ANKARA YILDIRIM BEYAZIT UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

A MODEL OF SUSPECTED ACTIVITY DETECTION AND APPLICATION USING BIG DATA ANALYTICS

M.Sc. Thesis by

Hüsrev Abdulcelil KARACABEY

Department of Computer Engineering

September, 2019

ANKARA

A MODEL OF SUSPECTED ACTIVITY DETECTION AND APPLICATION USING BIG DATA ANALYTICS

A Thesis Submitted to

The Graduate School of Natural and Applied Sciences of

Ankara Yıldırım Beyazıt University

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Engineering, Department of Computer Engineering

by

Hüsrev Abdulcelil KARACABEY

September, 2019

ANKARA

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"A MODEL OF SUSPECTED ACTIVITY DETECTION AND APPLICATION USING BIG DATA ANALYTICS"** completed by **HÜSREV ABDULCELİL KARACABEY** under the supervision of **ASST. PROF. DR. AHMET ERCAN TOPCU** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Ahmet Ercan TOPCU

Supervisor

Ankara Yıldırım Beyazıt University

Asst. Prof. Dr. Fahreddin Şükrü TORUN

Jury Member

Ankara Yıldırım Beyazıt University

Asst. Prof. Dr. Ali Osman ÇIBIKDİKEN

Jury Member

Necmettin Erbakan University

Prof. Dr. Ergün ERASLAN

Director

Graduate School of Natural and Applied Sciences

I hereby declare that, in this thesis which has been prepared in accordance with the Thesis Writing Manual of Graduate School of Natural and Applied Sciences,

- All data, information and documents are obtained in the framework of academic and ethical rules,
- All information, documents and assessments are presented in accordance with scientific ethics and morals,
- All the materials that have been utilized are fully cited and referenced,
- No change has been made on the utilized materials,
- All the works presented are original,

and in any contrary case of above statements, I accept to renounce all my legal rights.

Date: 24/09/2019 **Signature: ……..**

Name & Surname: Hüsrev Abdulcelil KARACABEY

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my supervisor, Asst. Prof. Dr. Ahmet Ercan TOPCU for his tremendous support and motivation during my study. His immense knowledge and precious recommendations constituted the milestones of this study. His guidance assisted me all the time of my research and while writing this thesis.

24/09/2019 **Hüsrev Abdulcelil KARACABEY**

A MODEL OF SUSPECTED ACTIVITY DETECTION AND APPLICATION USING BIG DATA ANALYTICS

ABSTRACT

This study proposes a big data analytics methodology to analyse data that are available from many institutions for providing safety to the citizens in their daily life. In the traditional protection system approach, tactics are mainly dependent on the experience of the law enforcements. However, it is really difficult to generalize their experiences for applying solutions in the real world. Also, the conventional approach usually loses its prevention capability because it would take more time to react to the incidents. If the crime happened, the damage would already be done to the victims. So, the best interest for the law enforcements is to prevent a crime before happening.

We believe that in order to have effective criminal activity prevention, law enforcement needs to implement data driven approaches, solutions and real time processing of data.

In this study we suggested to use CDR (Call Detail Records), ANPR (Automatic Number Plate Recognition), API/PNR (Advanced Passenger Information/Passenger Name Records) data. While combining these we have a methodology that is a force multiplier for law enforcement in their duty to counter criminal activities.

This study presents sample scenarios, use cases and methodologies while using big data for detecting suspected activities of individuals and objects for safety purposes. Hence, this study presents a model to analyze CDR, API/PNR and ANPR data using big data analytics for law enforcement usage to protect people from malicious activities.

Keywords: Big data, security, data, analysis, detection, law enforcement, CDR, ANPR, API/PNR.

BÜYÜK VERİ ANALİTİKLERİ KULLANARAK ŞÜPHELİ EYLEM TESPİTİ MODELİ VE UYGULAMASI

ÖZ

Bu çalışma, vatandaşlara günlük yaşamlarında güvenliğin sağlanabilmesi amacıyla pek çok kurumdan elde edilen verileri analiz edebilmek için bir büyük veri analitiği çözümü önermektedir. Geleneksel koruma sistemi yaklaşımında, taktikler temel olarak kanun uygulayıcıların deneyimine dayanmaktadır. Bununla birlikte dünyada çözümler uygulayabilmek amacıyla onların tecrübelerini genellemek oldukça zordur. Ayrıca, geleneksel yaklaşımın olaylara reaksiyon göstermesi daha fazla zaman alabileceğinden önleme kabiliyetini genellikle yitirmektedir. Eğer suç oluştu ise zarar mağdurlara çoktan verilmiştir. Bu sebepten kanun uygulayıcılar için en önemlisi suçu oluşmadan önce önlemektir.

Etkili bir şekilde suç faaliyetlerinin önlenmesi için, kolluk kuvvetleri veriye dayalı yaklaşımları, çözümleri ve gerçek zamanlı veri işlenmesini gerçekleştirmelidir.

Bu çalışmada CDR (Call Detail Records), ANPR (Automatic Number Plate Recognition), API/PNR (Advanced Passenger Information/Passenger Name Records) verilerini kullanmayı önerdik. Bunları birleştirerek, suç faaliyetlerine karşı koyma görevlerinde kolluk kuvvetlerine kuvvet çarpanı olabilecek bir metodoloji oluşturulmuştur.

Bu çalışma güvenlik amacıyla bireylerin ve objelerin şüpheli aktivitelerinin büyük veri kullanarak tespit edilmesi için örnek senaryolar, vakalar ve metodolojiler sunmaktadır. Bununla birlikte, bu çalışma kolluk kuvvetlerinin kullanımı için insanların kötü niyetli faaliyetlerden korunması amacıyla büyük veri analitikleri kullanarak CDR, API/PNR ve ANPR verilerinin analizini yapabilecek bir model sunmaktadır.

Anahtar Kelimeler: Büyük veri, güvenlik, veri, analiz, tespit, kolluk kuvvetleri, CDR, ANPR, API/PNR

NOMENCLATURE

Acronyms

CDR	Call Detail Record
ANPR	Automatic Number Plate Recognition
API	Advance Passenger Information
PNR	Passenger Name Records
LOTRF	Location and Time Risk Factor
PSF	Personal Suspicion Factor
VSF	Vehicle Suspicion Factor
TASF	Togetherness Analysis Suspicion Factor

LIST OF FIGURES

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Overview

The aim of this study is to build a methodology to find the suspected activities of individuals and objects using data analysing to prevent future criminal acts.

Big data consist of extremely large data sets that could be analysed computationally to reveal patterns, associations, etc. Big data characteristics are; volume, variety, veracity and velocity, this is known as the four Vs [1]. Big data are larger, more complex data sets than traditional ones. These massive volumes of data can be used to address security problems that we wouldn't have been able to solve before. Some of the data sets we will be mention in our study are an example of big data such as CDR data, API/PNR data and ANPR data.

The four Vs of big data;

Volume: The quantity of generated and stored data.

Variety: Variety refers to the many types of data that are available.

Veracity: The quality and applicability of the data.

Velocity: Velocity is the fast rate at which data are received and acted on.

Crime is the intentional commission of an act usually deemed socially harmful or dangerous and specifically defined, prohibited and punishable under criminal law [2]. It is an omission against the public which every state and law enforcement wishes to prevent. But, unfortunately, if the crime happened, the damage is already done to the victim or victims. It is in law enforcement's best interest to prevent a crime before it's happened instead of trying to solve the crime, which in fact does little to none on repairing or rehabilitating the victims lost, either physically or emotionally. In today's world, big data is very important for assisting a country's security forces in their

operation to find criminals by combining different data from different domains. Because of the reasons we give above, we believe finding a way to analyse and utilise big data is an apparent need for today's law enforcement.

1.2 Research Objectives

We want to prevent crime, for that we assume, we have to pinpoint the individuals who are most likely to commit a crime. To do that we must analyse real-time data (such as contacts, location information) and combine it with archive data (criminal records and other confidential data) to assess risk to an individual or an object such as a vehicle. After the assessment of risk then we will provide an alarm to the law enforcement so they can intervene in the situation.

1.3 Research Contributions

In the scope of this study, we created and proposed a flexible methodology for the detection of suspected activities for the prevention of criminal activities. We provide many examples for using various data types (such as; Call Detail Record, Automatic Number Plate Recognition, Advanced Passenger Information/Passenger Name Records) for crime prevention purposes. Since the methodology is flexible, it can be updated according to needs and changes for law enforcement work and our methodology can integrate and communicate with different systems to effectively assess the risk to an individual's or an object's activities in a modular fashion.

1.4 Structure of the Thesis

This study is composed of six chapters;

The first chapter is the introduction chapter which includes the overview, research objectives and contributions, the second chapter includes two important aspects of the methodology which are Suspicion Rate and Action Value.

The third chapter is about the theoretical background and the literature review. The fourth chapter explains the data inputs for the methodology such as CDR, ANPR, API/PNR, PSF, etc.

The fifth chapter explains the detection methodologies which are proposed for suspicion detection such as, CDR and Togetherness analysis, the sixth chapter contains a prototype implementation and the seventh chapter is about discussion and conclusion of the study.

CHAPTER 2

THEORETICAL BACKGROUND AND LITERATURE REVIEW

The conventional law enforcement approach is mainly dependent on the experience of the servicemen thus it is really difficult to generalize their experience and apply it to the real world. And above all the conventional approach usually loses its prevention capability because it can act very late. The Broken Window Theory [3] is an example of that.

We believe that to effectively prevent crimes, law enforcement needs to implement methods that's data-driven and continuous. Modern law enforcement especially the organizations which have federal or state-wide responsibility collects gigabytes of data every day from various resources (other state institutions, smart city systems, etc.), but they very rarely utilise the data effectively to prevent criminal activities.

The collected big data (petabytes of data, the institutions use many solutions to store and analyse the data, such as Relational Databases, NoSQL Databases or sometimes more advanced data warehouse solutions depending on their budget, priorities, infrastructure and analytic systems) is mainly used in investigations to solve criminal cases or to provide evidence to legal judicial system. The before mentioned law enforcement institutions usually have vast computer system resources to collect data and run vast computer networks within their area of responsibility, however, most of these expensive and sophisticated systems barely been used effectively or meet their full potential. With our study, we want to change this situation and reality, and provide an extensive and inclusive road map for those institutions, so they can use this study as a guide and continue and develop or evolve their computer systems to suit their needs.

Research about supporting law enforcement by using computer systems and big data analytics been always an intriguing topic. There have been studies [4] which map potential crime spots and use Machine learning techniques which assist crime

prevention, and a police big data analytics platform [5] are good examples for data-driven approaches for law enforcement work.

In this study we focused on mainly CDR (Call Detail Records), ANPR (Automatic Number Plate Recognition) and API/PNR (Advanced Passenger Information/Passenger Name Records) which used by law enforcement all around the world in their investigations. There are examples of solutions that use CDR data such as a Social Network Analysis system [7] which uses CDRs to classify influential subscribers, a system allows criminal investigations on CDR based on the criterions and queries by anti crime teams [8]. Another example could be a Criminal investigation solution for telecom that is developed by [9] and implemented on the cloud to reduce cost. CDR is a valuable data type for law enforcement since it has location data and shows interaction or communication between individuals. The ANPR data are created by the cameras that installed on the road crossings which include information about a vehicles licence plate number, time of the passage, etc., the main reason we choose this data type is the role of the vehicles in criminal and terrorist activities. ANPR also can be used for a variety of things such as vehicle activity classification [11], monitoring traffic signal violations [12] ,etc. The other data type is passenger data or API/PNR data which is mainly used by airline companies for commercial purposes or other research studies such as using PNR to predict passenger nationality [13]. In our case, the data are used for security purposes. We include this data type since the Foreign Terrorist Fighters (FTFs) use air travel to reach their target destination.

CHAPTER 3

SUSPICION RATE AND ACTION VALUE

3.1 SR – Suspicion Rate

Suspicion Rate (SR) is a term we define and propose for the algorithm. Everything whether an individual or an object has a SR associated with it. SR is a numerical value. It starts at zero and goes to infinity. We constantly increase or decrease the SR of the individual or object by analysing its activities. If the SR reaches a certain predetermined level, we can warn the law enforcement or security and preventive actions can be taken.

An increase or decrease of the Suspicion Rate is depended on the data associated with the object and its activities. Proposed data fields for SR represented in Figure 3.1. Details of the manipulation of SR will be explained in the Detection Methodologies section.

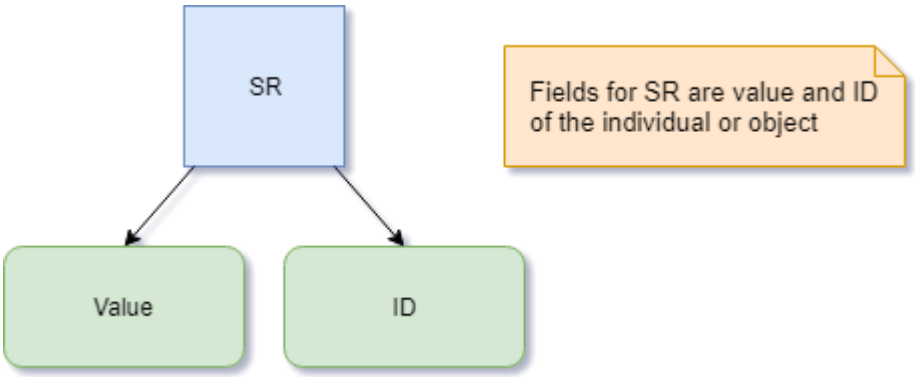

Figure 3.1 SR data fields

3.2 AV – Action Value

Action Value (AV) is another term we define and propose for the algorithm. Action Value is a numerical value. It starts from zero and goes to infinity. Everything whether an individual or an object has an AV associated with. It can be a predetermined or

periodically updated according to changes in Personal Suspicion Factor (PSF) or Vehicle Suspicion Factor (VSF). If Suspicion Rate for an individual or an object reaches or surpasses Action Value, it means there is a high possibility for a criminal action could take place that includes the individual or the object. Proposed data fields for AV represented in Figure 3.2. Details of the PSF and VSF will be explained in Chapter 4.

