Physical and Cyber Crime Detection using Digital Forensic Approach: A Complete Digital Forensic Tool

Abstract—Criminalization may be a general development that has significantly extended in previous few years. In order, to create the activity of the work businesses easy, use of technology is important. Crime investigation analysis is a section records in data mining plays a crucial role in terms of predicting and learning the criminals. In our paper, we’ve got planned an incorporated version for physical crime as well as cybercrime analysis. Our approach uses data mining techniques for crime detection and criminal identity for physical crimes and digitized forensic tools (DFT) for evaluating cybercrimes. The presented tool named as Comparative Digital Forensic Process tool (CDFPT) is entirely based on digital forensic model and its stages named as Comparative Digital Forensic Process Model (CDFPM). The primary step includes accepting the case details, categorizing the crime case as physical crime or cybercrime and sooner or later storing the data in particular databases. For physical crime analysis we’ve used k-means approach cluster set of rules to make crime clusters. The k-means method effects are a lot advantageous by the utilization of GMAPI generation. This provides advanced and consumer-friendly visual-aid to k-means approach for tracing the region of the crime. we have applied KNN for criminal identification with the help of observing beyond crimes and finding similar ones that suit this crime, if no past document is discovered then the new crime sample are introduced to the crime data-set. With the advancements of web, the network form has become much more complicated and attacking methods are further more than that as well. For crime analysis we’re detecting the attacks executed on host system through an outsider the usage of assorted digitized forensic tools to produce information security with the help of generating reports for an event which could need any investigation. Our digitized technique aids the development of the society by helping the investigation businesses to follow a custom-built investigative technique in crime analysis and criminal identification as opposed to manually looking the database to analyze criminal activities, and as a result facilitate them in combating crimes.

Keywords—Digitized Forensic Tools, K-means, KNN, GMAPI, cybercrime.

1. INTRODUCTION

Mankind has flourished in all of the spheres of lifestyles and is tough all the limits beyond the horizon, but at the equal time crime incidences are stoning up at a high rate and crime has now come to be a rampant act. Crime may be countrywide or worldwide but it’s always an offense in opposition to morality within the society. In latest years, many instances associated with murder, attacks, rapes, housebreaking, hacking, banking frauds, e-mail spamming and so on has emerged. Data till mid of 2016 has visualized that the crime index in all international locations is increasing and the protection index has decreased. Crime comes into being due to reasons like illiteracy, unemployment, over population, poverty. Many different factors like mindset of and character, upbringing, family heritage that can pressure an person to dedicate a crime. Because of the advancements in technology, electronic gadgets are available and reachable they are...
being misused with the aid of humans for their selfish motives. This has introduced the time period “Cyber Crime” into image. Physical and cyber-crimes are cropping up at special costs and special geographical locations.

A few well-known crimes are as follows:

a) Badaun gang-rape case in which two teenage cousin sisters, who have been allegedly gang-raped and hanged from a tree on 27 May additionally 2014.

b) An armed man shot more than one firearms on the target market inner Century movie theatre in Aurora, Colorado at the screening of “The dark Knight Rises”, killing 12 people.

c) In might also 2014, a collection of hackers controlled to steal or borrow non-public information of 233 million customers which blanketed names, electronic mail and postal deal with, smart phone variety etc. those crimes no longer simplest have an effect on the sufferer but additionally startles the country as a whole. for this reason, there is an urgency to combat the criminal incidences. The law enforcement companies are already chargeable for crime research from large amount of crime statistics. Many methodologies and strategies had been proposed to this point for crime investigation; however each of them has their personal set of benefits and sure intrinsic limitations. A number of the authors have targeted on detecting criminal styles or crook hotspots at the same time as a few used Gaussian aggregate models to pick out criminal which is only depending on the eye-witness. For cyber-crime investigation, some techniques followed have the usage of system and network analyzers to accumulate evidences and are expecting the assault using the association Mining. At the same time, a few papers have discussed the usage of frequent Mining algorithm to come across fraudulent e-mails.

This information monitors that they do not provide foolproof methods to address crime occurrences. Therefore, there may be a need to manufacture a consolidated and integrated machine for crime detection and criminal prediction in case of physical as well as cyber-crime.

