IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006044 |

A Review on Ontology Based E-Crime Detection

GOWTHAMI R, Mrs. VEENA B

MCA Student, UBDTCE Davanagere, Karnataka, India

Faculty, Department of MCA, UBDTCE Davanagere, Karnataka, India

ABSTRACT: Innumerable terror and suspicious messages are sent through Instant Messengers (IM) and Social Networking Sites (SNS) which are untraced, leading to hindrance for network communications and cyber security. We propose a Framework that discover and predict such messages that are sent using IM or SNS like *Facebook, Twitter, LinkedIn*, and others. Further, these instant messages are put under surveillance that identifies the type of suspected cyber threat activity by culprit along with their personal details. Framework is developed using Ontology based Information Extraction technique (OBIE), Association rule mining (ARM) a data mining technique with set of pre-defined Knowledge-based rules (logical), for decision making process that are learned from domain experts and past learning experiences of suspicious dataset like *GTD* (Global Terrorist Database). The experimental results obtained will aid to take prompt decision for eradicating cybercrimes.

KEYWORDS: Instant Messengers(IM); Social Networking Sites(SNS); Ontology based Information Extraction; Association Rule Mining(ARM); Knowledge based rules.

I. INTRODUCTION

Internet evolutions led to the growth of innumerable cybercrimes. Criminals adapted to send suspicious messages via mobile phones, Instant Messengers and Social Networking Sites, which is difficult to trace their criminal activities dynamically. The E-crime department must be improvised with the development of technology to find criminals. Many of the Instant Messaging Systems (IMS) developed restricted their limit for sending messages, video and audio conferencing. They are not well equipped to detect online suspicious messages.

Cybercrime activities are increasing day by day. The CIA, FBI and other federal agencies are actively collecting domestic and foreign intelligence information to prevent future cyber attacks. Recently the Internet Crime Complaint Center (IC3) released the report in 2012 of cybercrimes, with the latest data and trends of online criminal activity [1].We surveyed various architectures of Mobile Phones, Instant messengers and Social Networking sites [2, 3]. These studies helped us to develop a new Framework. WordNet, is a lexical database, contains a huge amount of information consisting of (155287 words organized in over 117000 Synesis for a total of 207000-word sense pairs) words that is useful for our study for scanning and filtering the text messages stored in TDB (Text Database) [4]. WordNet is used as features for classification of words from unstructured text. Similarly, WordNet Ontology based on information extraction technique is discussed in [5]. Our Contribution includes improving the existing IMS using data mining technique of Associative rules [6], Ontology based information retrieval technique, which is guided with pre-defined Knowledge based rules and ARM. Early detection of suspicious messages from instant messaging systems (Mobile Phone, IM and SNS) is possible with our proposed Framework to identify and predict the type of cyber threat activity and trace the criminal details.

This Section gives an overview of cybercrimes performed in IMS and deficiencies exist. The remainder of this paper is organized as follows:

- In Section II, we reviewed recent research advances in identifying criminals from cyberspace.
- The Section III illustrates the operational phases and its implementation of our proposed Framework to validate these instant messages sent are suspicious or not and steps for tracing the culprits.
- The experimental results are shown in Section IV, when tested with dataset collected from Global Terrorism Database (GTD) [25].
- Finally, Section V concludes the paper with an outlook towards future research directions for adding the features of our proposed Framework to current IMS.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006044 |

II. BACKGROUND

In this Section we explore the operational phases of Framework, shown in Fig. The Suspicious Pattern Detection (SPD) algorithm initiate the steps to capture the instant messages that are sent between the clients/users and stores them into database for identifying suspicious messages using Ontology based Information Extraction (OBIE) technique. The following steps are performed by the system:

Step I: Sign up and Login

Firstly, User has to sign up to the Instant Messaging Application by his username and password. User has to log in into application by inserting user name and password.

Step II: Analyze Input Message

System read the message from the input string.

Step III: Filtering of Message

Now the filtering of message is done by the following process: firstly, it checks for the suspicious words from the dataset, and then it checks the messages by NLP, and removes the non-action words and precedes the message.

Step IV: Ontology

After this, also check the type of words found in ontology, here words are comparing from safe ontology to suspicious ontology. If the safe counts are less than suspicious counts then it will consider being suspicious ontology.

Step V: User's messaging Behavior

Now check if the user has already sent suspicious message to the user. If this user has sent suspicious messages to more than 1 users previously. Then user will be in the list of suspicious users.