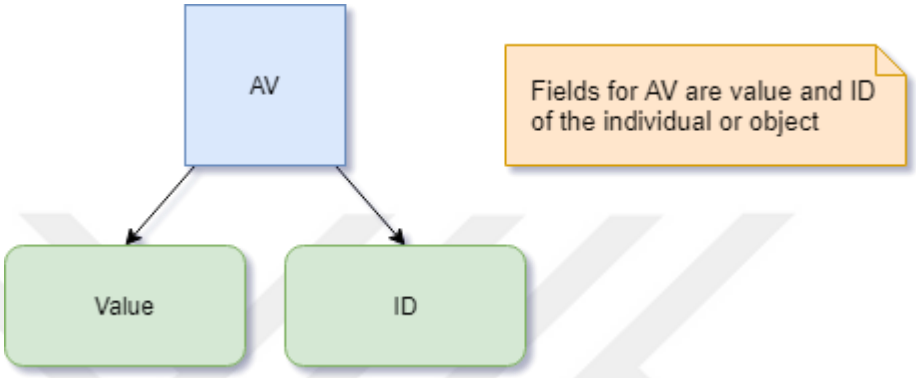

Figure 3.2 AV data fields

CHAPTER 4

DATA INPUTS

The data for this algorithm will consist of data sets which indicate an action of an individual or an object, archive records of the individual and other processed data. Such as CDR (Call Detail Records), ANPR (Automatic Number Plate Recognition Data), API/PNR (Advance Passenger Information/Passenger Name Records), Background Information (criminal records, etc.).

In the study we cannot acquire real data because of the legal reasons since to acquire the data, we need court permission or a warrant. But we know the fields of the data and what we could do with it.

Below we explain the data our algorithm will work on. The algorithm is not limited to the data we mentioned, we can incorporate other types of data to assist suspicion detection.

4.1 CDR-Call Detail Record

CDR is a record that contains detailed information about a telecom transaction, such as call start time, end time, duration, call parties, cell ID, requested websites [6]. It is produced by telephone call and documents the details of a transaction (s.a. duration, cell tower info, timestamp, source number and destination number). A subset of data fields of CDR represented in Figure 4.1. The Call Detail Records are being used by the telecom industry for their basic needs such as billing and charging, our focus here is to detect a movement and analyse connections.

CDR has the 3Vs characteristics of big data (Volume, Variety, Velocity), therefore it can be seen as a big data source [6].

A sample of CDR data is given as follows;

<CDR Record Sample>

<Imsi>223344556677889**</Imsi>**

<Imei>455667788999**</Imei>**

<Cell Tower Info>123768123**</Cell Tower Info>**

<Source Number>5050000000**</Source Number>**

<Destination Number>5050000001**</Destination Number>**

<Duration>247**</Duration>**

<Timestamp>1566140760**</Timestamp>**

<Date>01/08/2019**</Date>**

</CDR Record Sample>

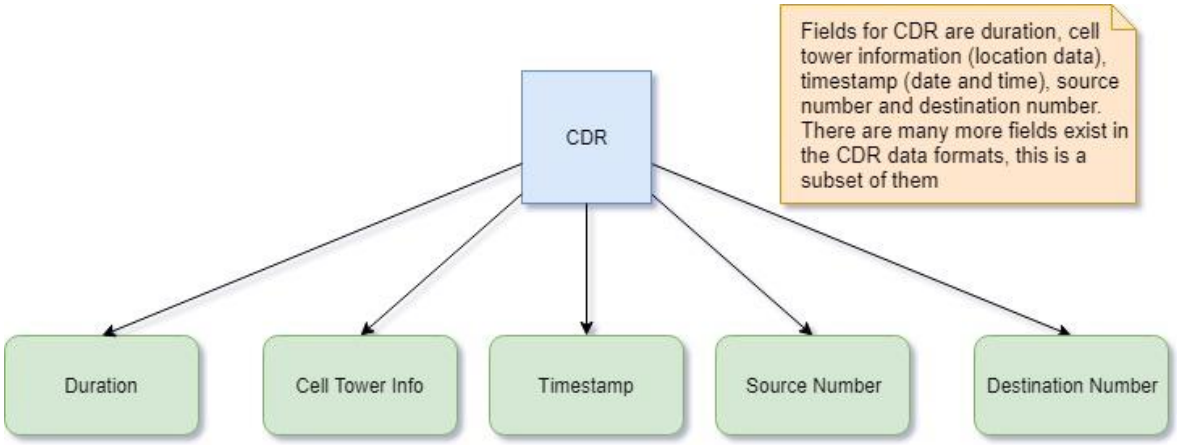

Figure 4.1 CDR data fields

4.2 ANPR-Automatic Number Plate Recognition

Automatic Number Plate Recognition (ANPR) data are generated by cameras that are installed at road crossings and transmitted to a central data center of the traffic management department continuously [10]. Automatic Number Plate Recognition is a technology that uses optical character recognition to read vehicle registration plates to create vehicle location data. It is used by law enforcement all around the world for investigations.

We can consider the ANPR data as big data because it has characteristics of big data (Volume, Veracity, Velocity), The data contains attributes like; **licence plate information, timestamp, tower information.** A subset of data fields of ANPR represented in Figure 4.2.

A sample of ANPR data is given as follows;

<ANPR Record Sample>

<Tower Info>185778134**</Tower Info>**

<Licence Plate Info>06AB111**</Licence Plate Info>**

<Timestamp>1566141000**</Timestamp>**

<**Date**>01/08/2019</**Date**>

</ANPR Record Sample>

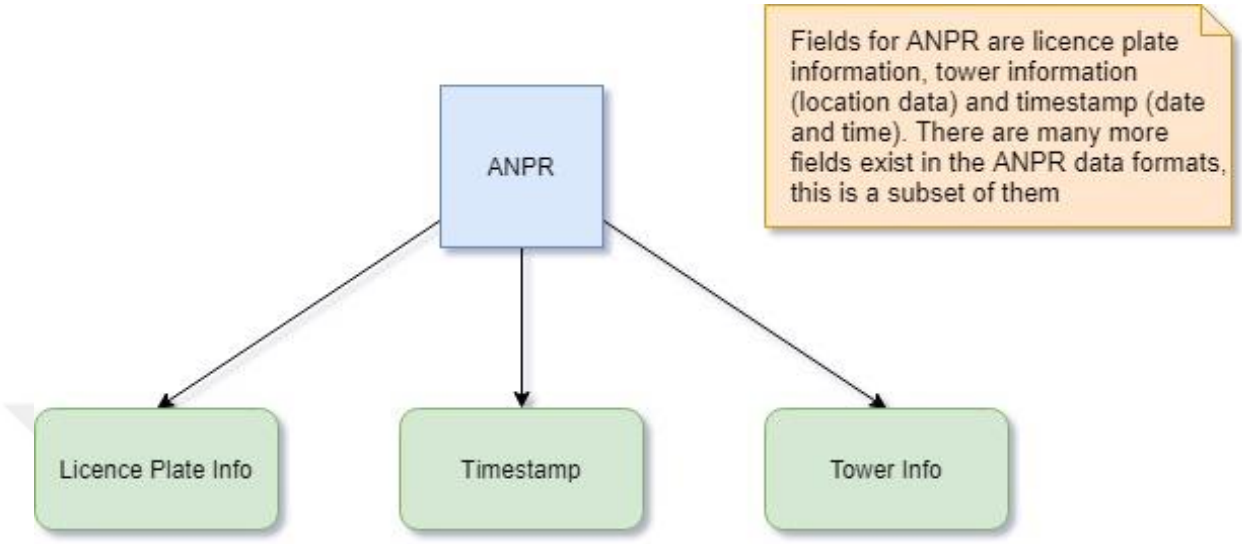

Figure 4.2 ANPR data fields

4.3 API-Advance Passenger Information, PNR-Passenger Name Records

The Advance Passenger Information Data or API is an electronic data interchange system data that contains information about a passenger's **full name, gender, date of birth, nationality, travel document number, booking data,** etc.

Passenger Name Record (PNR) is a set of data created when a travel reservation is made. It is generated by airlines or authorized agents such as travel agencies and standardized by the International Air Transport Association [13]. It archives the airline travel itinerary for the individual passenger and a group of passengers travelling together [14]. PNR consists of personal information for a passenger including; **full name, gender, date of birth, nationality, travel document number, booking data,** etc. A subset of data fields of API and PNR represented in Figure 4.3 and Figure 4.4.

For example, when travelling to Turkey passengers are required to provide advance passenger information (API) before they check in. And also travel companies has to

share passenger information with the countries they visit even before the passengers arrive according to regulations worldwide.

Passenger data are used by airliners for a variety of purposes for instance to find influential passengers, to offer better service for their customers to increase their loyalty, finding customer behavior or to being one step ahead of the competition. We can consider the Passenger data as big data because it has characteristics of big data (Volume, Variety, Velocity). The algorithm uses passenger data to detect a movement and manipulate the SR value of the passenger according to. A sample of API/PNR data is given as follows;

<API-PNR Record Sample>

<Fullname>Passenger Sam**</Fullname>**

<Gender>M**</Gender>**

<Nationality>USA**</Nationality>**

<Travel Document Number>25771915456340**</Travel Document Number>**

<Date of Birth>11/07/1980**</Date of Birth>**

</API-PNR Record Sample>

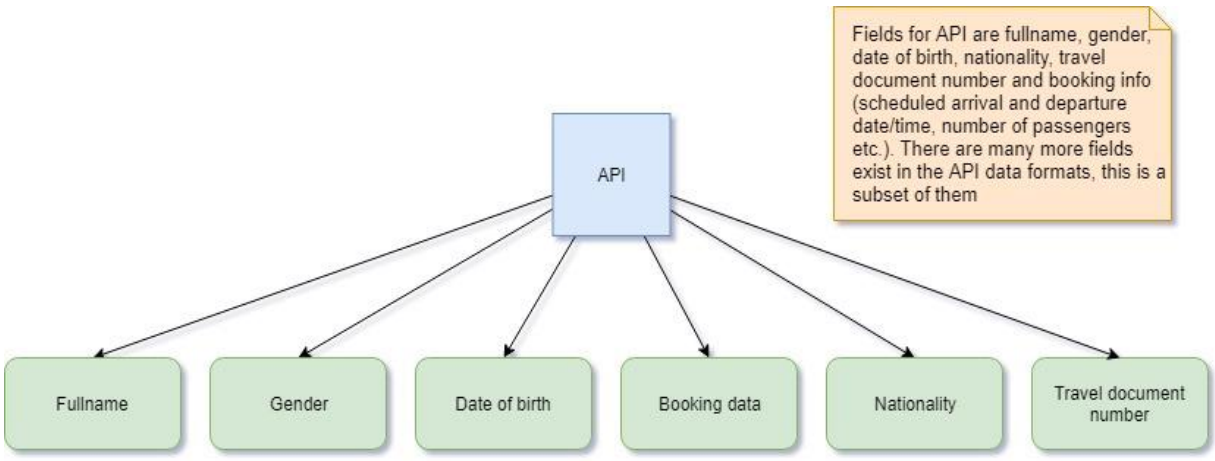

Figure 4.3 API data fields

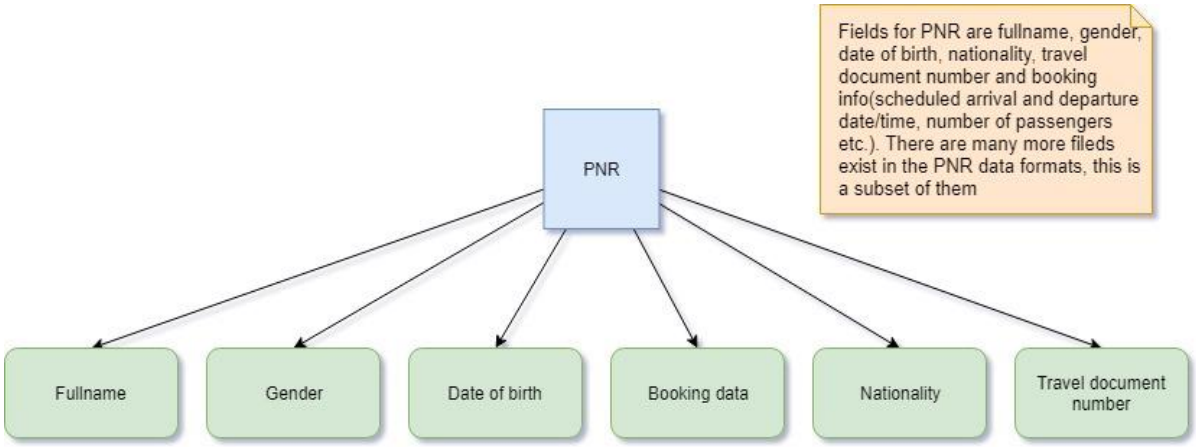

Figure 4.4 PNR data fields

4.4 LOTRF-Location and Time Risk Factor

The LOTRF (Location and Time Risk Factor) data are processed data that we create for the algorithm. It's about how much risk a location poses at a certain time. It consists of **location, time and value (an integer value)**. Proposed data fields for LOTRF represented in Figure 4.5.

The data are created by analysing the law enforcement records of criminal activities. We are interested in time, location and the criminal activities that took place at that location and time of those criminal activities.

We assume there is a correlation between future criminal activities and past criminal activities for a location. The data will have a numerical value for each time interval at a specific location. The LOTRF data will look like this; **time**: 13:00, **location**: Ulus/Ankara, **value**: 10.

The LOTRF value starts from zero and goes to infinity. The higher the value the more risk of criminal action will be taken place. We must update the LOTRF data continuously to be effective, the LOTRF data are one of the most important data that the algorithm will use to manipulate SR.