A. Review of Literature

In India, Malathi’s [1] work mainly focuses on developing a tool that analyses crime. The first step involved removing inconsistencies, extracting data from the crime data set and missing values from data. In the next step the data is clustered based on the ‘type of crime’ using a new hybrid algorithm which is a fusion of DBSCAN and K-means. The results are also modified and cities are grouped in high, medium and low crime zones. The clusters thus obtained represent the trend of each type of crime in respective cities. The last module involves predicting future crime patterns using classification techniques. The C4.5 decision tree algorithm was used to predict the crime trend for next year.

Ewart and Oatley [2] describe about the OVER project which is basically a software system to assist in combating high volume crime and burglary. Initially four database tables containing details of crime, stolen property, offender and victim were created. The two techniques namely general crime techniques and offender decision support techniques are supported by this model. The general crime techniques include displaying data as per distance and time between crimes and displaying data according to features of crime. The offender techniques display the offender profiles and their operations in particular area. A Kohonen neural network for matching crimes against a list of offenders is used and Bayesian belief network is developed for predicting re-victimization of a particular dwelling. The main advantage of this model is that it has the ability to identify the criminal and predict re-victimization. The drawback of the system is that it mainly focuses on burglary and needs to be extended to develop a full Bayesian model.

We combined data mining and digital forensics to create a model consisting of four modules: File forensic modules and Memory forensic modules for predicting the attacks and Network Forensic modules. File system analyzer modules shows all the details of directories and files which are present in the particular drive. Memory Forensic Analysis process consist of the browser history, list of all running processes in the system, list of ports, the login session details. Network forensic module monitors the traffic and consists of Source IP, Destination IP, Source MAC, Destination MAC, Method, Protocol, Captured Time, Captured Length, Frame Type, Version and Destination Host. The system uses the algorithms of K-means and Apriori algorithms for predicting the type of attack. When an incident is reported, the project investigates it and report is stored in the database. The type of the attack is identified and the administrator is alarmed about similar attacks in future using crime data mining tool. Disadvantage is it cannot track the location of the attacker.

We implemented cyber Forensics using Sequence Mining algorithm, by comparing it with association rule mining algorithm parameters. The project deals with investigating on fraud emails. If any email contents have been changed by attacker and then it is send to recipient. This is can be determined by the algorithm generated by the author called Fraud Detection algorithm to find out fraud email in inbox. When attacker changed the contents of email, then size of file changed, and this in turn changed the sequence of mails in inbox. For cyber Forensics investigation, GSP (Generalized Sequential pattern) sequence mining algorithm is used to rearrange emails and to find out which email is fraud and what changes have been made by attacker in that particular email. Sequence mining problems are mostly solved by the algorithms which are based on Apriori algorithm of association rules. Disadvantage is that this work only deals with detecting the Fraud emails from an inbox but cannot track the attacker involved in this fraud.

We used data mining techniques along with forensic and social mining, image processing. The eye witness gives the features and using the General Gaussian Mixture model, the features are mapped with that of the criminal who have committed the crime. It maps the criminal with the crime. Here, the eyewitness is considered important. If the similar criminal is not found then a new report is generated. Process starts with data which is collected from various police
stations. Database is created with the help of eyewitness and the clues obtained from the crime spot. Clues which are collected are used to identify the criminals; they are stored in the database. Crime variables such as type of crime, way of killings, criminal psychological behavior, time, modus operandi are used to analyze the dataset to identify the criminal. In this paper, k-means clustering algorithm is used to generate the clusters of crime using the attributes, crime variables etc. In this case, four clusters are considered i.e. robbery, kidnap, murder, riot etc. So whenever a crime takes place, the relevant cluster is considered in which this crime falls. The features given by eyewitness are given to Gaussian Mixture model and compare the PDF’s (Probability Density Function) of criminal’s features. So whenever the eyewitness submits the features, the PDF of these Features is calculated and compared with the PDF’s of those criminals in the database [6]. Those with nearer values are displayed. Output will be the CID. The MSE and PSNR of those CID’s are checked and the CID which has highest PSNR is most likely the criminal. Advantage of this process is that it clearly identifies the criminal whenever a crime is reported. It thereby helps in faster investigation. Disadvantage is that the Gaussian Mixture model does a tedious work for calculating the PDF’s for such a large number of criminals, which might be challenging.