Step VI. Checking for the number of malicious activities

If the numbers of suspicious activities are more than 2 then mark it as a malicious user and block that user. Otherwise using e-monitoring and data Ming techniques to store suspicious word in the database.

Step VII: Detected Suspicious Message

Detected suspicious word are store int the data base. The e-crime departments are trace the location of users details and take actions.

III. METHADOLOGY

The Fig shows the overall system structure of the proposed framework. As shown in the figure, data collection system follows ordinary web chat application's features forgroup chat along with encryption techniques to facilitate security. Suspicious word detectionsystem focuses on detecting suspicious words with the help of OBIE and data miningtechniques. Short word, code word and suspicious word databases are maintained. Finally,offender's details are displayed with the help of user's personal database.

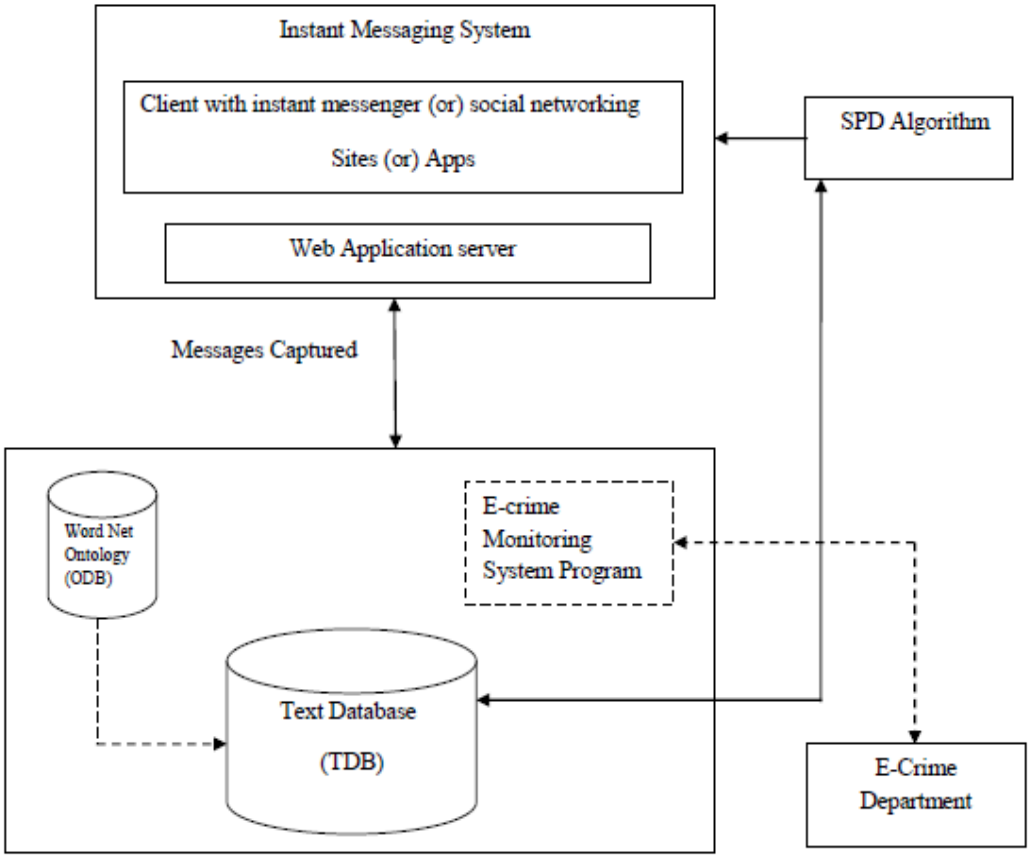
Framework of proposed system aids the E-crime department to identify suspicious
words from cyber messages and trace the suspected culprits. Currently existing InstantMessengers and Social Networking Sites lack these features of capturing significantsuspicious patterns of threat activity from dynamic messages and find relationships amongpeople, places and things during online chat, as offenders have adapted to it. The testbed isproven to be useful, for monitoring terror and suspicious crimes in cyberspace, whichprovides national and international security. If the proposed framework integrated withexisting IM and SNS at server-side for surveillance will change the world of cyberspace torest in peace without cybercrime.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006044 |



System framework

IV. RESULT

1. Evaluation method for datasets

We used Precision metric [20] to evaluate our Framework. The extracted suspicious words efficacy is based on two factors, the number of actual words available in the pre- defined database, to that of the number of words from user generated extracted testbeds.