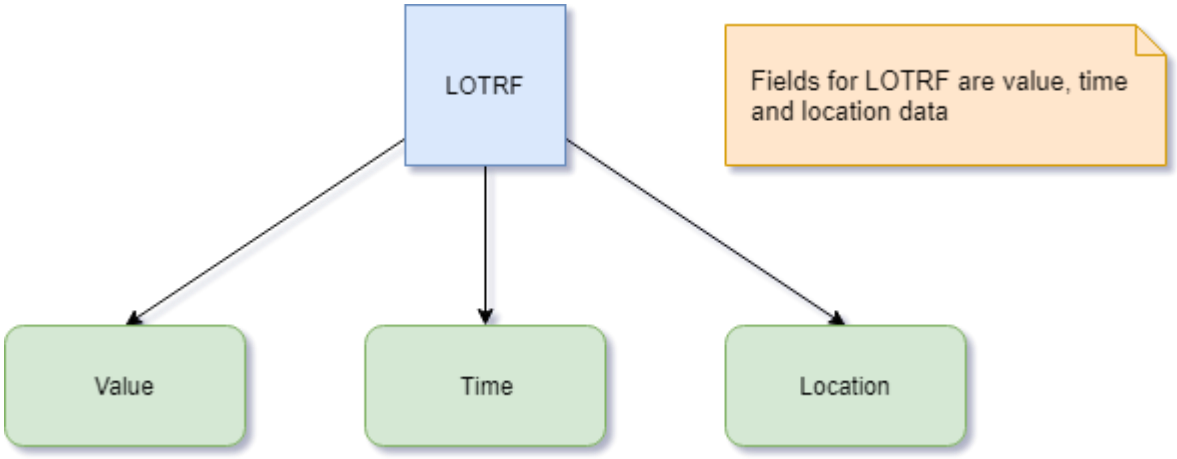

Figure 4.5 LOTRF data fields

4.5 PSF-Personal Suspicion Factor

The PSF (Personal Suspicion Factor) data are processed data that we create for the algorithm. It is a numerical value ranges from zero to infinity. Proposed data fields for PSF represented in Figure 4.6. To calculate the PSF we must analyse the law enforcement criminal records of individuals. The number of criminal records and type of criminal activities should affect the PSF value differently.

PSF value must be updated as criminal records change. We assume there is a strong correlation between an individual's criminal past and future criminal activities. If an individual's PSF is higher, the law enforcement must give more attention to that individual’s activities.

PSF will affect the SR, higher the PSF value means higher the SR value. The PSF data will look like this **ID**: XY123, **value**: 5.

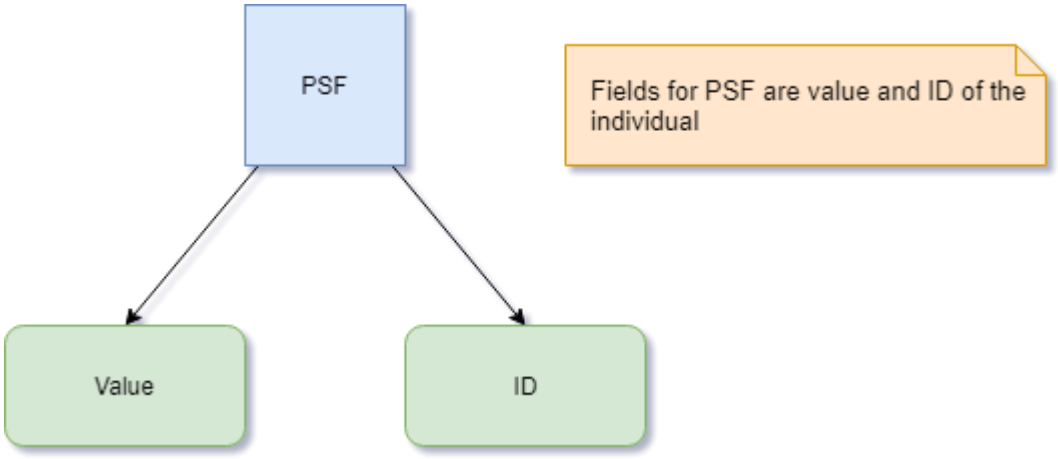

Figure 4.6 PSF data fields

4.6 VSF-Vehicle Suspicion Factor

The VSF (Vehicle Suspicion Factor) data are processed data that we create for the algorithm. It is a numerical value ranges from zero to infinity. Proposed data fields for VSF represented in Figure 4.7. To calculate the VSF, we must analyse law enforcement criminal records. If the vehicle is used in a crime, we will increase its VSF value. Number of criminal records and type of the criminal activities should affect the VSF value differently.

VSF value must be updated as criminal records change. VSF will affect the SR, higher the VSF value means higher the SR value. The VSF data will look like this **ID**: AS1443, **value**: 7.

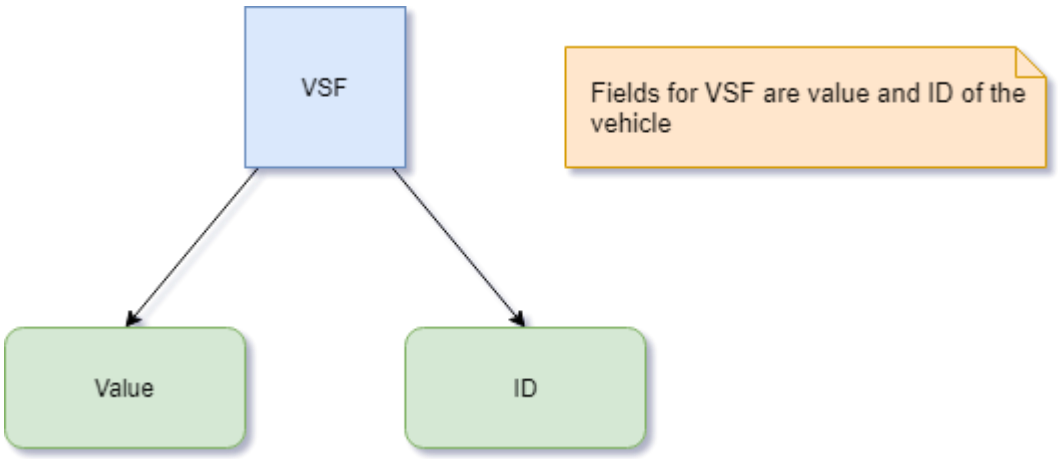

Figure 4.7 VSF data fields

4.7 TASF-Togetherness Analysis Suspicion Factor

The TASF (Togetherness Analysis Suspicion Factor) data are processed data that we create for the algorithm. It is a numerical value ranges from zero to infinity. To calculate the TASF, the algorithm uses CDR and ANPR data. Proposed data fields for TASF represented in Figure 4.8.

We assume if the individuals or vehicles move together, that indicates a risk for criminal activities. We assume If multiple individuals or vehicles move or act as a group and all or some of the individuals or vehicles have high PSF or VSF values they are likely to commit a crime and law enforcement should intercept their activities.

To find togetherness within a group of people the algorithm uses CDR data. Since CDR data indicate a connection between people and it has location information. The algorithm could look into the connections and interactions between the group and analyse their location information to see if they are moving together. After detection of togetherness, the algorithm will calculate the TASF, which is the sum of all PSF values of the individuals within the group then manipulate their SR.

To find togetherness within a group of vehicles the algorithm uses ANPR data. Since ANPR data has information about time, location about the vehicle's movements, the algorithm could detect if multiple vehicles move together. There are studies about detecting multi-vehicle convoy scenarios using ANPR data [15].

After the algorithm detects the togetherness, it will look for the VSF value for each vehicle, calculate the TASF, which is the sum of all VSF values of the vehicles within the group and manipulate their SR with TASF value. Criminals (for example smugglers) usually put a vehicle as a "police detector" in front of the real carrier vehicle that's called "pioneer" that looks for the police [16]. When the "pioneer" vehicle encounters the police, it informs the real carrier car or their partner and they can get away from the police. Criminals also track vehicles such as bank cash carriers as a group to commit robbery [10]. Our approach can eliminate threats that use these kinds of tactics.

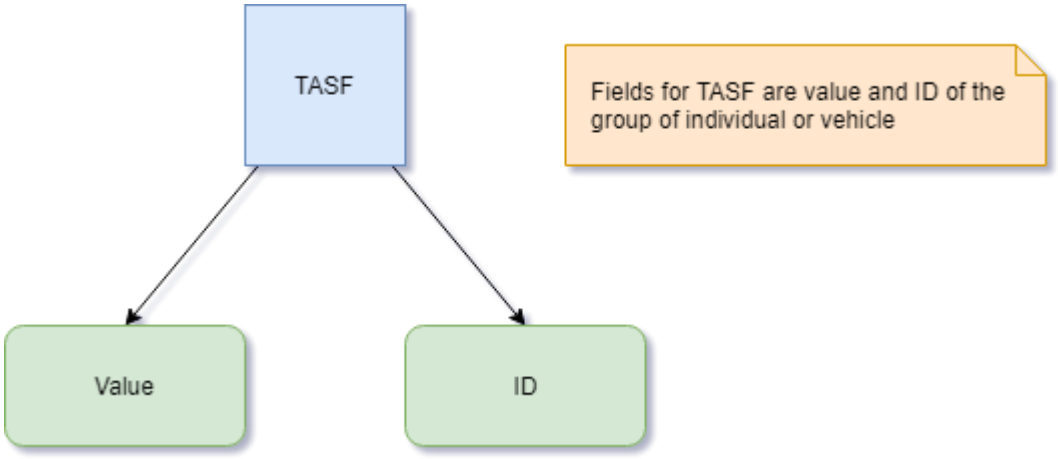

Figure 4.8 TASF data fields

CHAPTER 5

DETECTION METHODOLOGIES

Our aim is to construct an algorithm. The algorithm can run at certain intervals, constantly or fired by an action (a phone call, an ANPR detection, a check-in at the airport). After the initiation, the algorithm starts to analyse the data and manipulate the SR. The algorithm modeled to work by analysing big data gathered from many different systems and sources, such as CDR data from telecom companies, ANPR from traffic management authorities, API/PNR data from airline companies and authorities.

SR value will be manipulated separately by each different activity. All activities will increase the SR either with zero or any other positive integer value. The reduction of the SR values will be depended on time that there is no increase in SR value for the individual or object.

The reduction rate or amount could be constant or dynamic. The reduction of the SR values could be different for each individual or object, it could dynamically be calculated according to their PSF or other related data. For example; reduction rate or amount could be slower or less for the individual who has high PSF value or a vehicle that has high VSF value. To find the ideal reduction method we have to analyse

feedback from the field and form an approach to find a sweet spot for each individual or object.

After the alarm is given, law enforcement units can patrol the area, make ID checks or move their personal to critical areas to be a deterrent for possible crimes.

We have to mention that the alarm does not prove that there is a criminal activity is taking place by the ones that are detected, as we stated before the alarm is based on suspicion detection and do not prove actuality of a real criminal action, it gives law enforcement a chance to prevent a possible instance of a crime by acting proactively. And we believe that by acting proactively we can reduce the crime rates and improve the safety and well being of citizens.

5.1 CDR-Call Detail Record

Call Detail Records shows an interaction between individuals, it also has attributes that show the time and location of the interaction, in our methodology after the detection of the individuals by CDR data, the PSF (Personal Suspicion Factor) will be retrieved for both, then LOTRF (Location and Time Risk Factor) will be retrieved according to their location and time information. Then the algorithm will increase each party's SR (Suspicion Rate) according to the sum of their PSF and locations LOTRF values.

For instance, there are two individuals A and B, the A calls or sends a text message to the B, this interaction will generate a CDR record, after the generation of the CDR record, it must be analysed by our algorithm in real or near-realtime to be effective. After the interaction, the algorithm calculates the sum of each individual's PSF and their locations LOTFR values, then it will increase the Suspicion Rate of A and B by that value. Steps for CDR analysis represented in Figure 5.1.

We assume that a high PSF value is an indicator of a possible tendency for criminal behaviour, and people who are interacting with people with high PSF values should be given special attention to them, if their SR value reaches their respective AV (Action Value). We also assume that individuals locations are also something to consider, being at a place at a time that has historically high rates of crime should be seen as a possible sign of criminal activity.

After the increase of A and B's Suspicion Rate, if any of the individuals SR reach his/her Action Value, an alarm should be created for both of them and appropriate actions by law enforcement should be taken accordingly.