II. PROPOSED SYSTEM

Digital forensic is a moderately new field in the sphere of digital forensic. In recent years numerous digital forensic processes asserted as a complete investigation model in digital investigation world. Numerous software developers have grabbed the chance to build up a tool same as a portion of the proposed forms. Indeed some have even built up their own digital forensic methodology, for example, Encase. But most of these tools provide partial solutions based on their understanding of the essential process components and they are limited to very few steps of investigation methodology. Some of them focus on different phases or sub-components of an investigation.

However, no joined approach, which consolidated physical crime and cyber crime detection, towards digital forensic investigation has been set up yet. The deficiencies of the investigation procedure are for the most part misused in court where both the digital evidence and the digital forensic process are investigated by the court and contradicting partners. In digital forensic field there is lack of feedback system and there is no history viewer technique is present, which increase the time of investigation. There is intense need of keeping up entire data about the case and its investigation procedure and also the result which will fill in as wellspring of data for new investigators.

The primary objective of current research work is to investigate existing digital forensic models published in well known literature and find out whether these all can be compared and added into a single digital forensic process model named as Comparative Digital Forensic Process Model (CDFPM) [7] and also find out whether any module which is need of current digital forensic world can be added into the proposed forensic model. The CDFPM will also be integrated with two new modules named as feedback module and case history keeper module. Evidence produced by implementing this process model in an investigation will en-6 sure that it can withstand legal scrutiny in a court of law. The digital evidence presented is thus the result of a rigorous digital evidence collection process.

To achieve this objective we integrate and compared various digital forensic model and finally generate a new digital forensic model [7] named as Comparative Digital Forensic Process Model and its conversion in Tool named as Comparative Digital Forensic Process Tool [8]. The phases of CDFPM can be represented as sequential logic equation:

\[
\text{CDFPM} = (\text{Proper Plan} \rightarrow \text{Authentication} \rightarrow \text{Evidence Finder} \rightarrow \\
\text{Determination of crime type and level} \rightarrow \text{Evidence Report} \rightarrow \text{Future Update})
\]

Where

- **Proper Plan** = (Plan \rightarrow \text{Readiness} \rightarrow \text{Policy} \rightarrow \text{Triggering the case})
- **Authentication** = (\text{Authenticate} \rightarrow \text{Become aware} \rightarrow \text{Classify} \rightarrow \text{Assess} \rightarrow \text{Notify} \rightarrow \text{Incident Response})
- **Evidence Finder** = (\text{Acquire} \rightarrow \text{Analyze} \rightarrow \text{Approach Strategy} \rightarrow \text{Physical Investigation} \rightarrow \text{Search} \rightarrow \text{Reconstruct})

Fig. 1 Phases of CDFPM

A secondary objective is to investigate whether complete process model can be automated, without compromising the validity of the process model with two new extra modules. Complete comparative digital forensic model will be converted into a automated comparative digital forensic process tool (CDFPT) [8]. The purpose here is to alleviate the time-consuming investigative processes on behalf of the investigators. Such processes are generally aimed at reducing the data to be analyzed during the investigation.

A digital forensic process, in recent few years, has been proposed to claim authenticate and reliable process in court of law. Many software companies developed the digital forensic tool to manifest few proposed process. Such as Encase, many authors developed a tool same as their own digital forensic methodology in digital forensic investigation.
And as per the knowledge of methodology proposed, most of the tools provide partial solution to the investigation. Some suggest new components to the existing forensic processes and added to the existing ones. In court of law where both the digital process and evidences are rejected by the court, many processes claimed to be complete are exploited. However, no compared and combined approach has been established with feedback and history keeper approach. To overcome the problem we proposed Comparative digital Forensic Process Tool CDFPT [8] whose primary objective is to learn existing top 25 digital forensic tools and identify the phases which these tools can be complete of proposed framework.

The Third objective of this paper is integrate the detection technique of physical crime and cyber crime which should also intimate a normal user about the status of crime rate in each area and also generate the list of suspects based on previous cases registered. The block diagram of the system is given in the figure 2.

![Proposed System](image)

Fig 2. Proposed System

Like for type 1 cases (Cyber Crime) we use digitized forensic tools like IP to domain converter, Domain to IP converter, IP locator etc. to identify the attacks on the system and the possible attackers. The report thus generated is sent to the intended user as well as the administrator to deal with similar type of cases in future.

