Role of Data Mining in solving Cyber Crime Problem in India

Shruti Bajaj[1], Dr. Rajesh Kumar Singh[2]
PHD STUDENT[1], PRINCIPAL[2]
[1]PUNJAB TECHNICAL UNIVERSITY KAPURTHALA, [2]SUSCET TANGORI MOHALI

Abstract: Data mining is the process of posing queries and extracting patterns, often previously unknown from large quantities of data using pattern matching or other reasoning techniques. Cyber security is the area that deals with protecting from cybercrime. The paper provides an overview of data mining techniques and cybercrime and discusses developments in applying data mining for cyber security. Cyber security is concerned with protecting computer and network systems from corruption due to malicious software including Trojan horses and viruses' .Cyber security standards are security standards which enable organizations to practice safe security techniques to minimize the number of successful cyber security attacks. Data mining can be used to model crime detection problems. Cyber Crimes are a social nuisance and cost our society dearly in several ways. They are IT-based criminal offense. They are a new class of crimes rapidly increasing due to extensive use of Internet and I.T. enabled services. Here we look how data mining approach helps to detect cybercrime patterns and speed up the process of solving it. The purpose of this paper is to suggest ways by which data mining can analyze large volume of data using as one of the ways of getting active solutions for cybercrime investigation in India.

Keywords-- Data Mining; Cyber Crime; Cyber Security; Cyber Attacks.

I. INTRODUCTION

With the rapid advancement of information discovery techniques data mining continues to play an important role in cyber security. Cyber security is a complex issue that cuts across multiple domains and calls for multi-dimensional, multilayered initiatives and responses. It has proved a challenge for governments because different domains are typically administered through ministries and departments. Data mining is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data and it is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining plays an important role in terms of prediction and analysis. Data mining and web mining may be used to detect and possibly prevent cyber-attacks and cybercrime. Cybercrime is an illegal act committed using a computer network. It is a subset of computer crime and involves a computer and a network. It is ubiquitous and has become a major security issue. The computer may have been used in the commission of a crime, or it may be the target [1]. They are offences that are committed against individuals or groups of individuals with a criminal motive to intentionally harm the reputation of the victim or cause physical or mental harm, or loss, to the victim directly or indirectly, using modern telecommunication networks such as Internet (Chat rooms, emails, notice boards and groups) and mobile phones (SMS/MMS)"[2].

II. CYBER CRIME IN INDIA

Cybercrime is one of the dangerous factors for any country. It is impossible to find a country which has a crime- free society. In the Indian scenario with e commerce becoming popular in the last few years cybercrimes is a term used to broadly describe criminal activity in which computers or computer networks are a tool, a target, or a place of criminal activity and include everything from electronic cracking to denial of service attack. It is also used to include traditional crimes in which computers or networks are used to enable the illicit activity.

A key finding of the Economic Crime Survey 2006 was that a typical perpetrator of economic crime in India was male (almost100%), a graduate or undergraduate and 31-50 years of age. Further, one third of the frauds were from insiders and over 37% of them were in senior managerial positions [3]. The present society is filled with various kinds of cybercrimes. This paper presents a brief overview of all about cyber criminals and cybercrime with its evolution, types, case study, preventive majors and the department working to combat this crime using data mining.

This paper is organized as follows. Section 1 gives the introduction about data mining, cybercrime and cyber security. Section 2 describes cybercrime in India. Section 3 explains the role of data mining in cyber security in India. Concluding remarks are given in Section 4.

With increasing internet penetration, cybercrimes have also increased in the last few years. Between 2011 and 2015, the number of cybercrimes registered in the country has gone up 5 times. With increasing mobile and internet penetration in the country, cybercrimes have also increased proportionately. Between 2011 and 2015, more than 32000 cybercrimes were reported across the country. More than 24000 of these cases are registered under the IT Act and the remaining under the various sections of IPC and other State Level Legislations (SLL). Fig no 1 shows the cases registered under IT Act and IPC Act. Cyber Crimes in India are registered under two different acts, the IT Act and the Indian Penal Code (IPC). The cases registered under the IT Act include:

- Tampering computer source documents (Section 65 IT Act)
- Loss /damage to computer resource/utility (Section 66 (1) IT Act)
- Hacking (Section 66 (2) IT Act.
- Obscene publication/transmission in electronic form (Section 67 IT Act)
- Failure of compliance/orders of Certifying Authority (Section 68 IT Act)

Fig no 2 shows Cyber Crime in India and the number of cases registered under IT Act.