These are main steps for CDR analysis:

After a CDR record is created for A and B;

- Step 1: Check the PSF of A
- Step 2: Check the PSF of B
- Step 3: Check the LOTRF value of A's location with respect to time
- Step 4: Check the LOTRF value of B's location with respect to time
- Step 5: Calculate the SR increase of A by adding the PSF and the LOTRF value
- Step 6: Calculate the SR increase of B by adding the PSF and the LOTRF value
- Step 7: Calculate the sum of the SR increase of A and B, then increase the SR of A and B with the calculated value
- Step 8: Check if either A or B reaches their respective Action Value
- Step 9: Alarm the law enforcement if either A or B reach their respective AV

Sample pseudocode for CDR analysis:

```
WHEN a CDR record is created for A and B:

begin

  #Retrieve the PSF value for A and B from the database then
  initialize variables

  var_psf_of_A = getPSFfromArchieveStorage(ID of A)

  var_psf_of_B = getPSFfromArchieveStorage(ID of B)

  #Retrieve the LOTRF value for A's and B's location from the
  database then initialize variables
```

```
#Location and Time information should be extracted from CDR
record

var_lotrf_for_A = getLOTRFfromArchieveStorage(time and
location info)

var_lotrf_for_B = getLOTRFfromArchieveStorage(time and
location info)

#Calculate the SR increase for A and B then initialize
variables

var_sr_inc_for_A = sum(var_psf_of_A, var_lotrf_for_A)

var_sr_inc_for_B = sum(var_psf_of_B, var_lotrf_for_B)

#Calculate the sum of the SR increase of A and B then
initialize the variable

var_sr_inc_for_A_and_B = sum(var_sr_inc_for_A,
var_sr_inc_for_B)

#Update the SR values of A and B with the sum of their SR
increase

updateSR_PersonObjectStorage(ID of A,
var_sr_inc_for_A_and_B)

updateSR_PersonObjectStorage(ID of B,
var_sr_inc_for_A_and_B)

#Retrieve the Action Value and updated Suspicion Rate from
the database for A and B

var_av_of_A = getAVfromPersonObjectStorage(ID of A)

var_av_of_B = getAVfromPersonObjectStorage(ID of B)

var_sr_of_A = getSRfromPersonObjectStorage(ID of A)

var_sr_of_B = getSRfromPersonObjectStorage(ID of B)

#Check if either A or B reach their respective Action Value

If var_sr_of_A greater than or equal to var_av_of_A

  Send alarm to the law enforcement for A and B
```

```
  endif
  If var_sr_of_B greater than or equal to var_av_of_B
    Send alarm to the law enforcement for A and B
  endif
end
```

CDR analysis for suspicion detection:

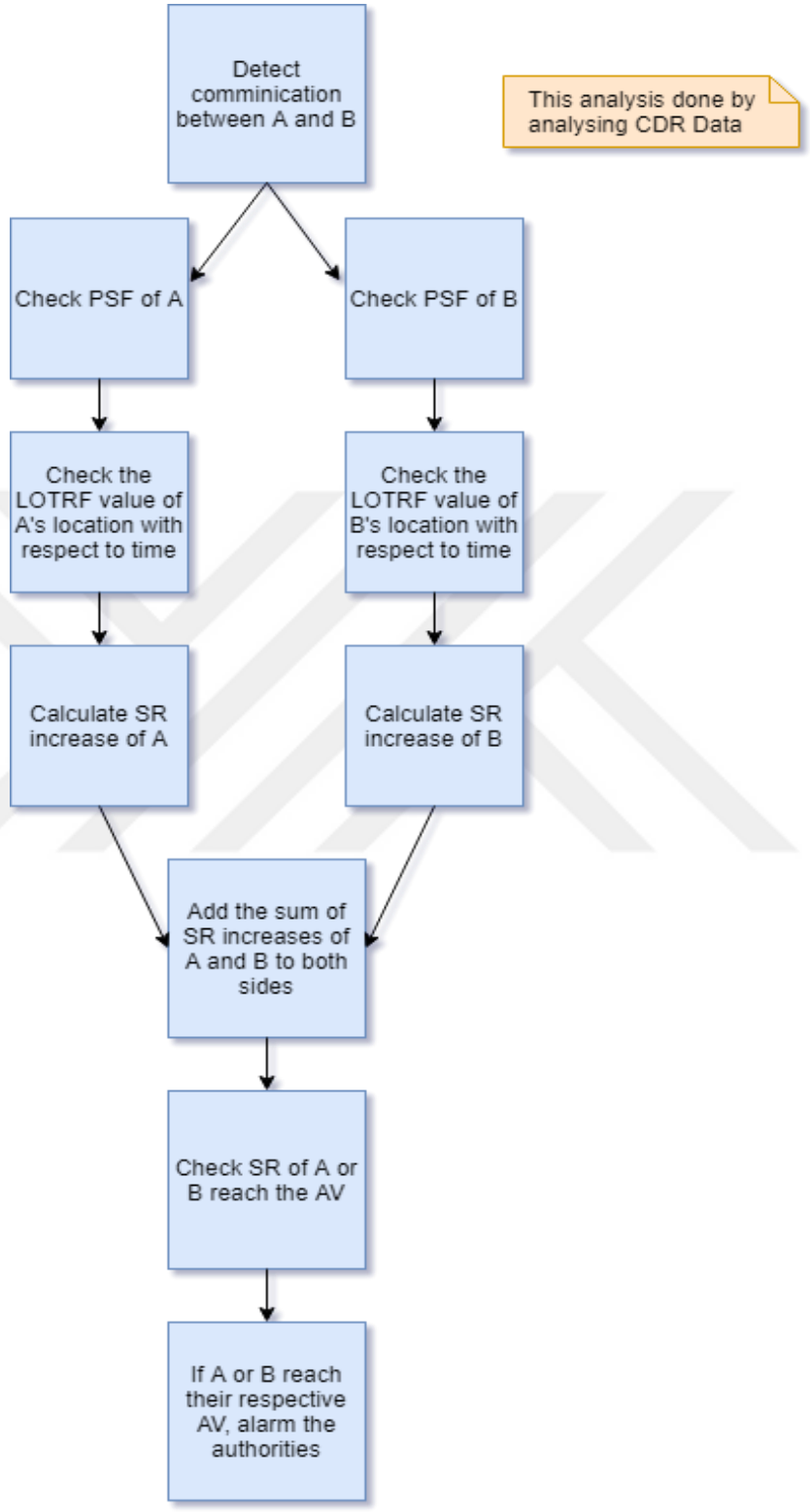

Figure 5.1 CDR data analysis for suspicion detection

5.2 ANPR-Automatic Number Plate Recognition

Automatic Number Plate Recognition data shows the movement of vehicles. It has attributes that include information about the time, location of the cameras, plate number, etc. In our methodology after the detection of movement of the vehicle, the VSF (Vehicle Suspicion Factor) of the vehicle, PSF (Personal Suspicion Factor) of the owner and the LOTRF (Location and Time Risk Factor) will be retrieved. Then the algorithm will increase the vehicles and owners SR (Suspicion Rate) by the sum of the mentioned VSF, PSF and LOTRF values.

For instance, a vehicle pass through an ANPR system and an ANPR record is created accordingly, this data must be analysed by our algorithm in real or near-realtime to be effective. After the detection, the algorithm calculates the sum of the vehicles VSF, owners PSF, and the LOTRF value of vehicle's location, then it will increase the Suspicion Rate of both vehicle and the owner by the calculated value. The LOTRF value will be calculated according to time and location information that exists in the ANPR records. Steps for ANPR analysis represented in Figure 5.2.

We assume that, if a vehicle that has high VSF value, shows a potential that the vehicle could be used for future criminal activities and its movement must be taken seriously, if its SR value reaches its Action Value. We assume that the driver of the car is the same as the owner of the car. We also assume that location is an important factor dealing with the criminal activities that include vehicles, that's why we include the LOTRF in the methodology.

After the manipulation of the vehicle's and the owner's Suspicion Rate, if either the vehicles or the owners SR reaches its respective Action Value, an alarm should be created for both of them and appropriate actions by law enforcement should be taken.

These are main steps for ANPR analysis:

After an ANPR record is created for the vehicle A;

- Step 1: Check the LOTRF value of A's location with respect to time

- Step 2: Increase the SR of the A and owner of the car with the LOTRF value
- Step 3: Check the VSF of A
- Step 4: Increase the SR of the A and owner of the car with the VSF value
- Step 5: Check the PSF of the owner of the car
- Step 6: Increase the SR of the A and owner of the car with the PSF value
- Step 7: If either A or the owner reach their respective Action Value, alarm the law enforcement

Sample pseudocode for ANPR analysis:

```
WHEN an ANPR record is created for the vehicle A:

begin

  #Retrieve the LOTRF value for A's location from the database
  then initialize the variable

  #Location and Time information should be extracted from ANPR
  record

  var_lotrf_for_A = getLOTRFfromArchieveStorage(time and
  location info)

  #Update the SR values of A and the Owner with the LOTRF
  value

  updateSR_PersonObjectStorage(ID of A, var_lotrf_for_A)

  updateSR_PersonObjectStorage(ID of Owner, var_lotrf_for_A)

  #Retrieve the VSF value for A from the database then
  initialize the variable

  var_vsf_of_A = getVSFfromArchieveStorage(ID of A)
```

```
#Update the SR values of A and the Owner with the VSF value

updateSR_PersonObjectStorage(ID of A, var_vsf_of_A)

updateSR_PersonObjectStorage(ID of Owner, var_vsf_of_A)

#Retrieve the PSF value of the Owner from the database then
initialize the variable

var_psf_of_owner = getPSFfromArchieveStorage(ID of Owner)

#Update the SR values of A and the Owner with the PSF value

updateSR_PersonObjectStorage(ID of A, var_psf_of_owner)

updateSR_PersonObjectStorage(ID of Owner, var_psf_of_owner)

#Retrieve the Action Value and updated Suspicion Rate from
the database for A and the Owner

var_av_of_A = getAVfromPersonObjectStorage(ID of A)

var_av_of_owner = getAVfromPersonObjectStorage(ID of Owner)

var_sr_of_A = getSRfromPersonObjectStorage(ID of A)

var_sr_of_owner = getSRfromPersonObjectStorage(ID of Owner)

#Check if either A or the Owner reach their respective
Action Value

If var_sr_of_A greater than or equal to var_av_of_A

  Send alarm to the law enforcement for A and the Owner

endif

If var_sr_of_owner greater than or equal to var_av_of_owner
```

```
    Send alarm to the law enforcement for A and the Owner
  endif
end
```

ANPR analysis for suspicion detection:

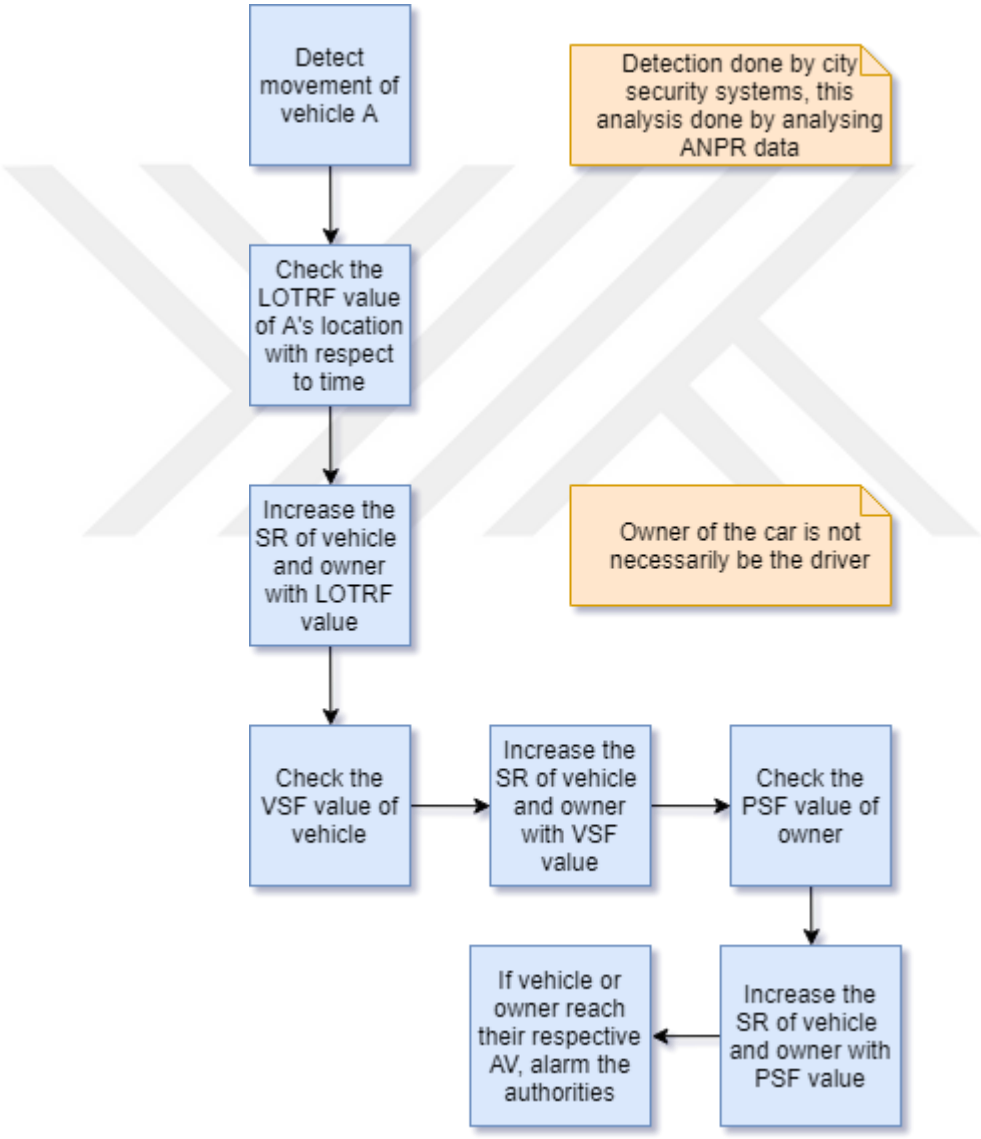

Figure 5.2 ANPR data analysis for suspicion detection

5.3 API/PNR- Advance Passenger Information/Passenger Name Records

Advance Passenger Information and Passenger Name Records data show us the movement of passengers regionally and around the globe, they have attributes that include information about passenger's personal data, booking details, travel documents, etc. In our methodology after the detection of the movement, passenger's PSF (Personal Suspicion Factor) will be retrieved. Then the algorithm will increase the passenger's SR (Suspicion Rate) by his/her PSF value. Steps for API/PNR analysis represented in Figure 5.3.

For example, a passenger data either API or PNR record is created and transferred from airline authorities to law enforcement units, this record must be delivered and analysed by our algorithm in real or near-realtime to be effective. After the detection, the algorithm retrieves the PSF value for the passenger then increases his/her SR with PSF value. We assume that a passenger with a high PSF value is a potential security concern both for the airliners and the destinations that passenger is traveling to.

