(Bagging). The accuracy of the classifier considering Blockchain and without Blockchain is illustrated in fig 9. The J48 shows the highest accuracy(92.64%) for this dataset without using the Blockchain. Fig 9 shows that accuracy is increased for each classifier except Naive Bayesian(BN) in the Blockchain based model. Random Forest and

Multilayer Perceptron's accuracy significantly improve if the Blockchain is used to maintain the training dataset.

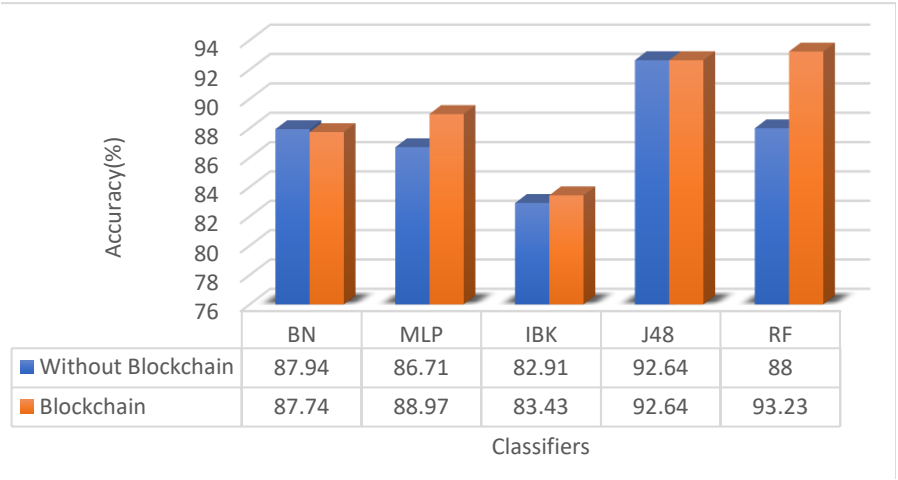


Fig. 9: Accuracy of five classifiers to detect cyberbullying

5 Conclusions

Technological methodologies might be able to protect children from cyberbullying to some extent. Cyberbullying can be mitigated by changing the norm, thinking, and behaviors of adolescents and respecting other peers. Teachers and parents could play an important role in preventing cyberbullying. In this paper, we propose a novel architecture to detect cyberbullying and combine two approaches called NLP and ML to efficiently detect cyberbullying. We are deploying the cyber detection and mitigation technique in the Fog devices. So, both users and servers are safe from bullying attacks as well as other network attacks such as data theft, privacy disclosure, etc. In the future, we will implement a prototype of the architecture to analysis the performance of Blockchain in detecting bullying and use publicly available datasets under Twitter's Terms of Service.

Declarations

Availability of data and materials

All data that has been used in the article are available online and publicly accessible. The appropriate references for the used data have been provided in the paper.

Competing interests

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding

Not Applicable

Authors' Contributions

Md Anwar Hossen Wadud collected data, studied and implemented the proposed architecture. Md Ashraf Uddin contributed to designing the idea, writing the proposed methodology and compiling the article. Shamima Parvez wrote introduction and reviewed literature. Motiur, Ammar and Omi went through the articles and reviewed it.

Acknowledgement

Authors are grateful to Internet Commerce Security Laboratory, Federation University Australia for supporting resources to implement the framework.

References

1. Sweta Agrawal and Amit Awekar. Deep learning for detecting cyberbullying across multiple social media platforms. In *European Conference on Information Retrieval*, pages 141–153. Springer, 2018.
2. John Chapin. Adolescents and cyber bullying: The precaution adoption process model. *Education and information technologies*, 21(4):719–728, 2016.
3. Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. Mean birds: Detecting aggression and bullying on twitter. In *Proceedings of the 2017 ACM on Web Science Conference*, pages 13–22. ACM, 2017.
4. Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. Detecting offensive language in social media to protect adolescent online safety. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 71–80. IEEE, 2012.
5. Donna Cross, Helen Monks, Marg Hall, Therese Shaw, Yolanda Pintabona, Erin Erceg, Greg Hamilton, Clare Roberts, Stacey Waters, and Leanne Lester. Three-year results of the friendly schools whole-of-school intervention on children's bullying behaviour. *British Educational Research Journal*, 37(1):105–129, 2011.
6. Maral Dadvar, Franciska MG de Jong, Roeland JF Ordelman, and Rudolf Berend Trieschnigg. Improved cyberbullying detection using gender information. In *Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012)*. Ghent University, 2012.