On the other hand, cases are also registered under the IPC and those include:

- Offences by/against Public Servant (Section 167, 172, 173, 175 IPC)
- False electronic evidence (Section 193 IPC)
- Destruction of electronic evidence (Section 204, 477 IPC)
- Forgery (Section 463, 465, 466, 468, 469, 471, 474, 476, 477A IPC)
- Criminal Breach of Trust (Section 405, 406, 408, 409 IPC)
- Counterfeiting Property Mark (Section 482, 183, 483, 484, 485 IPC). Fig no 3 shows Cyber Crime in India and the number of cases registered under IT Act.

Year	IT Act		IPC Act	
	Cases Registered	Persons Arrested	Cases Registered	Persons Arrested
2011	1791	1184	422	446
2012	2876	1522	601	549
2013	4356	2098	1337	1203
2014	7201	4246	2272	1224
2015	8045	5102	3422	2867
Total	**24269**	**14152**	**8054**	**6289**

Fig no: 1. Number of cases registered and persons arrested under IT Act and IPC Act.

The numbers of cases registered under the IT Act and IPC Act have been growing continuously. The cases registered under the IT act grew by more than 350% from 2011 to 2015. There was almost a 70% increase in the number of cybercrimes under the IT act between 2013 and 2014. The cases registered under the IPC increased by more than 7 times during the period between 2011 and 2015. Similar trend is observed in the number of persons arrested. The government also acknowledges the increase in the number of such crimes and that the introduction of technologies, devices including smart phones and complex applications, and rise in usage of cyber space for businesses has resulted in such an increase [4] .

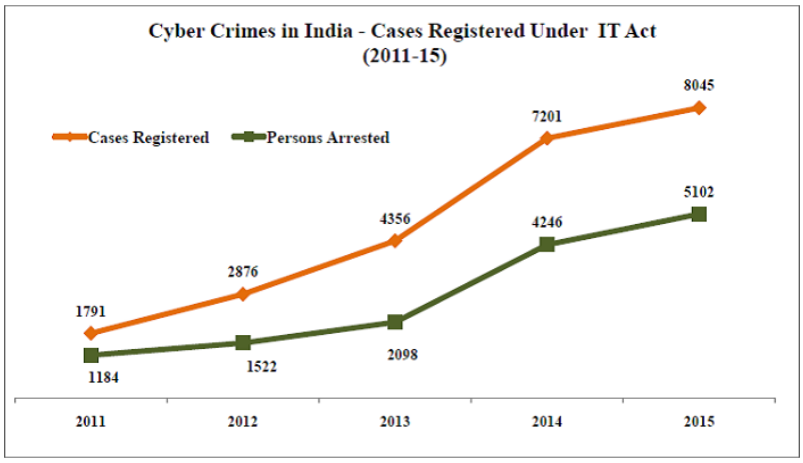

Fig no: 2. Number of cases registered under IT Act [4].

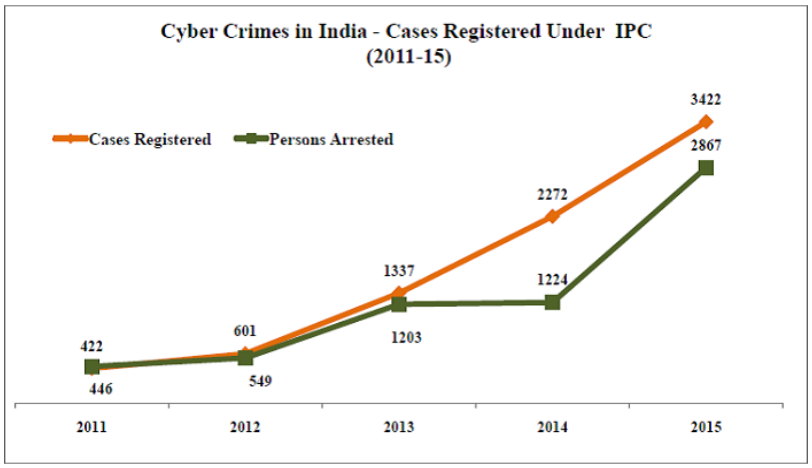

Fig no: 3. Number of cases registered under IPC Act [4]

Maharashtra & Uttar Pradesh on the top

The list of states with the highest incidence of cybercrime for the period 2011 to 2015 throws no surprises. Maharashtra tops the list with more than 5900 cases in the 5 years followed by Uttar Pradesh with close to 5000. Karnataka is third with more than 3500 cases. The top states in this list are the ones with a greater internet subscriber base. The bottom 10 are relatively smaller states with lower population and lower internet penetration [4]. Summary of Cybercrimes in India is represented in Fig no 4.