For instance, in recent years, Foreign Terrorist Fighters (FTF) use air travel to move to the intended destination for terrorism. We can define FTF's as "individuals who travel to a State other than their States of residence or nationality for the purpose of the perpetration, planning, or preparation of, or participation in, terrorist acts or the providing or receiving of terrorist training" [17]. FTF's can represent great security threat either for their destination and to the airliner, such as planting an improvised explosive device (IED) in aircraft with the intention of detonating the aircraft mid-air [17].

Because of the reasons we mentioned, we assume if the passenger's Suspicion Rate reaches his/her Action Value, an alarm should be created for the passenger, then an appropriate and decisive action should be taken at the airport. We believe our algorithm could improve and support the security of the passengers and the air travel industry.

These are main steps for API/PNR analysis:

After an API/PNR record is created for passenger A;

- Step 1: Check PSF of A
- Step 2: Increase the SR of the A with PSF value
- Step 3: If passenger A reaches his/her respective Action Value, alarm the law enforcement

Sample pseudocode for API/PNR analysis:

```
WHEN an API/PNR record is created for the passenger A:
begin
  #Retrieve the PSF value for the A from the database then
  initialize the variable
  var_psf_of_A = getPSFfromArchieveStorage(ID of A)
  #Update the SR value of A with the PSF value
  updateSR_PersonObjectStorage(ID of A, var_psf_of_A)
  #Retrieve the Action Value and updated Suspicion Rate from
  the database for A
  var_av_of_A = getAVfromPersonObjectStorage(ID of A)
  var_sr_of_A = getSRfromPersonObjectStorage(ID of A)
  #Check if A reaches his/her their Action Value
  If var_sr_of_A greater than or equal to var_av_of_A
    Send alarm to the law enforcement for A
```

```
  endif
end
```

API/PNR analysis for suspicion detection:

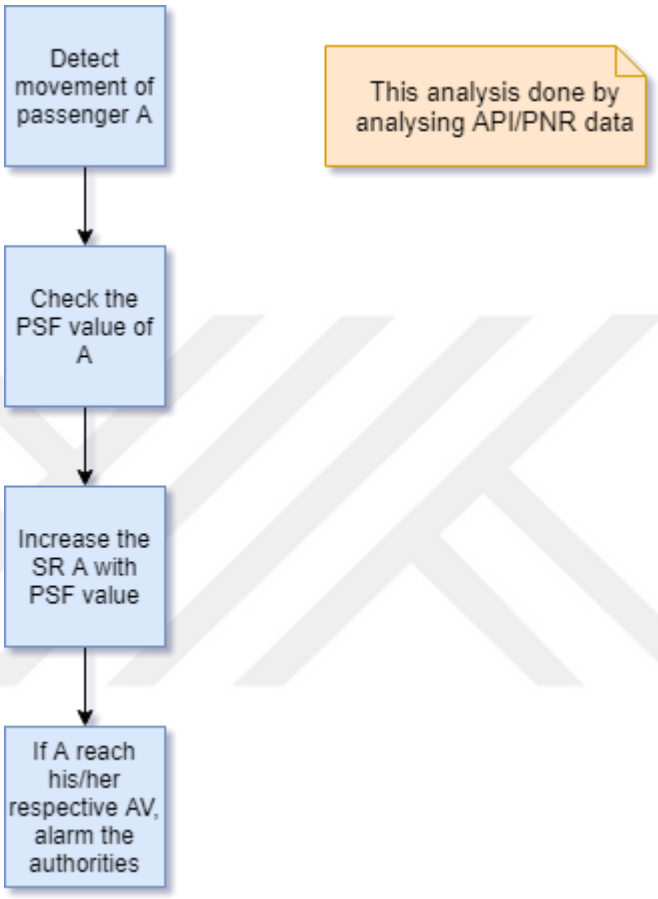

Figure 5.3 API/PNR data analysis for suspicion detection

5.4 TASF-Togetherness Analysis Suspicion Factor

For detecting togetherness, the algorithm must do analysis on CDR and ANPR data constantly and in real time. We assume a group of people with high PSF values, or a group of vehicles with high VSF values in total is an indicator of likely criminal activity. TASF value is the sum of the PSF values of the individuals or sum of the VSF values of the vehicles that move together. When the algorithm detects togetherness within a group of people or vehicles, it will increase their SR with the calculated TASF value. Then if any of the individual or vehicle within the detected group reaches its

Action Value, an alarm for all of the group of individuals or vehicles should be created and law enforcement should intercept the group's activities. Steps for TASF analysis using ANPR and CDR data represented in Figure 5.4 and Figure 5.5.

To detect togetherness of vehicles, COINCIDENT (Closed Travelling Companion Discovery on ANPR Data Stream) algorithm can be used to instantly and continuously discover vehicles that move together as travelling companions on live data stream effectively in near real time [10]. Then our algorithm can continue by checking VSF values of vehicles within the group, calculate TASF than manipulate Suspicion Rate of vehicles.

To detect the togetherness of people, we propose to use CDR data. The main reason is it shows an interaction between people through telecom systems and has location information. There are studies about finding criminal networks using archive CDR data, that proposes graph based models [18]. Our algorithm has to do analysis in real time and automatically, instead of a visual analysis done by law enforcement. Our algorithm could use clustering techniques on mobile users based on location information in CDR data, then combining it with communication (interaction) data within the CDR to detect togetherness in real time or near-real time. We believe by the above mentioned approach, togetherness of individuals based on their location and interaction with each other could be detected. Then our algorithm could check all of the PSF values of the detected group of people, calculate TASF value and manipulate their Suspicion Rate.

We believe this approach is important because of the realities of organized crime organizations. There are many types of organized criminals all around the world, for instance biker gangs which are involved in many criminal activities such as illegal drug distributions, firearms violations, racketeering, extortion [19] etc. or street gangs that involved in robberies, assaults, drug distribution and gang fights [20] or terrorist organizations. Such organized criminals mainly operate in urban areas and disrupt the safety and stability of cities.

Because of the reasons we mentioned above, we assume after togetherness is detected for a group of individuals or vehicles and the SR manipulation, if any of the individual

or vehicle reaches its Action Value within those groups, we should alarm the law enforcement about the incident. We believe our algorithm will be a great asset for the law enforcement to tackle organized criminal activities thus increases the safety and security of citizens.

These are main steps for Togetherness (TASF) analysis:

After detection of togetherness of individuals from the CDR data;

- Step 1: Check all the PSF values of the group of individuals
- Step 2: Calculate the TASF (sum of all PSF values)
- Step 3: Increase the SR of all individuals in the group by the TASF value
- Step 4: If any of the individual within the group reaches his/her AV, alarm the law enforcement

After detection of togetherness of vehicles from ANPR data;

- Step 1: Check all the VSF values of the group of vehicles
- Step 2: Calculate the TASF (sum of all VSF values)
- Step 3: Increase the SR of all vehicles in the group by the TASF value
- Step 4: If any of the vehicle within the group reaches its Action Value, alarm the law enforcement

Sample pseudocode for TASF analysis:

```
WHEN a togetherness detected for a group of individuals from
the CDR data:

begin

  #Create a variable for TASF value

  var_tasf
```

```
#Loop through each member of the group to calculate TASF
value

for each member of the group do

  #Retrieve the PSF value of the individual A from the
  database then initialize the variable

  var_psf_of_A = getPSFfromArchieveStorage(ID of A)

  #Add the PSF value to the TASF value

  var_tasf += var_psf_of_A

endfor

#Loop through each member of the group to increase their SR
by the TASF value and check their SR and AV, if any of the
individuals within the group reaches his/her AV, alarm the
law enforcement

for each member of the group do

  #Increase the individual A's SR by the TASF value

  updateSR_PersonObjectStorage(ID of A, var_tasf)

  #Retrieve the Action Value and updated Suspicion Rate
  from the database for A

  var_av_of_A = getAVfromPersonObjectStorage(ID of A)

  var_sr_of_A = getSRfromPersonObjectStorage(ID of A)

  #Check if A reaches his/her Action Value

  If var_sr_of_A greater than or equal to var_av_of_A

    Send alarm to the law enforcement for all of the group
```

```
    endif
  endfor
end
```

WHEN a togetherness detected for a group of vehicles from the ANPR data:

```
begin
  #Create a variable for TASF value
  var_tasf
  #Loop through each member of the group to calculate TASF value
  for each member of the group do
    #Retrieve the VSF value of the vehicle A from the database
    then initialize the variable
    var_vsf_of_A = getVSFfromArchieveStorage(ID of A)
    #Add the VSF value to the TASF value
    var_tasf += var_vsf_of_A
  endfor
  #Loop through each member of the group to increase their SR
  by the TASF value and check their SR and AV, if any of the
  vehicles within the group reaches its AV, alarm the law
  enforcement
  for each member of the group do
    #Increase the vehicle A's SR by the TASF value
```

```
    updateSR_PersonObjectStorage(ID of A, var_tasf)

    #Retrieve the Action Value and updated Suspicion Rate from
    the database for A

    var_av_of_A = getAVfromPersonObjectStorage(ID of A)

    var_sr_of_A = getSRfromPersonObjectStorage(ID of A)

    #Check if A reaches its Action Value

    If var_sr_of_A greater than or equal to var_av_of_A

      Send alarm to the law enforcement for all of the group

    endif

  endfor

end
```

TASF analysis with ANPR data for suspicion detection:

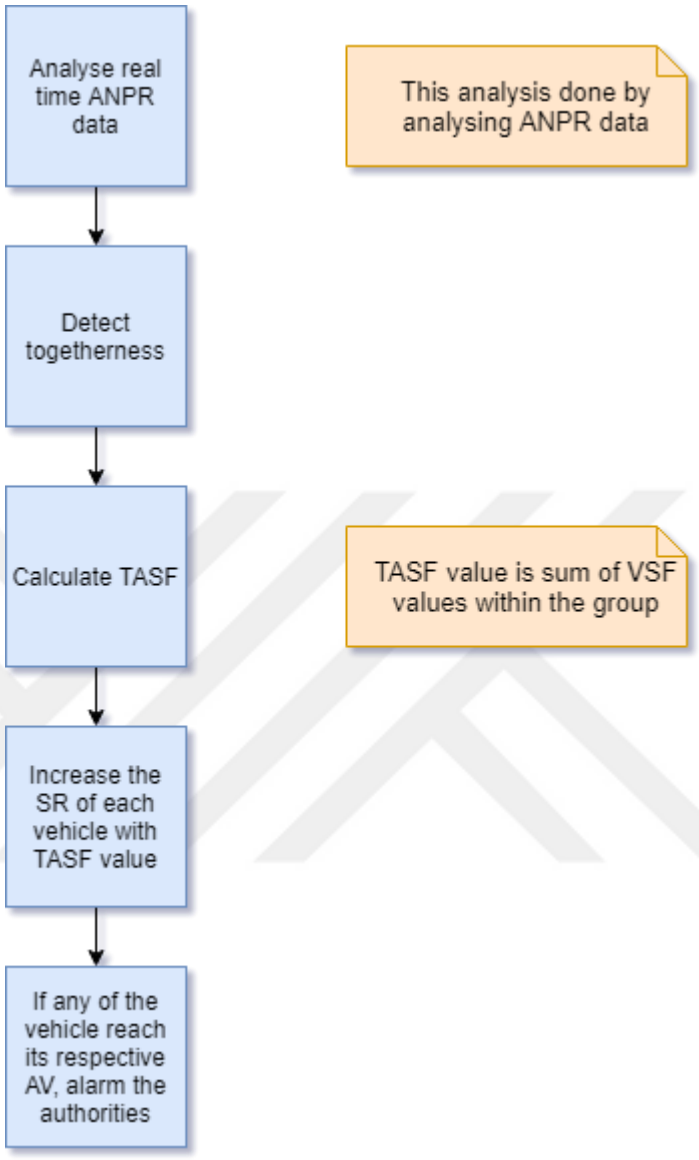

Figure 5.4 Calculation of TASF from ANPR data and suspicion detection

TASF analysis with CDR data for suspicion detection:

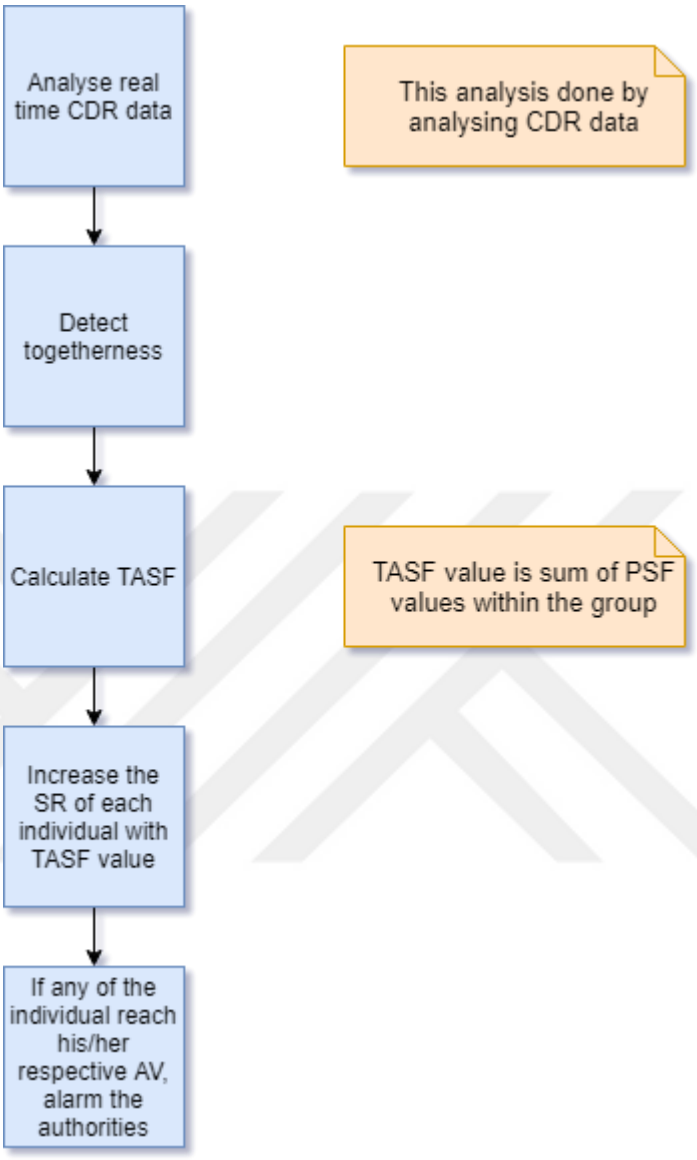

Figure 5.5 Calculation of TASF from CDR data and suspicion detection

5.5 Overall Detection Methodology and System Architecture

Overall, in our algorithm all the events will cause an increase in Suspicion Rate, as we mentioned before our basic assumption is; to commit a crime, a movement or an activity is needed. There is only one way to reduce the Suspicion Rate and it is dependent on the time period that there is no increase in SR. When the SR for an individual or an object reaches its Action Value, an alarm will be created for law

enforcement. The overall architecture of the algorithm is represented in Figure 5.6 and Figure 5.7. In the methodology, the events that increase the SR are as follows;

- Detection in CDR analysis
- Detection in ANPR analysis
- Detection in API/PNR analysis
- Detection of togetherness in TASF analysis

The above mentioned four analysis types will increase SR, but the algorithm is extendable and not limited to those four, we can introduce more real-time data and analysis to the algorithm that can contribute in suspicion detection.