7. Maral Dadvar, Dolf Trieschnigg, Roeland Ordelman, and Franciska de Jong. Improving cyberbullying detection with user context. In *European Conference on Information Retrieval*, pages 693–696. Springer, 2013.
8. Karthik Dinakar, Roi Reichart, and Henry Lieberman. Modeling the detection of textual cyberbullying. In *fifth international AAAI conference on weblogs and social media*, 2011.
9. Pernille Due, Juan Merlo, Yossi Harel-Fisch, Mogens Trab Damsgaard, Mag scient soc, Bjørn E Holstein, Mag scient soc, Jørn Hetland, Candace Currie, Saoirse Nic Gabhainn, et al. Socioeconomic inequality in exposure to bullying during adolescence: a comparative, cross-sectional, multilevel study in 35 countries. *American journal of public health*, 99(5):907–914, 2009.
10. Andrew T Fong, Constance H Katelaris, and Brynn Wainstein. Bullying and quality of life in children and adolescents with food allergy. *Journal of paediatrics and child health*, 2017.
11. Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12), 2009.
12. Sameer Hinduja and Justin W Patchin. Bullying, cyberbullying, and suicide. *Archives of suicide research*, 14(3):206–221, 2010.
13. April Kontostathis, Kelly Reynolds, Andy Garron, and Lynne Edwards. Detecting cyberbullying: query terms and techniques. In *Proceedings of the 5th annual acm web science conference*, pages 195–204. ACM, 2013.
14. Redowan Mahmud, Ramamohanarao Kotagiri, and Rajkumar Buyya. Fog computing: A taxonomy, survey and future directions. In *Internet of everything*, pages 103–130. Springer, 2018.
15. Faye Mishna, Michael Saini, and Steven Solomon. Ongoing and online: Children and youth's perceptions of cyber bullying. *Children and Youth Services Review*, 31(12):1222–1228, 2009.
16. Manish Kumar Mishra, Sumit Kumar, Abhishek Vaish, and Satya Prakash. Quantifying degree of cyber bullying using level of information shared and associated trust. In *India Conference (INDICON), 2015 Annual IEEE*, pages 1–6. IEEE, 2015.
17. Tadashi Nakano, Tatsuya Suda, Yutaka Okaie, and Michael John Moore. Analysis of cyber aggression and cyber-bullying in social networking. In *Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on*, pages 337–341. IEEE, 2016.
18. Sourabh Parime and Vaibhav Suri. Cyberbullying detection and prevention: Data mining and psychological perspective. In *Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference on*, pages 1541–1547. IEEE, 2014.
19. Nektaria Potha and Manolis Maragoudakis. Cyberbullying detection using time series modeling. In *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on*, pages 373–382. IEEE, 2014.
20. Kelly Reynolds, April Kontostathis, and Lynne Edwards. Using machine learning to detect cyberbullying. In *2011 10th International Conference on Machine learning and applications and workshops*, volume 2, pages 241–244. IEEE, 2011.
21. Huascar Sanchez and Shreyas Kumar. Twitter bullying detection. *ser. NSDI*, 12:15–15, 2011.
22. V Sugumaran, V Muralidharan, and KI Ramachandran. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical systems and signal processing*, 21(2):930–942, 2007.

23. Junming Sui. *Understanding and fighting bullying with machine learning*. PhD thesis, The University of Wisconsin-Madison, 2015.
24. Md Ashraf Uddin, Andrew Stranieri, Iqbal Gondal, and Venki Balasubramanian. An efficient selective miner consensus protocol in blockchain oriented iot smart monitoring.
25. Zeerak Waseem and Dirk Hovy. Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL student research workshop*, pages 88–93, 2016.
26. Ellery Wulczyn, Nithum Thain, and Lucas Dixon. Ex machina: Personal attacks seen at scale. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1391–1399, 2017.
27. Peter A Wyman, C Hendricks Brown, Mark LoMurray, Karen Schmeelk-Cone, Mariya Petrova, Qin Yu, Erin Walsh, Xin Tu, and Wei Wang. An outcome evaluation of the sources of strength suicide prevention program delivered by adolescent peer leaders in high schools. *American journal of public health*, 100(9):1653–1661, 2010.
28. Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu, and Amy Bellmore. Learning from bullying traces in social media. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 656–666. Association for Computational Linguistics, 2012.
29. Jun-Ming Xu, Xiaojin Zhu, and Amy Bellmore. Fast learning for sentiment analysis on bullying. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, page 10. ACM, 2012.
30. Dawei Yin, Zhenzhen Xue, Liangjie Hong, Brian D Davison, April Kontostathis, and Lynne Edwards. Detection of harassment on web 2.0. *Proceedings of the Content Analysis in the WEB*, 2:1–7, 2009.
31. Yi Zhang, Artur Dubrawski, and Jeff G Schneider. Learning the semantic correlation: An alternative way to gain from unlabeled text. In *Advances in Neural Information Processing Systems*, pages 1945–1952, 2009.
32. Rui Zhao and Kezhi Mao. Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder. *IEEE Transactions on Affective Computing*, 8(3):328–339, 2017.