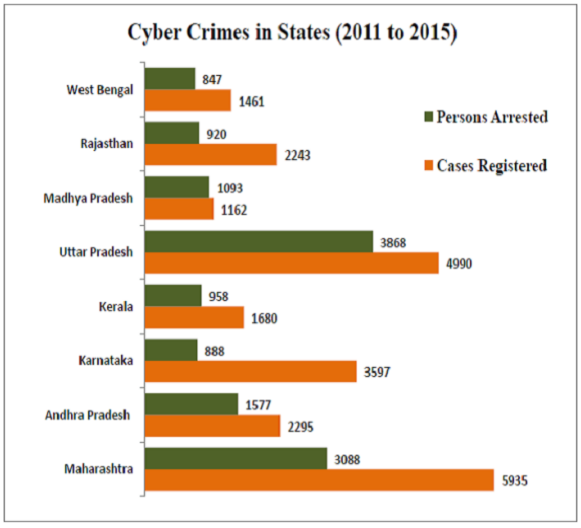

Fig no 4: Summary of Cyber Crimes in India [4]

Mumbai: Cybercrimes have steadily increased in Maharashtra in the last three years, doubling in 2014 over the previous year. Most of the 9,322 cases registered under the IT Act and IPC Act during 2014 were registered in Maharashtra.

New Delhi: Cyber Crime cases in the country registered under IT Act surged nearly 300 percent between 2011 and 2014. The study revealed that in the past, the attacks have been mostly initiated from countries like US, Turkey, China, Brazil, Pakistan, Algeria, Europe and UAE, with growing adoption of internet and smartphones India has emerged one of the primary targets among criminals.

Hyderabad: An email allegedly from India's central bank, asking to secure their bank account details with the RBI is fake, and an attempt by new-age fraudsters to con people into giving away bank account details and lose hard-earned money, security experts said. The email says RBI has launched a new security system, asking users to click on a link to open a page with list of banks in place. Once anyone chooses a particular bank, it asks for all net banking details, including card numbers and the secret three digit CVV number, among others. The incidence of Cyber Crime cases during 2014 is shown in Fig no 5. Also Fig no 6.1 and 6.2 shows the state wise analysis of Cyber Crime in India.