The overall architecture:

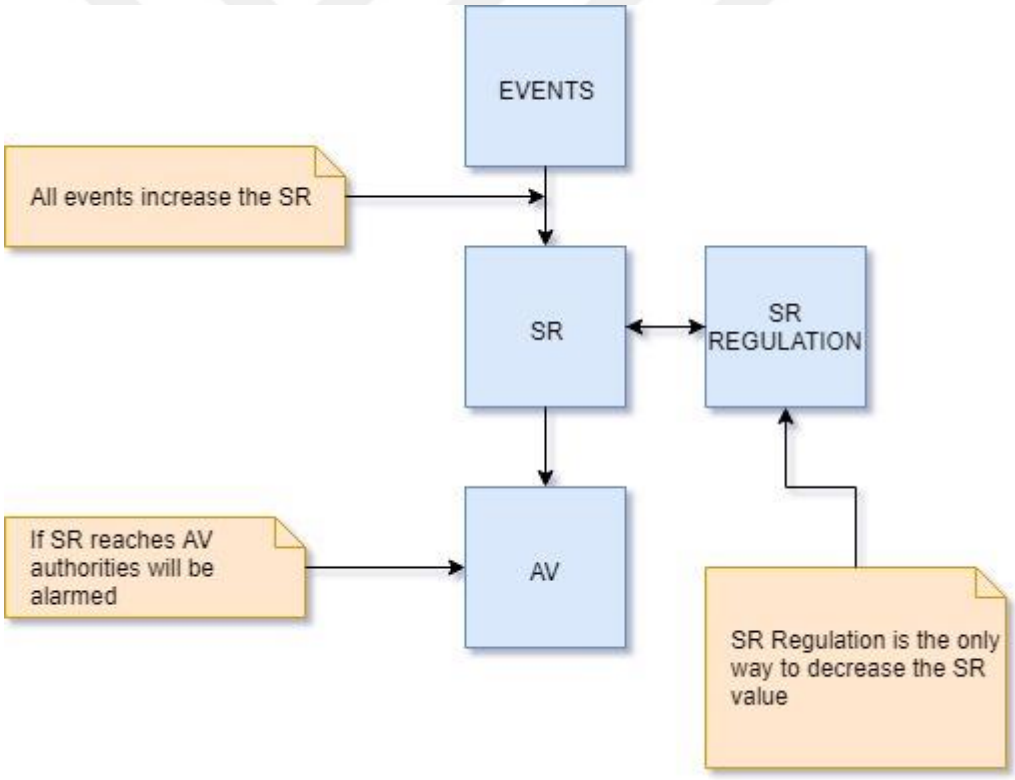

Figure 5.6 Overall architecture

The overall architecture with respect to analyse types:

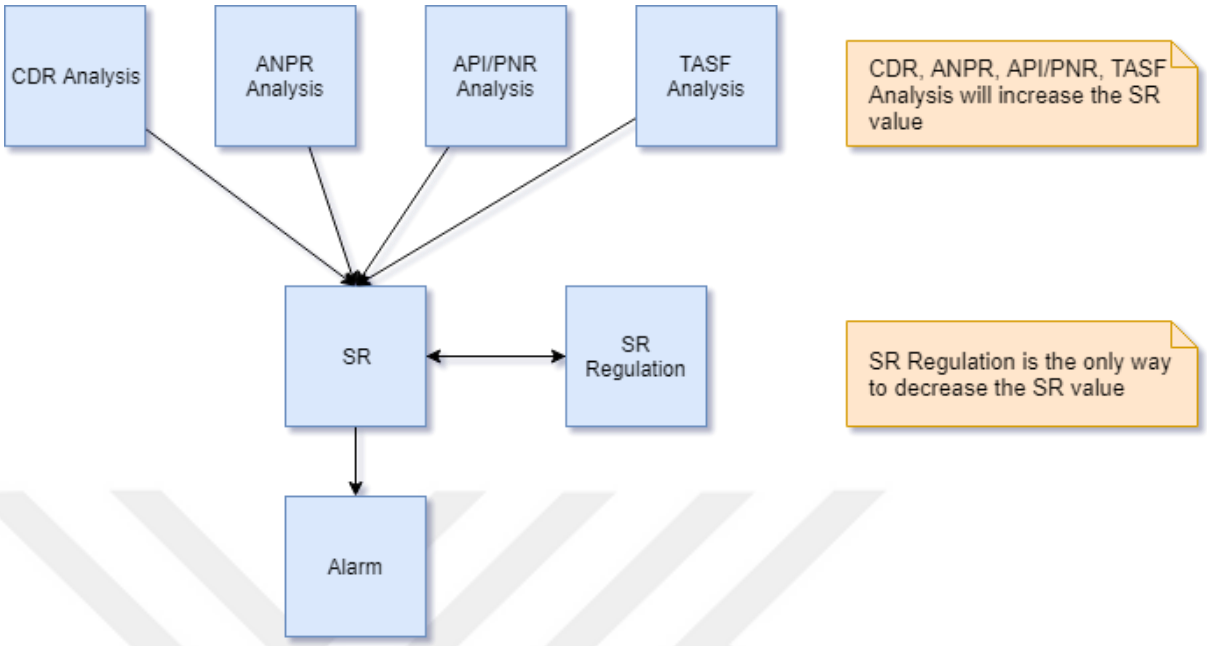

Figure 5.7 Architecture of the algorithm with respect to analyse types

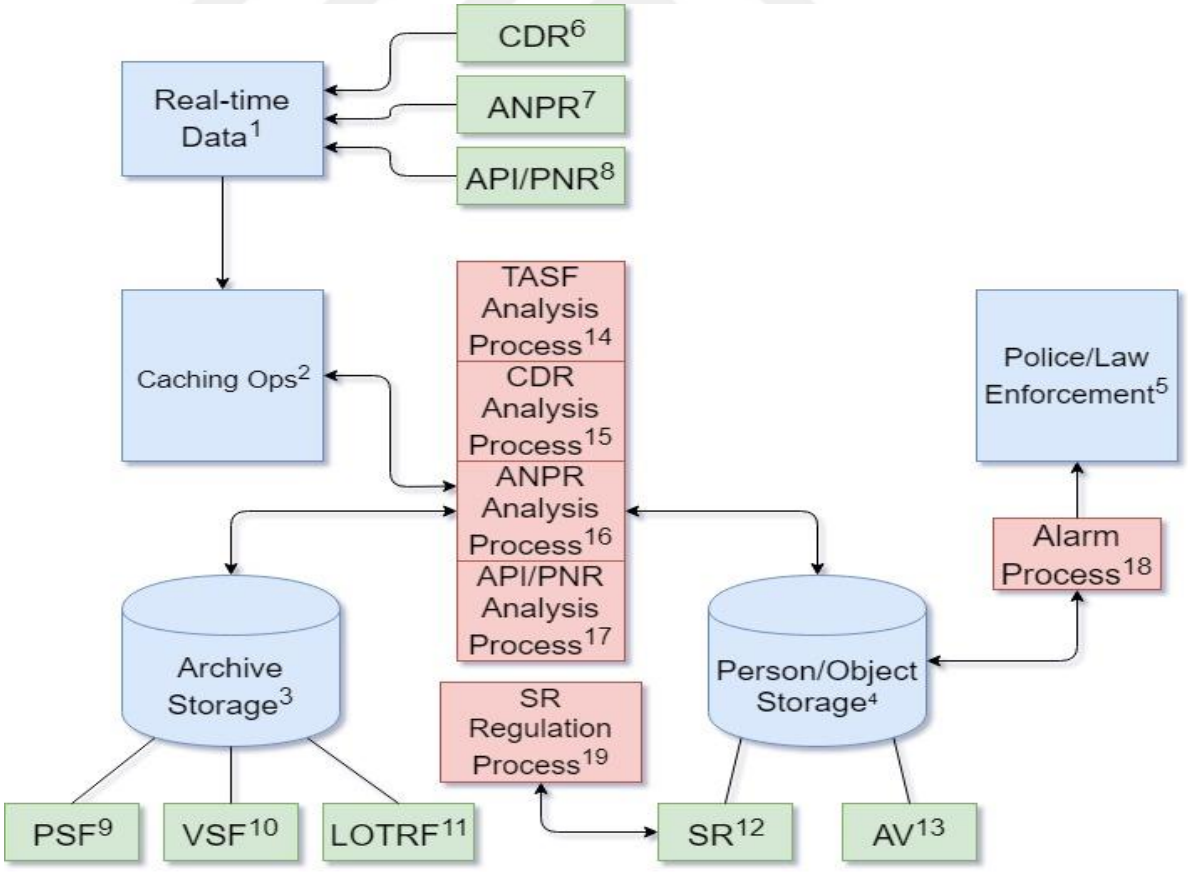

Figure 5.8 Architecture of the proposed system

In Figure 5.8 we represent and propose the system architecture for our detection methodology. These are the parts of the architecture;

1. **Real-time Data**: The real-time data come to the system from multiple sources and organizations. For instance telecom companies, airliners, traffic management authorities. The data will be sent to the memcache system.
2. **Caching Ops**: We need to analyse real-time data and because of that we don't have to store it for a long period of time. We suggest storing the data from 15 to 30 minutes. After that period the data will be deleted. For memcache, we suggest to use Redis. But any kind of in-memory data structure can be used as a temporary database for the real-time data. To host the memcache virtualization technology could be used for better resource management.
3. **Archive Storage**: In Archive Storage, three types of data that we created for the algorithm should be stored. Those are PSF (Personal Suspicion Factor), VSF (Vehicle Suspicion Factor) and LOTRF (Location and Time Risk Factor). For storage technology, any kind of Relational or NoSQL database can be used such as Oracle, PostgreSQL, MongoDB, etc.
4. **Person/Object Storage**: In Person/Object Storage two types of data should be stored that we will manipulate and acted on for each individual and object. Those are SR (Suspicion Rate), AV (Action Value). For storage technology any kind of Relational or NoSQL database can be used such as Oracle, PostgreSQL, MongoDB, etc.
5. **Police/Law Enforcement**: In the system, we need to have a connection to the police and law enforcement forces constantly so we can deliver the alarms that's created according to individuals and objects activities.
6. **CDR**: CDR (Call Detail Record) data come to the system from telecom companies. And it will be stored in the memcache temporarily and will be analysed by TASF and CDR analysis processes.
7. **ANPR**: ANPR (Automatic Number Plate Recognition) data come from traffic management authorities. And it will be stored in the memcache temporarily and will be analysed by TASF and ANPR analysis processes.

8. **API/PNR**: API/PNR (Advanced Passenger Information/Passenger Name Records) data come from airline companies. And it will be stored in the memcache temporarily and will be analysed by API/PNR analysis processes.
9. **PSF**: PSF (Personal Suspicion Factor) is a data we created for the algorithm, it will be stored in Archive Storage. It is a numerical value, we assume higher PSF value is an indicator for the possible tendency to involve in criminal activities for an individual.
10. **VSF**: VSF (Vehicle Suspicion Factor) is a data we created for the algorithm, it will be stored in Archive Storage. It is a numerical value, we assume higher VSF value is an indicator for the possibility that the vehicle could be involved in future criminal activities.
11. **LOTRF**: LOTRF (Location and Time Risk Factor) is a data we created for the algorithm, it will be stored in Archive Storage. It is a numerical value, it should be created by analysing records of criminal activities. We assume that a high LOTRF value for a location for a specific time represents a high risk of crime that could take place in that location at that time.
12. **SR**: SR (Suspicion Rate) is a numerical value we store in the Person/Object Storage, everything whether an individual or an object has a SR value. SR will be manipulated by TASF, CDR, ANPR, API/PNR Analysis Processes and SR Regulation process. We constantly manipulate SR based on the real-time data that comes to the system.
13. **AV**: AV (Action Value) is a numerical value we store in the Person/Object Storage, everything whether an individual or an object has an AV value. If SR of an individual or an object reaches its AV, the system will alarm the law enforcement.
14. **TASF Analysis Process**: TASF AP is a process that is run constantly to detect togetherness between individuals or objects. It uses the data (CDR and ANPR) that is stored in memcache. If togetherness is detected, TASF AP will calculate TASF value and manipulate the SR of the group of individuals or objects with that value. It can be written by any programming language such as Java and Python.