$$\text{Precision } (P) = \frac{\text{Correctly Extracted}}{\text{Total Extracted Correctly}}$$

$$\text{Recall}(R) = \frac{\text{Correctly Extracted}}{\text{Total no. of Possible Words}}$$

2. Preparation of datasets and results obtained

The Terrorist Atta ck (Domain) dataset is taken from Global Terrorism Database (GTD) which has recorded information on terrorist events around the world sin ce 1970 to till date. The complete representation related to terrorist attacks is found using *CODEBOOK [21]*. We obtained dataset using brainstorming session from domain experts using GTD that consists of 59787 rows, size of 30MB, and 7 columns taken out of 99 column s, named it as User Generated Content (UGC) i.e. UGC-testbed-1 and tested with our Framework.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006044 |

The outputs obtained are shown in Table II.

Terms	FrameworkOut
TotalExtractedCorrectly	**1779**
CorrectlyExtracted	**1703**
TotalPossiblewordsextracted	**1732**
Precision	**95.72**
Recall	**98.32**

Dataset used are manually created by brain storming session from domain experts using GTD, as we could not able to get re al suspicious contents that are stored in hi story from IM and SNS, due to authorization restriction

3. Comparison of our framework with existing IMS

Currently none of Instant Messengers has the ability to detect suspicious messages during online chat. The features based on which our Framework is compared with IM(ICM) are shown in Table III.

Table III. Comparison of our Framework with (IM, SNS & Apps

Features	IM	ProposedFramework
CyberthreatActivity Detection	StaticDetection (timeconsumed)	DynamicDetection
Report Generationfor E-crimedepartment	NoReport	Report with details (Email-id,PhoneNo.,etc.)
Ontologysupport	No	Yes
Dynamic LocationMapping based onISP andIPaddress	No	Yes (usingR2DWrapper)
Efficiency	VeryGood	Moderate (as onlinemessages aremonitored&stored)
Database&Data Miningsupport	No	Yes
SystemArchitecture	EasytoDesign	Complextodesign
Code words and shortwords	Notdetected	Detected
Encryption anddecryption	No	Yes

V. CONCLUSION

Framework aids the E-crime department to identify suspicious words from cyber messages and trace the suspected culprits. Currently existing Instant Messengers lack these features of capturing significant suspicious patterns of threat activity from dynamic messages and find relationships among people, places and things during online chat, as criminals have adapted to it [10]. The User Generated Content (UGC) testbed is proven to be useful, for monitoring terror and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006044 |

suspicious crimes in cyberspace which provides national and international security. We used simple English terms like kill, murder, etc. But, in practical scenarios these words are in specific coding language, for example "picnic" is used instead of *"*kill*"*.

Issues and challenges of our Framework are:

1. The suspicious words sent in Steganography techniques are not detected and hence neglected as ignore words.
2. Support for Multilingual languages to be included [24]. The media may also actively participate in transmitting messages to terrorists and criminals indirectly, via newspapers and TV channels (Text, Audio and Video) unknowingly [25].
3. Integration with HADOOP to solve Big Data problems.

   If the proposed Framework integrated with existing IM at Server-side, for surveillance will change the world of cyberspace to rest in peace without cyber crime [10].

REFERENCES

[1] http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet.
[2] http://en.wikipedia.org/wiki/Computer_crime
[3] "FrameworkforSurveillanceofInstantMessagesinInstantmessengersandSocialnetworkingsitesusingDataMiningandOntology" by Mohammed Mahmood Ali, Mohammed MahmoodAli, LakshmiRajamani.
[4] http://www.interpol.int/Crime-areas/Cybercrime/Cybercrime.
[5] 3GPP2partners,"ShortMessageServiceoverIMS:3rdGenerationPartnershipProject2,"developedunder3GPP2,publishedin2007.
[6] Daya C. Wimalasuriya, and Dejing Dou, "Ontology-Based InformationExtraction:AnIntroductionandaSurveyofCurrentApproaches,"JournalofInformationScience, Volume36, No.3,pp.306-323,2010.
[7] M. Mahmood Ali, and L. Rajamani, "Framework for surveillance ofinstantmessages,"publishedbyindersciencein IJITST,vol.5,2013.
[8] Michael Robertson, Yin Pan, and Bo Yuan, "A Social Approach toSecurity:UsingSocialNetworkstoHelpDetectMaliciousWebContent,"publishedbyIEEEin2010.
[9] (2012). [Online]. Available: http://www.digitaltrends.com/social-media/Facebook-scans-chats-and-comments-looking-for-criminal-behavior/

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details