For type 2 cases (Physical Crime), our approach uses K-means clustering to group crime instances into clusters with similar attributes for crime detection. These clusters are visually displayed using GMAPI. We have formulated a crime data set containing pre-defined crime records from web sources. Our tool implements KNN classification for filtering out possible list of criminals from our crime data set for a particular crime case. This suspect list is then displayed to the user.

### III. PROPOSED ALGORITHM

Finally we represent the crime trends graphically for both type 1 and type 2 cases based on crime location.

The proposed algorithm is as follows:

1. Categorize the registered crime case into type 1 or type 2.
2. Store the registered crime details into respective databases.
3. Evaluating according to type of crime:
   a. For type 1 crime case, create crime cluster from database using Kmeans clustering. Display the clusters using GMAPI.
   b. For type 2 crime cases, apply digitized forensic tools to assess the crime details and generate the reports.
4. Use KNN classification to muster the possible suspect list for type 1 case. If no match is found then new crime pattern is added to the data set.
5. Graphically display the crime pattern trends for type 1 and type 2 cases.

### IV. WORKING FLOW OF PROPOSED SYSTEM

The workflow of our proposed system starts with the user registering to our system as given in figure 3 and then logging in to the system as shown in Figure 3.
The Login can be done using either as registered user or as Administrator. The user will not able to access the system until the Administrator will not approve and verify all the details of the user.

Once the Verification and approval done by the administrator, the user can login to the system using User ID and Password. She/he can register case details of either cyber crime or physical crime.

The case is registered by providing a unique case ID to the user and the case details are stored in our database. This database is then supplied to either physical crime case (type 1) or cybercrime case (type 2) modules.

The Administrator Menu have 3 main Module as shown in Figure 5 first Data Loader which keeps all details like registered user, cases and previous cases and its solution and so on. This module actually works like history viewer and keep all feedback of all previous records. The investigator can send related data or suggestion to the users using that module. Second Module is all Digital Forensic Tools shown in figure 6. Final module is Report Module which contains all types of records, case reports, year wise report and some pictorial representation for admin purpose.
V. ADVANTAGES AND DISADVANTAGES

A. Advantages

1. Provides Security of personal information by the Admin and User login, thus our project is more secured.
2. Our system will help the law enforcing agencies to reduce the crime rate by identifying the crime percent of a particular city.
3. GMAPI used will speed up the crime investigation to enforce security measures in those effected locations with enhanced visualization.
4. The approach is not limited to a particular place or type of crime.
5. The user can register the crime even at home no need to run to the police station.
6. It will also speed up the crime solving process by processing and filtering the voluminous crime data within a short span of time.
7. Our system is for both investigating physical crime as well as for the cybercrime.
8. The KNN method used will help in determining the criminal based on the past records available.
9. System can help the police and justice departments to narrow down the identification of criminals. This in turn will reduce the cost and time of crime investigation.
10. It can aid the law enforcement agencies to enforce the security of citizens.

B. Disadvantages

1. In order to register the case the user first needs to login the system.
2. The approach for cyber crime detection can only trace the location not the actual criminal.
3. If the criminal is new in the records then it will become difficult to identify him.
4. Instead a new crime record is created in our system to avoid the crimes in future.

CONCLUSION

One of the critical undertakings of police associations is Crime Investigation. In today's IT empowered period numerous techniques are utilized for prevention and investigation of crime. Data mining practices is one part of Crime Investigation. We have utilized classification and clustering techniques to analyze data from the database. Crime investigation agencies seek the database of culprits which is a dreary procedure. So to contribute towards combating crimes and to recognize culprits, we propose a coordinated innovation for both cyber and physical crime. We can state that with a specific end goal to recognize physical crime KNN is utilized for criminal distinguishing proof by filtering the number of suspected lawbreakers. Lately, cybercrimes has turned out to be progressively advanced making it hard to recognize and alleviate. So in cybercrime we have utilized digitized forensic tools that are created using digital forensic to identify cyber crime.

REFERENCES

[4] Priyanka V. Kayarkar, Mining Frequent Sequences for Emails in Cyber Forensics Investigation


[34] GuidanceSoftware.Encase.searchtechnologyvalidated