15. **CDR Analysis Process**: CDR AP is a process that is run constantly and analyse the CDR data which is stored in memcache. CDR AP has access to both Archive Storage and Person/Object Storage. When it detects a connection, it retrieves LOTRF and PSF data from the Archive Storage, then manipulates the SR of the individuals. It can be written by any programming language such as Java and Python.
16. **ANPR Analysis Process**: ANPR AP is a process that is run constantly and analyse the ANPR data which is stored in memcache. ANPR AP has access to both Archive Storage and Person/Object Storage. When it detects a connection, it retrieves LOTRF and VSF data from the Archive Storage, then manipulates the SR of the individuals and vehicles. It can be written by any programming language such as Java and Python.
17. **API/PNR Analysis Process**: API/PNR AP is a process that is run constantly and analyse the API/PNR data which is stored in memcache. API/PNR AP has access to both Archive Storage and Person/Object Storage. When it detects a connection, it retrieves PSF data from the Archive Storage then manipulates the SR of the individuals. It can be written by any programming language such as Java and Python.
18. **Alarm Process**: Alarm Process is a process that is run constantly and compares the SR and AV values of individuals and objects in the Person/Object Storage. When it detects that SR for an individual or an object reaches or surpasses its AV, this process will inform the law enforcement about the incident. It can be written by any programming language such as Java and Python.
19. **SR Regulation Process**: SR Regulation Process run constantly and decreases the SR values in Person/Object Storage, reduction of SR depends on the time period that there is no increase in SR of an individual or an object. It can be written by any programming language such as Java and Python.

CHAPTER 6

PROTOTYPE IMPLEMENTATION

In this section, we present a sample prototype implementation of the methodology. The implementation is primarily done in Python programming language. Architecture of the prototype implementation is represented in Figure 6.1.

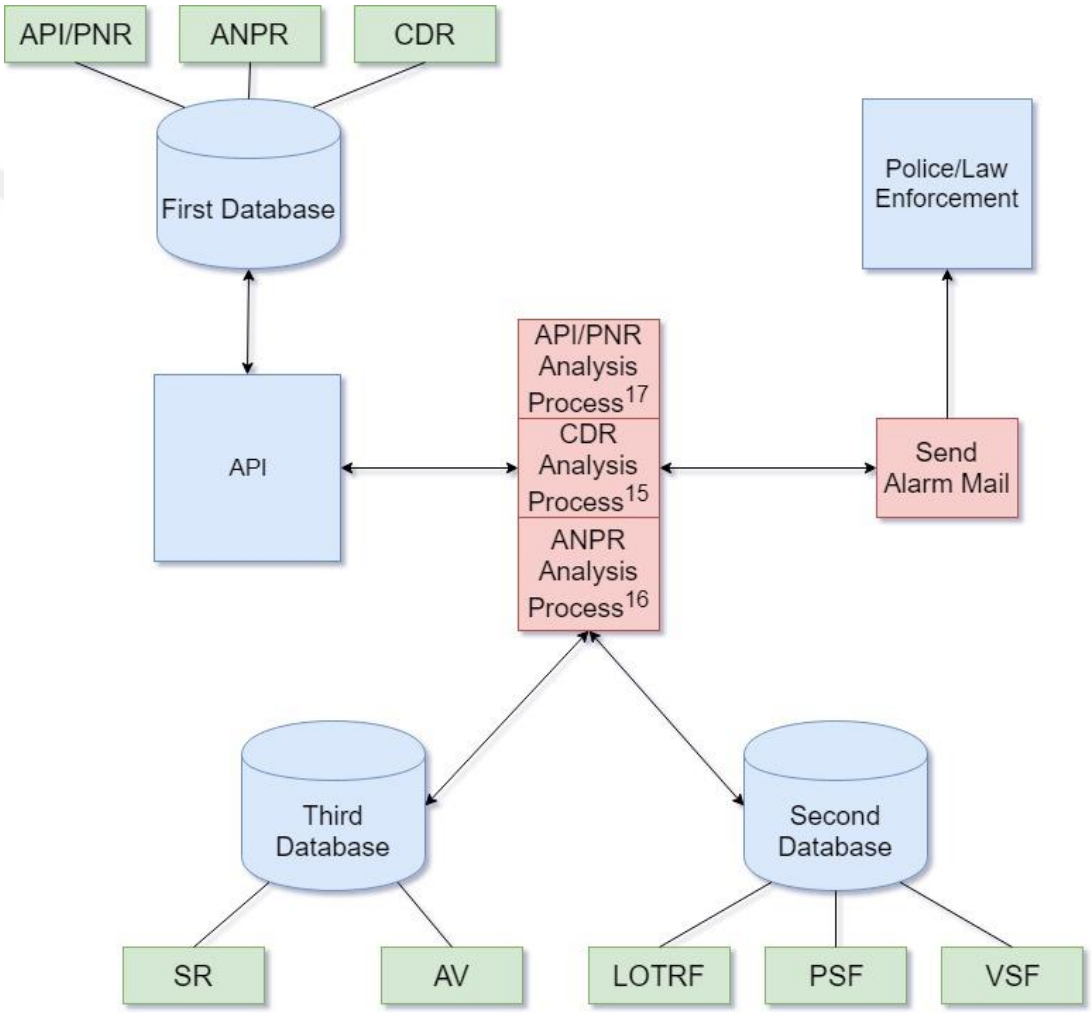

Figure 6.1 Architecture of the prototype implementation

6.1 Database

For database technology, we choose SQLite for rapid development and simplicity. SQLite is one of the most used database engines in the world and it's free to use. In the prototype, we have three different databases. The first database contains the ANPR, API/PNR, and CDR data and has the following tables; ANPR, API_PNR and CDR. The second database contains the LOTRF, PSF and VSF data and has the following tables; LOTRF, PSF and VSF. The third database contains the Action Value and Suspicion Rate data and has the following tables; Action Value and Suspicion Rate. The Entity Relationship Diagram of the databases is represented in Figure 6.2.

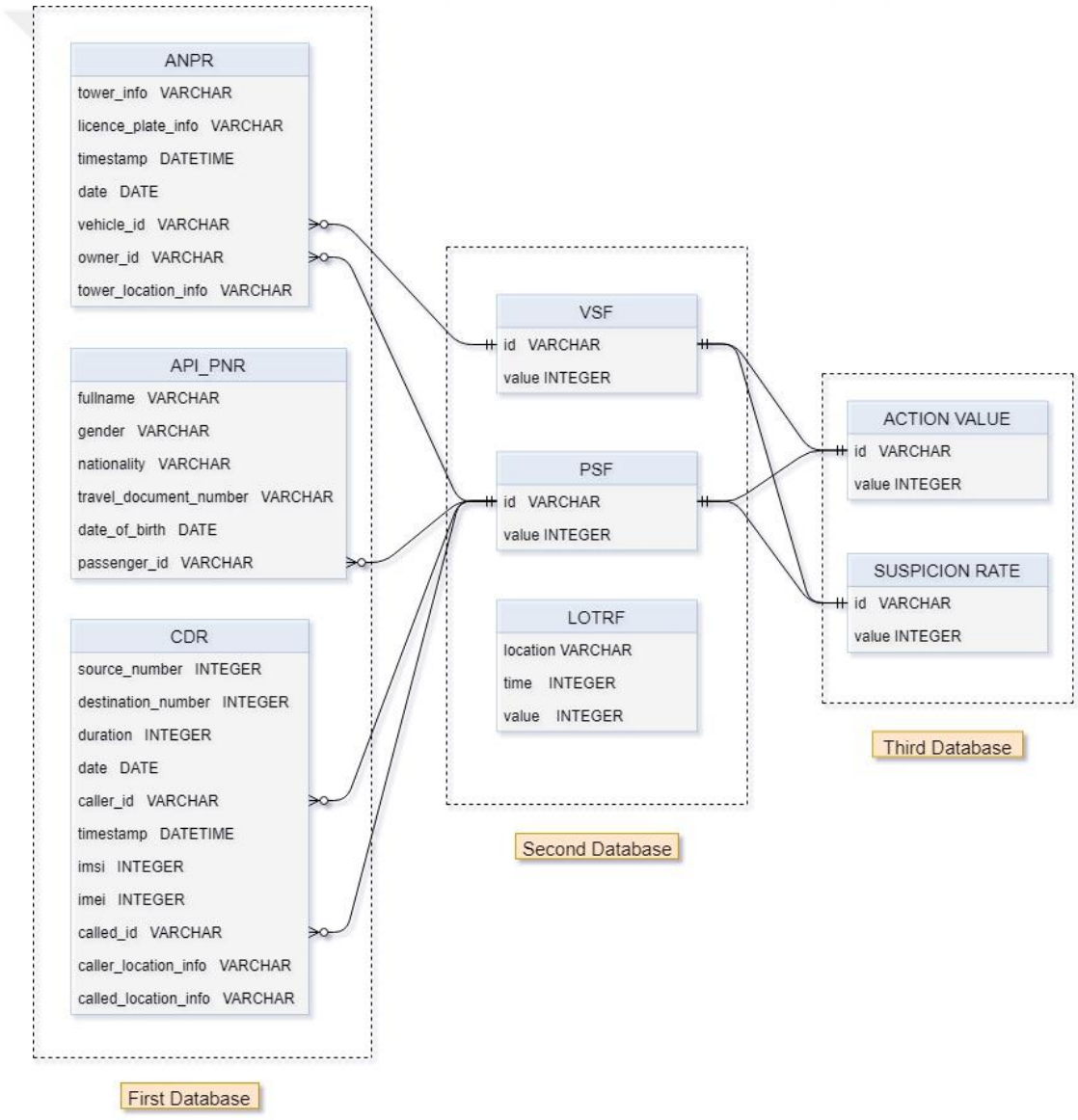

Figure 6.2 Entity relationship diagram of the databases

6.2 API

For the development of API, we decided to implement a RESTful API in Flask Framework. Flask is a micro web framework and it's written in Python. This API returns API, ANPR and API/PNR data. The console output of the API implementation is represented in Figure 6.3.

Implementation code of the API:

```
from flask import Flask, render_template, url_for, request,
redirect, jsonify

from flask_sqlalchemy import SQLAlchemy

from datetime import datetime

import json

import sqlite3

# initialize the app

app = Flask(__name__)

# initialize the database

DB = "./data.db"

def get_data(sql_query):

  conn = sqlite3.connect( DB )

  conn.row_factory = sqlite3.Row

  db = conn.cursor()

  rows = db.execute(sql_query).fetchall()

  conn.commit()
```

```
    conn.close()
    return json.dumps( [dict(ix) for ix in rows] )
# get cdr data
@app.route('/getcdr', methods=['GET'])
def getcdr():
    return get_data("select * from cdr")
# get anpr data
@app.route('/getanpr', methods=['GET'])
def getanpr():
    return get_data("select * from anpr")
# get api/pnr data
@app.route('/getapipnr', methods=['GET'])
def getapipnr():
    return get_data("select * from api_pnr")
if __name__ == "__main__":
    app.run(debug=True)
```

The console output after three requests:

```
$ python api.py
 * Serving Flask app "api" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: on
 * Restarting with stat
 * Debugger is active!
 * Debugger PIN: 314-048-968
 * Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
127.0.0.1 - - [17/Sep/2019 12:16:59] "GET /getcdr HTTP/1.1" 200 -
127.0.0.1 - - [17/Sep/2019 12:17:03] "GET /getanpr HTTP/1.1" 200 -
127.0.0.1 - - [17/Sep/2019 12:17:13] "GET /getapipnr HTTP/1.1" 200 -
```

Figure 6.3 Console output after three requests from the API

In Figure 6.3, workings of the API and logs of three different data requests (CDR, API/PNR and ANPR data requests) are represented.

6.3 Suspicion Analysis

In the prototype, we implemented the API/PNR, ANPR and CDR analysis. Implementations and console outputs of a use case scenario (console outputs are represented in Figure 6.4 to Figure 6.22) are given below.

Sample implementation code of the API/PNR analysis:

```
#retrieval of data

response = requests.get('http://127.0.0.1:5000/getapipnr')

data = response.json()

#analysis of data

for key in data:

  passenger_id = key["passenger_id"]
```

```
var_psf_of_passenger =
getPSFfromArchieveStorage(passenger_id)

updateSR_PersonObjectStorage(passenger_id,
var_psf_of_passenger)

var_av_of_passenger =
getAVfromPersonObjectStorage(passenger_id)

var_sr_of_passenger =
getSRfromPersonObjectStorage(passenger_id)

alarm = False

if var_sr_of_passenger >= var_av_of_passenger:

  alarm = True

if alarm == True:

  content = "A passenger with ID of {} has detected for
  suspicious activities!".format(passenger_id)

  send_alarm.send_email("Suspicion Detected!", content)

else:

  print("\nPassenger has not reached his/her respective
  Action Value...")
```

The console outputs:

```
$ python api-pnr_analysis.py
The retrieved API/PNR data: [{'fullname': 'Sam Wallace', 'gender': 'M', 'nationality': 'USA', 'travel_d

Passenger ID: P991955637

Step 1: Check PSF of passenger

Getting the PSF of passenger...

PSF of the Passenger: 5
```

Figure 6.4 Console output of step 1 of API/PNR analysis

As it is shown in Figure 6.4, we retrieve the API/PNR data from the API via a web request then get the PSF value of the passenger from the database according to API/PNR data we get from the API.

```
Step 2: Increase the SR of the passenger with PSF value

Updating the SR value of passenger with the PSF value...

ID of P991955637's SR Value increased to 5! (previous SR was 0)

Getting the Action Value and updated Suspicion Rate from the database for the passenger...

Passenger ID: P991955637, Action Value: 60, Suspicion Rate: 5
```

Figure 6.5 Console output of step 2 of API/PNR analysis

As it is shown in Figure 6.5, we increase the SR of the passenger with the PSF value then get the passenger's AV and updated SR from the database.

```
Step 3: If passenger reaches his/her respective Action Value, alarm the law enforcement

Checking if passenger reaches his/her respective Action Value...

Passenger has not reached his/her respective Action Value...
```

Figure 6.6 Console output of step 3 of API/PNR analysis

As it is shown in Figure 6.6, we compare the passenger's updated SR and passenger's AV to see if the passenger's SR reaches his/her AV, in this case, the passenger did not reach his/her AV.