Figures

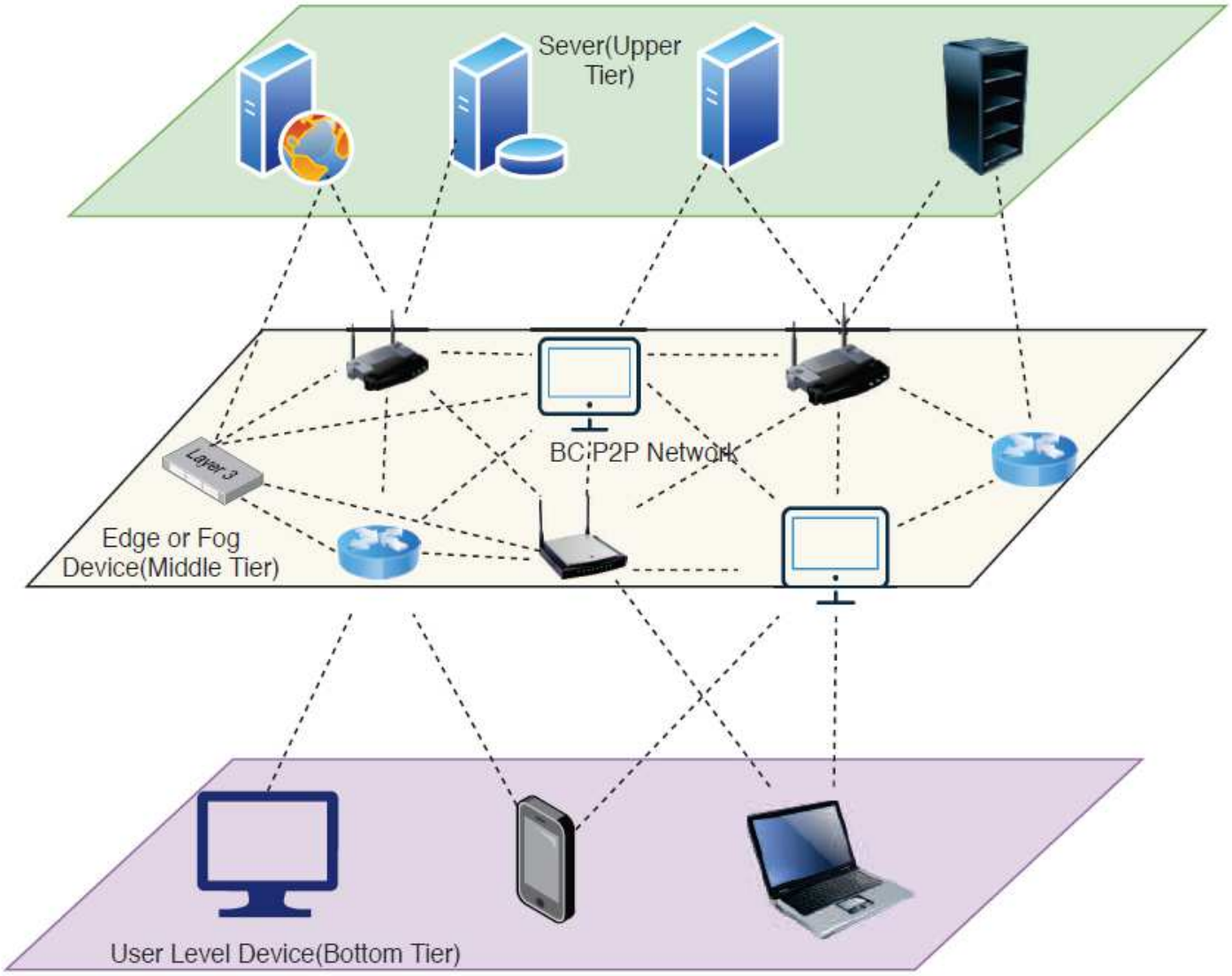


Figure 1

A tier based Blockchain leveraged architecture

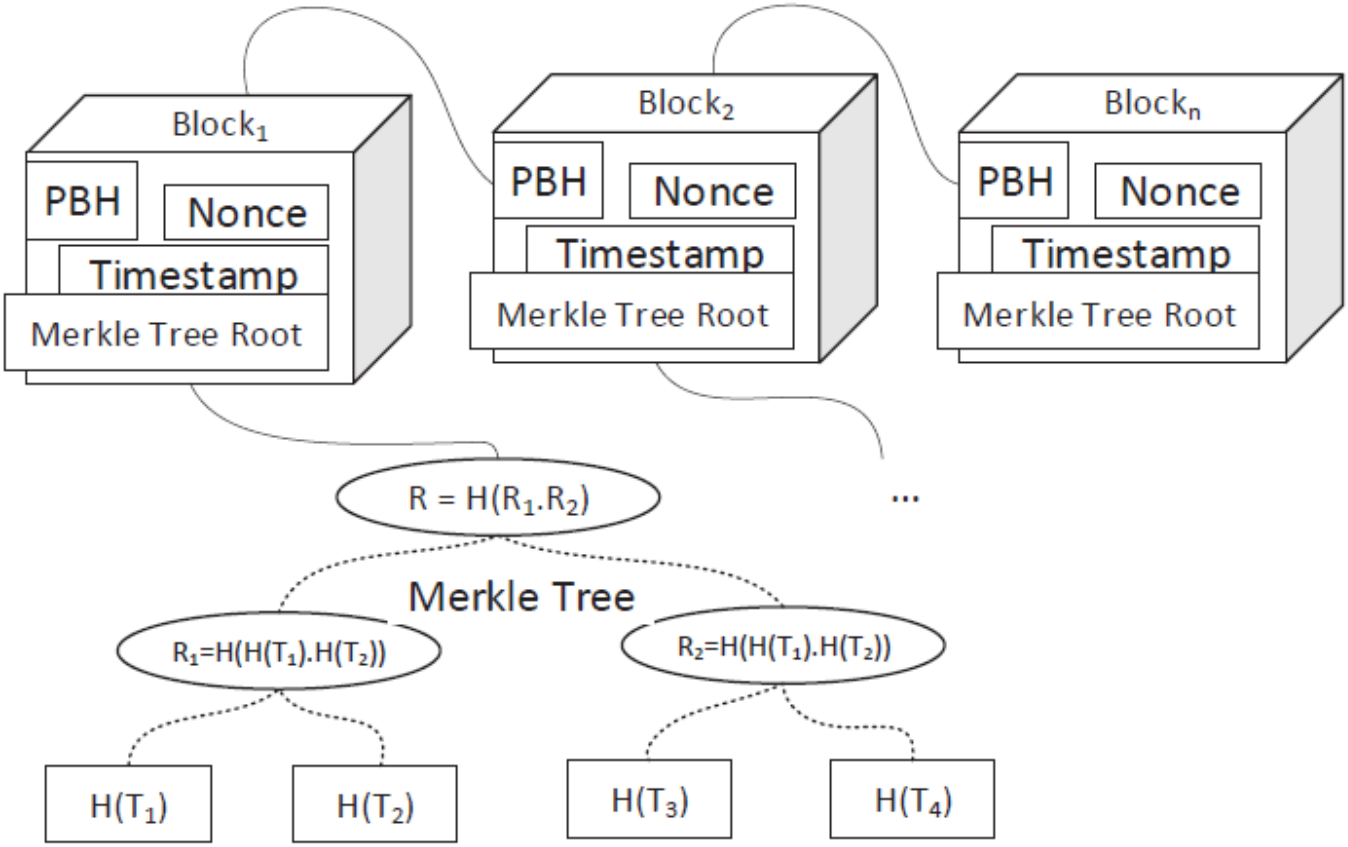


Figure 2

A typical Blockchain

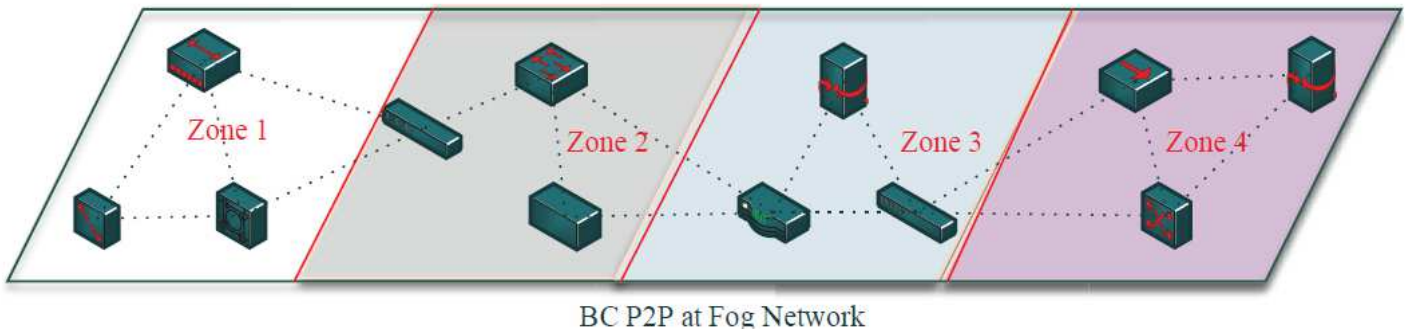


Figure 3

Different zone of a Fog based Blockchain network

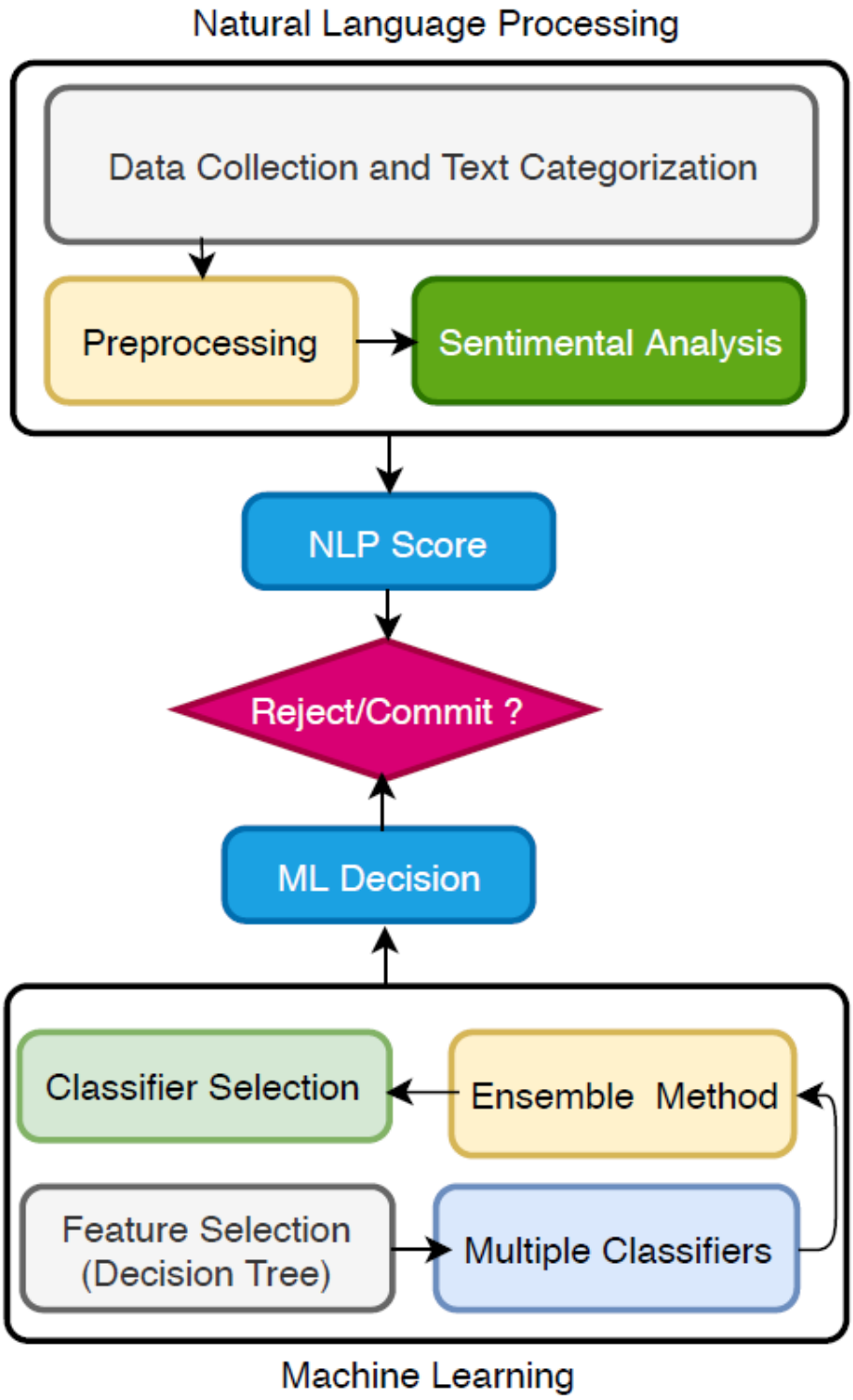


Figure 4