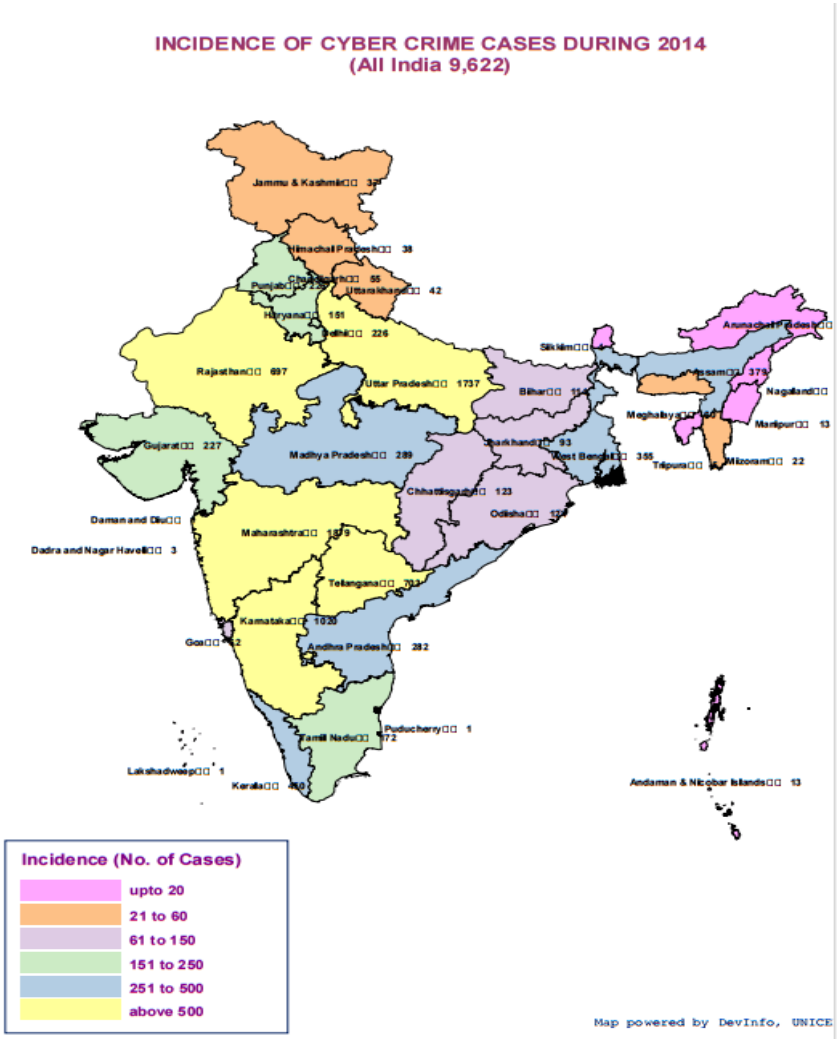

Fig no 5: Incidence of Cyber Crime cases during 2014.

State/UT	IT Act		IPC	
	Cases Registered	Persons Arrested	Cases Registered	Persons Arrested
Maharashtra	1458	976	403	345
Andhra Pradesh (Including Telangana)	1413	708	64	111
Karnataka	1076	194	54	29
Kerala	845	437	95	47
Uttar Pradesh	678	518	367	428
Madhya Pradesh	514	414	128	63
Rajasthan	508	335	89	42
West Bengal	449	142	259	206
Punjab	277	247	36	33
Delhi	257	76	76	44

Fig no 6.1: State wise analysis of Cyber Crime in India [3].

Bottom 10

State/UT	IT Act		IPC	
	Cases Registered	Persons Arrested	Cases Registered	Persons Arrested
Nagaland	0	0	0	0
Lakshadweep	0	0	0	0
Manipur	1	0	0	0
Daman & Diu	2	3	0	0
Mizoram	3	1	0	0
Sikkim	3	1	1	0
Dadra & Nagar Haveli	3	1	3	1
Puducherry	11	7	0	0
Andaman & Nicobar Islands	20	3	0	0
Tripura	28	23	0	0

Fig no 6.2: State wise analysis of Cyber Crime in India [3].

Considering the increasing trends of the crimes the Bureau has collected a comprehensive data on cybercrimes in 2014 using revised Performa of 'Cybercrime in India'. The IT Act, 2000, specifies the acts which are punishable.

In India, cybercrime cases registered are less compared to the US, Europe, etc. The Internet Crime Complaint Center (IC3) 2006 ranks the US (60.9%) as first among the nations in hosting perpetrators followed by the UK (15.9%). Many countries, including India, have established Computer Emergency Response Teams (CERTs) with an objective to coordinate and respond during major security incidents/events. These organizations identify and address existing and potential threats and vulnerabilities in the system and coordinate with stakeholders to address these threats [3].

III. ROLE OF DATA MINING IN CYBER SECURITY

Data mining for cyber security applications. For example, anomaly detection techniques could be used to detect unusual patterns and behaviors. Link analysis may be used to trace the viruses to the perpetrators. Classification may be used to group various cyber-attacks and then use the profiles to detect an attack when it occurs. Prediction may be used to determine potential future attacks depending in a way on information learnt about terrorists through email and phone conversations.

Anomaly Detection: Anomaly detection approaches build models of normal data and detect deviations from the normal model in observed data. Anomaly detection applied to intrusion detection and computer security has been an active area of research since it was originally proposed by Denning. They have the advantage that they can detect emerging threats and attacks as deviations from normal usage. Also, unlike misuse detection schemes anomaly detection algorithms do not require an explicitly labeled training data set, which is very desirable, as labeled data is difficult to obtain in a real network setting [5].

Profiling Network Traffic Using Clustering: Clustering is a widely used data mining technique which groups similar items, to obtain meaningful groups/clusters of data items in a data set. These clusters represent the dominant modes of behavior of the data objects determined using a similarity measure. A data analyst can get a high level understanding of the characteristics of the data set by analyzing the clusters. Clustering provides an effective solution to discover the expected and unexpected modes of behavior and to obtain a high level understanding of the network traffic.