Cyberbullying detection module at fog layer

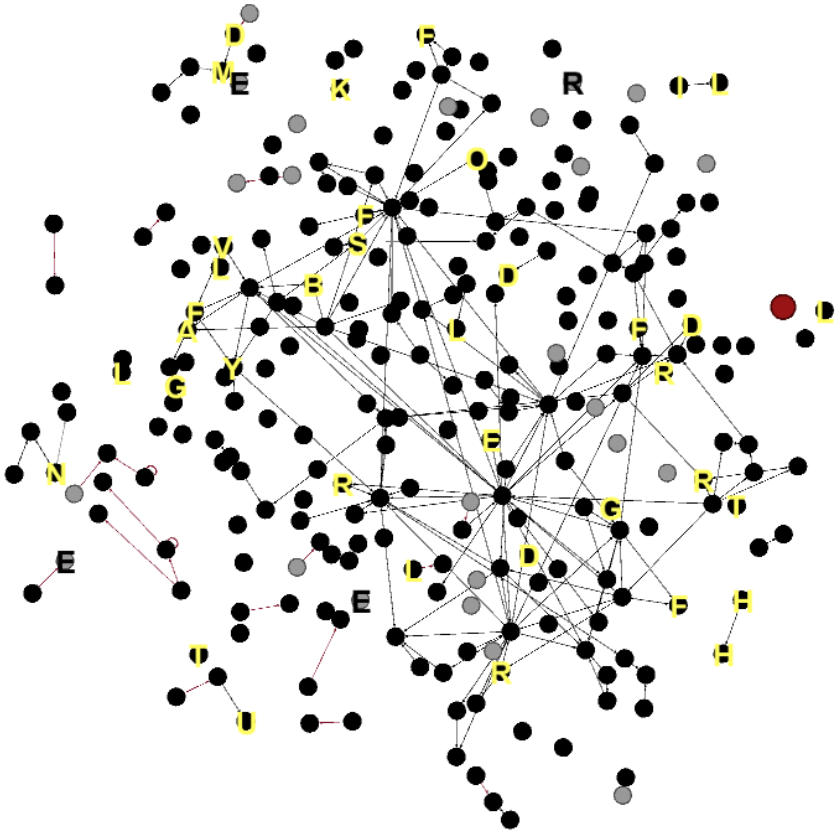


Figure 5

Visualization of proposed Blockchain network

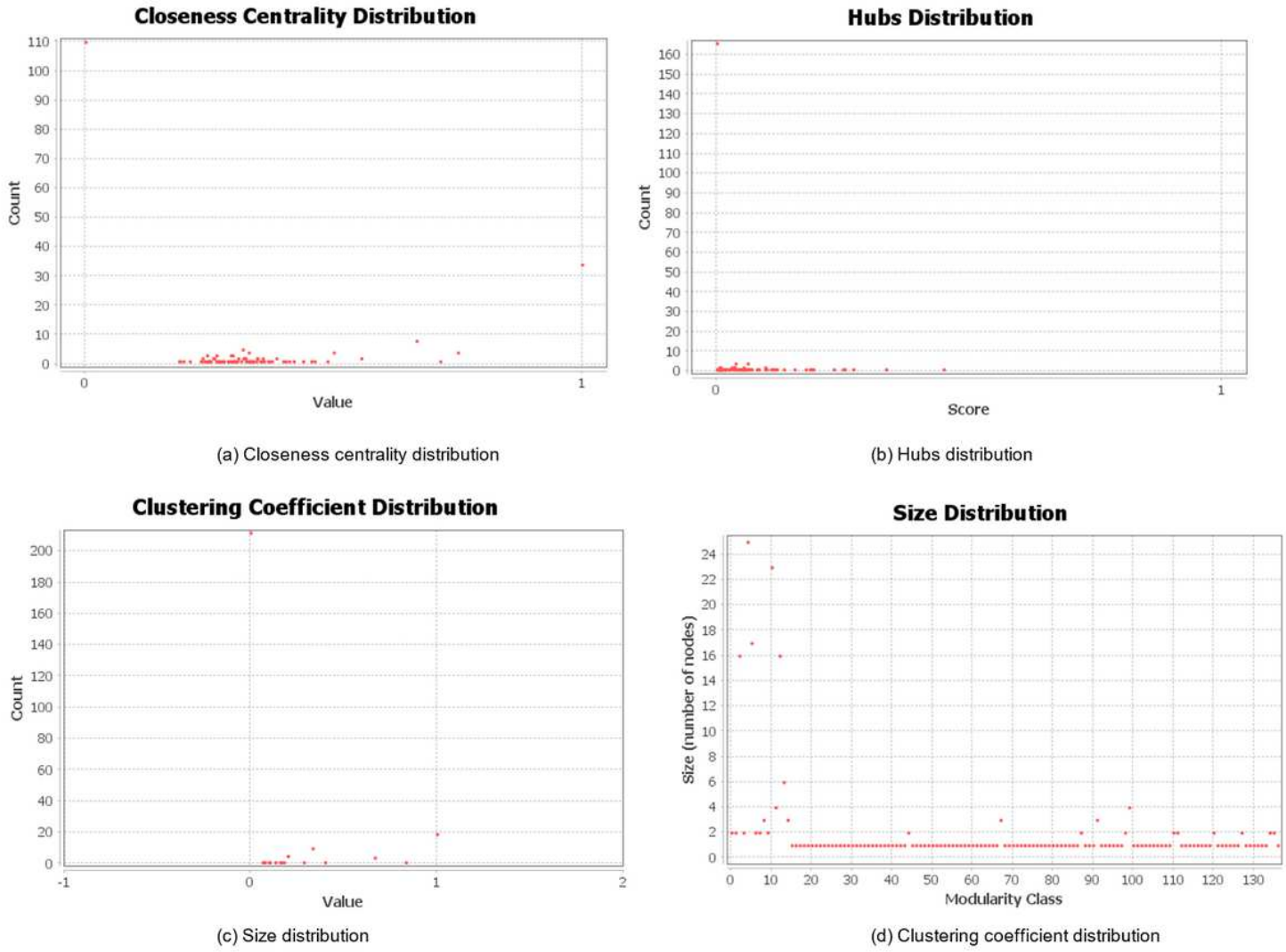


Figure 6

Distribution graph of the proposed network

```
def mine():

    if len(blockchain.transactions):
        msg = blockchain.transactions[0]['value']
        badWord = blockchain.chain
        isCyberbullying = detectCyberbulling(badWord, msg)
        if isCyberbullying:
           sender.setWorningMessage("Cyberbulling Detected")
           last_block = blacklist.chain[-1]
           nonce = blacklist.proof_of_work()
           previous_hash = blacklist.hash(last_block)
            blockchain.create_block(nonce, previous_hash)
      else:
           last_block = blockchain.chain[-1]
           nonce = blockchain.proof_of_work()
           previous_hash = blockchain.hash(last_block)
           blockchain.submit_transaction(sender_address=MINING_SENDER,
           recipient_address=blockchain.node_id, value=MINING_REWARD,
           signature="")
           previous_hash = blockchain.hash(last_block)
           block = blockchain.create_block(nonce, previous_hash)
```

Figure 7

A sample code from the implementation

Input:

Sender & Recipient Address, sender private key, encoded message.

Output:

Set Waring to sender Address & discard to send message if there have any cyberbullying related information. Else send message to recipient address.

1. Select `random_node` from n fog nodes as miners
2. `decoded_message = message_decoder (private_key, encoded_message)`
3. **Go to** step 4 to 6 for cyberbullying detection
4. Get all cyberbullying related keyword as `black_list` from cloud database
5. Apply machine learning model on `decoded_message` to detect cyberbullying based on `black_list`
6. **If** `(isCyberbullying)` then execute line 7 to 9:
7. Store new `cyberbullying_word` to `cloud_database` for further processing
8. `setWaring(sender address, "waring_message")`
9. **Exit**
10. **Else** execute following 11 to 16 line

```
nonce = blockchain. proof_of_work()
blockchain.submit_transaction (sender_address, recipient
_address, message, signature)
previous_hash = blockchain.hash(last_block)
block = blockchain.create_block(nonce, previous_hash)
```

Set `block` to recipient address

17. **Exit**

Figure 8

Pseudocode of Cyberbullying Detection and Prevention Model

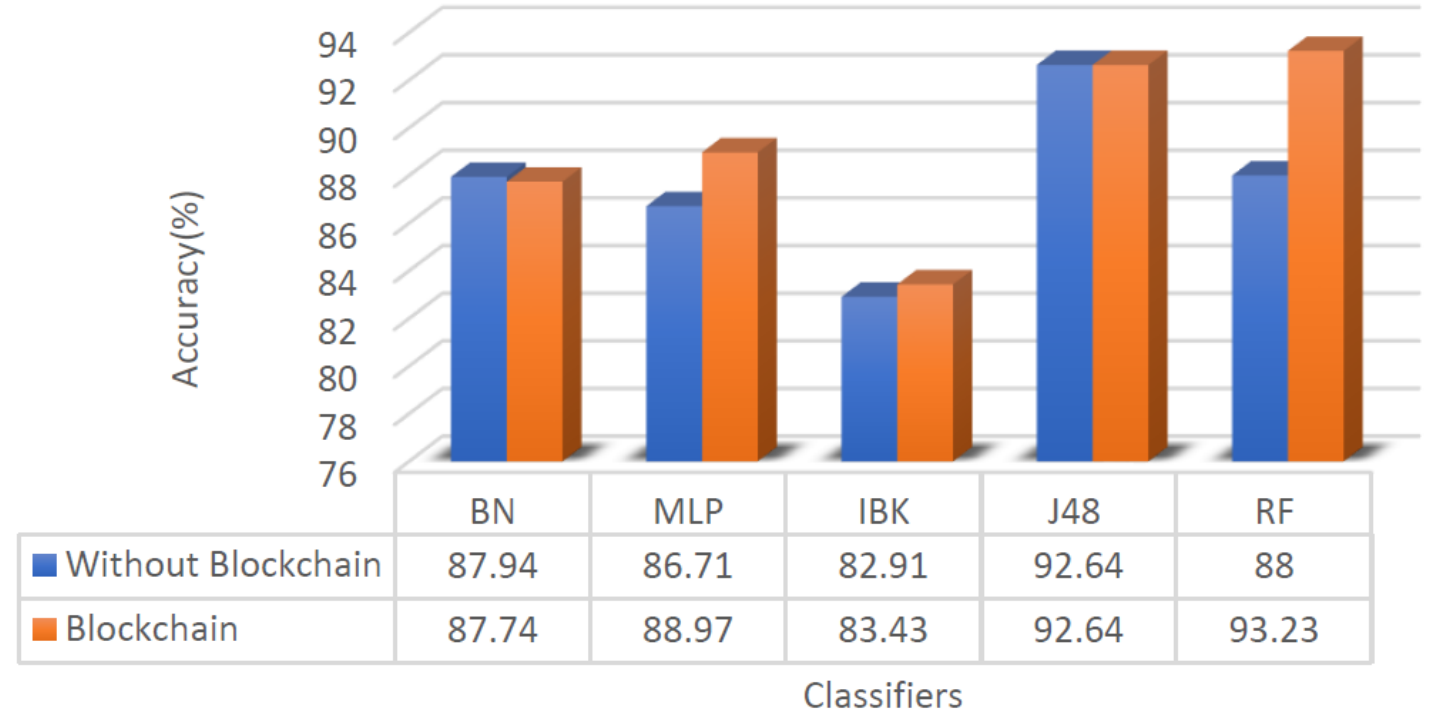


Figure 9

Accuracy of five classiers to detect cyberbullying