Scan Detection: A precursor too many attacks on networks is often a reconnaissance operation, more commonly referred to as a scan. Identifying what attackers are scanning for can alert a system administrator or security analyst to what services or types of computers are being targeted. Knowing what services are being targeted before an attack allows an administrator to take preventative measures to protect the resources e.g. installing patches, firewalling services from the outside, or removing services on machines which do not need to be running them.

Methodology: Currently solution is a batch-mode implementation that analyzes data in windows of 20 minutes. For each 20-minute observation period, we transform the Net Flow data into a summary data set. With our focus on incoming scans, each new summary record corresponds to a potential scanner that is pair of external source IP and destination port (SIDP). For each SIDP, the summary record contains a set of features constructed from the raw Net flows available during the observation window. Observation window size of 20 minutes is somewhat arbitrary. It needs to be large enough to generate features that have reliable values, but short enough so that the construction of summary records does not take too much time or memory.

Cyber-terrorism, Insider Threats, and External Attacks: Cyber-terrorism is one of the major terrorist threats posed to our nation today. This threat is exacerbated by the vast quantities of information now available electronically and on the web. Attacks on our computers, networks, databases and the Internet infra-structure could be devastating to businesses. It is estimated that cyber-terrorism could cause billions of dollars to businesses. A classic example is that of a banking information system. If terrorists attack such a system and deplete accounts of funds, then the bank could lose millions and perhaps billions of dollars. Even a simple power outage at work through some accident could cause several hours of productivity loss and as a result a major financial loss. Therefore it is critical that the information systems be secure.

Credit Card Fraud and Identity Theft: In the case of credit card fraud, an attacker obtains a person's credit card and uses it to make unauthorized purchases. By the time the owner of the card becomes aware of the fraud, it may be too late to reverse the damage or apprehend the culprit. A similar problem occurs with telephone calling cards. A more serious theft is identity theft. Here one assumes the identity of another person by acquiring key personal information such as social security number, and uses that information to carry out transactions under the other person's name.

Attacks on Critical Infrastructures: Attacks on critical infrastructures could cripple a nation and its economy. Infrastructure attacks include attacking the telecommunication lines, the electric, power, gas, reservoirs and water sup-plies, food supplies and other basic entities that are critical for the operation of a nation. Attacks on critical infrastructures could occur during any type of attack whether they are non-information related, information related or bio-terrorism attacks. For example, one could attack the software that runs the telecommunications industry and close down all the telecommunication lines. Similarly, software that runs the power and gas supplies could be attacked. Attacks could also occur through bombs and explosives. That is, the telecommunication lines could be physically attacked. Attacking transportation lines such as highways and railway tracks are also attacks on infrastructures. Infrastructures could also be attacked by natural disaster such as hurricanes and earth quakes. The purpose is to examine data mining and related data management technologies to detect and prevent such infrastructure attacks [5].

IV. CONCLUSION

Data mining has many applications in security including in national security as well as in cyber security. The threats to national security include attacking buildings and destroying critical infrastructures such as power grids and telecommunication systems. Data mining techniques are being used to identify suspicious individuals and groups, and to discover which individuals and groups are capable of carrying out terrorist activities. Cybercrime is all about the crimes in which communication channel and communication device has been used directly or indirectly as a medium whether it is a Laptop, Desktop, PDA, Mobile phones, Watches, Vehicles. It

is harder to detect and hardest to stop once occurred causing a long term negative impact on victims. Cyber security is concerned with protecting computer and network systems from corruption due to malicious software including Trojan horses and viruses. In this paper we focused mainly on data mining for cyber security applications in India. We also concluded role of data mining to observe confidential data to preserve cyber security. For cyber security and national security data mining is a very wide & active area to research. Data mining helps users to make all kinds of correlations and leads to privacy concerns. Based on the data mining of sample data, this paper provides a reasonable method which can provide a future development direction of this field research.

V. REFERENCES

[1]. [Online].Available:https://www.idsa.in/system/files/book/book_indiacybersecurity.pdf.
[2]. [Online].Available:https://en.wikipedia.org/wiki/Cybercrime.
[3]. [Online].Availaible:http://www.serc.iisc.ernet.in/graduation-theses/CizaThomas-PhD-Thesis.pdf.
[4]. [Online].Availaible:https://factly.in/cyber-crimes-in-india-which-state-tops-the-chart/.
[5]. [Online].Availaible:http://jems.net.in/Roll%20of%20Data%20Mining%20in%20Cyber%20Security.pdf.