Sample implementation code of the ANPR analysis:

```
#retrieval of data

response = requests.get('http://127.0.0.1:5000/getanpr')

data = response.json()

#analysis of data

for key in data:

  vehicle_id = key["vehicle_id"]

  owner_id = key["owner_id"]

  time_info = datetime.strptime(key["timestamp"], '%d/%m/%y
  %H:%M:%S')

  location_info_vehicle = key["tower_location_info"]

  var_lotrf_for_vehicle =
  getLOTRFfromArchieveStorage(time_info.hour,
  location_info_vehicle)

  updateSR_PersonObjectStorage(owner_id,
  var_lotrf_for_vehicle)

  updateSR_PersonObjectStorage(vehicle_id,
  var_lotrf_for_vehicle)

  var_vsf_of_vehicle = getVSFfromArchieveStorage(vehicle_id)

  updateSR_PersonObjectStorage(owner_id, var_vsf_of_vehicle)
```

```
updateSR_PersonObjectStorage(vehicle_id, var_vsf_of_vehicle)

var_psf_of_owner = getPSFfromArchieveStorage(owner_id)

updateSR_PersonObjectStorage(owner_id, var_psf_of_owner)

updateSR_PersonObjectStorage(vehicle_id, var_psf_of_owner)

var_av_of_vehicle = getAVfromPersonObjectStorage(vehicle_id)

var_av_of_owner = getAVfromPersonObjectStorage(owner_id)

var_sr_of_vehicle = getSRfromPersonObjectStorage(vehicle_id)

var_sr_of_owner = getSRfromPersonObjectStorage(owner_id)

alarm = False

if var_sr_of_vehicle >= var_av_of_vehicle:

  alarm = True

if var_sr_of_owner >= var_av_of_owner:

  alarm = True

if alarm == True:

  content = "An individual with ID of {} and a vehicle with
  ID of {} have detected for suspicious
  activities!".format(owner_id,vehicle_id)

  send_alarm.send_email("Suspicion Detected!", content)

else:

  print("\nNeither of Vehicle or Owner reaches their
  respective Action Value...")
```

The console outputs:

```
$ python anpr_analysis.py
The retrieved ANPR data: [{'tower_info': 'TWR859774059', 'licence_plate_info': '56SD14', 'timestamp': '
 'LOC245251046'}]

Owner ID: P991955637, Vehicle ID: V764260552

Step 1: Check the LOTRF value of Vehicle's location with respect to time

Getting the LOTRF value of Vehicle's location with respect to time...

LOTRF value of the Vehicle's location: 10
```

Figure 6.7 Console output of step 1 of ANPR analysis

As it is shown in Figure 6.7, we retrieve the ANPR data from the API via a web request then get the LOTRF value of the vehicle's location from the database according to ANPR data we get from the API.

```
Step 2: Increase the SR of the A and owner of the car with the LOTRF value

Updating the SR values of Owner and Vehicle's with the LOTRF value...

ID of P991955637's SR Value increased to 15! (previous SR was 5)

ID of V764260552's SR Value increased to 10! (previous SR was 0)
```

Figure 6.8 Console output of step 2 of ANPR analysis

As it is shown in Figure 6.8, we increase the SR of the vehicle and the owner with the LOTRF value.

```
Step 3: Check the VSF of Vehicle

Getting the VSF of Vehicle...

VSF of the Vehicle: 1
```

Figure 6.9 Console output of step 3 of ANPR analysis

As it is shown in Figure 6.9, we get the vehicle's VSF value from the database.

```
Step 4: Increase the SR of the A and owner of the car with the VSF value

Updating the SR values of Owner and Vehicle's with the VSF value...

ID of P991955637's SR Value increased to 16! (previous SR was 15)

ID of V764260552's SR Value increased to 11! (previous SR was 10)
```

Figure 6.10 Console output of step 4 of ANPR analysis

As it is shown in Figure 6.10, we increase the SR of the vehicle and the owner with the VSF value.

```
Step 5: Check the PSF of the owner of the car

Getting the PSF of Owner...

PSF of the Owner: 5
```

Figure 6.11 Console output of step 5 of ANPR analysis

As it is shown in Figure 6.11, we get the owner's PSF value from the database.

```
Step 6: Increase the SR of the vehicle and owner of the car with the PSF value

Updating the SR values of Owner and Vehicle's with the PSF value...

ID of P991955637's SR Value increased to 21! (previous SR was 16)

ID of V764260552's SR Value increased to 16! (previous SR was 11)

Getting the Action Value and updated Suspicion Rate from the database for Owner and Vehicle...

Owner ID: P991955637, Action Value: 60, Suspicion Rate: 21

Vehicle ID: V764260552, Action Value: 20, Suspicion Rate: 16
```

Figure 6.12 Console output of step 6 of ANPR analysis

As it is shown in Figure 6.12, we increase the SR of the vehicle and the owner with the PSF value then get the vehicle's and the owner's AV and updated SR from the database.

```
Step 7: If either vehicle or the owner reach their respective Action Value, alarm the law enforcement

Checking if either Vehicle or Owner reaches their respective Action Value...

Neither of Vehicle or Owner reaches their respective Action Value...
```

Figure 6.13 Console output of step 7 of ANPR analysis

As it is shown in Figure 6.13, we compare the vehicle's and the owner's updated SR and AV to see if either vehicle or owner reaches their AV. In this case, neither the vehicle or owner reaches their respective AV.

Sample implementation code of the CDR analysis:

```
#retrieval of data

response = requests.get('http://127.0.0.1:5000/getcdr')

data = response.json()

#analysis of data

for key in data:

  caller_id = key["caller_id"]

  called_id = key["called_id"]

  var_psf_of_caller = getPSFfromArchieveStorage(caller_id)

  var_psf_of_called = getPSFfromArchieveStorage(called_id)

  time_info = datetime.strptime(key["timestamp"], '%d/%m/%y
  %H:%M:%S')

  location_info_caller = key["caller_location_info"]
```

```
var_lotrf_for_caller =
getLOTRFfromArchieveStorage(time_info.hour,
location_info_caller)

location_info_called = key["called_location_info"]

var_lotrf_for_called =
getLOTRFfromArchieveStorage(time_info.hour,
location_info_called)

var_sr_inc_for_caller = var_psf_of_caller +
var_lotrf_for_caller

var_sr_inc_for_called = var_psf_of_called +
var_lotrf_for_called

var_sum_of_sr_inc = var_sr_inc_for_caller +
var_sr_inc_for_called

updateSR_PersonObjectStorage(caller_id, var_sum_of_sr_inc)

updateSR_PersonObjectStorage(called_id, var_sum_of_sr_inc)

var_av_of_caller = getAVfromPersonObjectStorage(caller_id)

var_av_of_called = getAVfromPersonObjectStorage(called_id)

var_sr_of_caller = getSRfromPersonObjectStorage(caller_id)

var_sr_of_called = getSRfromPersonObjectStorage(called_id)

alarm = False

if var_sr_of_caller >= var_av_of_caller:

  alarm = True

if var_sr_of_called >= var_av_of_called:
```

```
        alarm = True

    if alarm == True:

        content = "Individuals with ID of {} and {} have detected
        for suspicious activities!".format(caller_id,called_id)

        send_alarm.send_email("Suspicion Detected!", content)

    else:

        print("\nNeither of Caller or Called reaches their
        respective Action Value...")
```

The console outputs:

```
$ python cdr_analysis.py
The retrieved CDR data: [{'source_number': 4935404470, 'destination_number': 3628736639, 'duration': 10
7488', 'caller_location_info': 'LOC245251046', 'called_location_info': 'LOC245251046'}]

Caller ID: P991955637, Called ID: P617877488

Step 1: Check the PSF of Caller

Getting the PSF of Caller...

PSF of the Caller: 5
```

Figure 6.14 Console output of step 1 of CDR analysis

As it is shown in Figure 6.14, we retrieve the CDR data from the API via a web request then get the PSF value of the Caller from the database.

```
Step 2: Check the PSF of Called

Getting the PSF of Called...

PSF of the Called: 15
```

Figure 6.15 Console output of step 2 of CDR analysis

As it is shown in Figure 6.15, we get the PSF value of the Called from the database.

```
Step 3: Check the LOTRF value of Caller's location with respect to time

Getting the LOTRF value of Caller's location with respect to time...

LOTRF value of the Caller's location: 10
```

Figure 6.16 Console output of step 3 of CDR analysis

As it is shown in Figure 6.16, we get the LOTRF value of the Caller's location from the database according to CDR data we get from the API.

```
Step 4: Check the LOTRF value of Called's location with respect to time

Getting the LOTRF value of Called's location with respect to time...

LOTRF value of the Called's location: 10
```

Figure 6.17 Console output of step 4 of CDR analysis

As it is shown in Figure 6.17, we get the LOTRF value of the Called's location from the database according to CDR data we get from the API.

```
Step 5: Calculate the SR increase of caller by adding the PSF and the LOTRF value

SR increase for Caller: 15
```

Figure 6.18 Console output of step 5 of CDR analysis

As it is shown in Figure 6.18, we calculate the SR increase of the Caller.

```
Step 6: Calculate the SR increase of called by adding the PSF and the LOTRF value

SR increase for Called: 25
```

Figure 6.19 Console output of step 6 of CDR analysis

As it is shown in Figure 6.19, we calculate the SR increase of the Called.

```
Step 7: Calculate the sum of the SR increase of A and B, then increase the SR of A and B with the calculated value

Sum of SR increase for both parties: 40

Updating the SR values of Caller and Called's with the sum of their SR increase...

SR value of Person with ID of P991955637 increased to 61! (previous SR was 21)

SR value of Person with ID of P617877488 increased to 40! (previous SR was 0)

Getting the Action Value and updated Suspicion Rate from the database for Caller and Called...

Caller ID: P991955637, Action Value: 60, Suspicion Rate: 61

Called ID: P617877488, Action Value: 50, Suspicion Rate: 40
```

Figure 6.20 Console output of step 7 of CDR analysis

As it is shown in Figure 6.20, we calculate the sum SR increases of both parties then increase the SR of the Caller and Called with the calculated value. Then we get the Caller's and the Called's AV and updated SR from the database.

```
Step 8: Check if either caller or called reaches their respective Action Value

Checking if either Caller or Called reaches their respective Action Value...

Caller reaches his/her respective Action Value!
```

Figure 6.21 Console output of step 8 of CDR analysis

As it is shown in Figure 6.21, we compare the Caller's and the Called's updated SR and AV to see if either Caller or Called reaches their AV. In this case, the Caller reaches his/her respective AV.

```
Step 9: Alarm the law enforcement if either caller or called reach their respective AV

Alarming the Law enforcement about the detection...

An alarm E-mail is successfully sent!
```

Figure 6.22 Console output of step 9 of CDR analysis

As it is shown in Figure 6.22, we alarm the law enforcement about the detection of the Caller and Called with an alarm E-mail.

CHAPTER 7

DISCUSSION AND CONCLUSION

In this study, we propose a methodology for detecting suspected activities. In this methodology the basic assumption is; to commit a crime, a movement or an activity is needed. For this study the following data types are mentioned; CDR (Call Detail Record), ANPR (Automatic Number Plate Recognition), API/PNR (Advanced Passenger Information/Passenger Name Records). These three data types have one thing in common they represent a movement, and the methodology regulates the Suspicion Rates of individuals and objects accordingly. With our study, we proposed a model and we demonstrated a prototype implementation that many different data types from different domains can be analysed with in the same platform and this platform can be used to help security problems of many countries worldwide. Moreover, we demonstrated that our architecture can support big data analytics for investigating criminals and preventing criminal activities and supporting law enforcement to enhance their tactics and procedures.

The methodology is pretty flexible and it is not bound to only above mentioned data types for SR manipulation. It can be extended and updated according to needs by introducing new data types or changes in technology, such as 5G cellular network technology. In future work, we can enhance this methodology with machine learning techniques and we can apply it to big scale analysis to solve security problems more effectively and precisely.

We believe that law enforcement agencies should implement this methodology and find a way to fuse different data into it for crime prevention purposes, we believe that this methodology can improve the ability of law enforcement services to counter criminal activities in this era of big data. Combining this methodology with the experience of the law enforcement servicemen could improve the security and safety of the citizens of the world and provide a hopeful insight into the future.

REFERENCES

[1] Saha, B., & Srivastava, D., Data quality: The other face of Big Data. 2014 IEEE 30th International Conference on Data Engineering. Presented at the 2014 IEEE 30th International Conference on Data Engineering (ICDE), IEEE, 1294-1297. doi: 10.1109/icde.2014.6816764, 2014

[2] Bernard, Thomas J., Allot A. Nicolas, Thomas David A., Clarke Donald C., Edge Ian D., "*Crime"*. Retrieved May 6, 2019, from https://www.britannica.com/topic/crime-law

[3] Kelling, George L., Wilson, James Q., "*Broken Windows*". (1982, March). The Atlantic, Retrieved May 6, 2019, from https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/, 1982

[4] Lin, Y.-L., Chen, T.-Y., & Yu, L.-C., Using Machine Learning to Assist Crime Prevention. 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI). Presented at the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), IEEE, 1029-1030. doi: 10.1109/iiai-aai.2017.46, 2017

[5] Yu, H., & Hu, C., A Police Big Data Analytics Platform: Framework and Implications. 2016 IEEE First International Conference on Data Science in Cyberspace (DSC). Presented at the 2016 IEEE First International Conference on Data Science in Cyberspace (DSC), IEEE, 323-328. doi: 10.1109/dsc.2016.84, 2016

[6] Elagib, S. B., Hashim, A.-H. A., & Olanrewaju, R. F., CDR analysis using Big Data technology. 2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE). Presented at the 2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE), IEEE, 467-471. doi: 10.1109/ICCNEEE.2015.7381414, 2015

[7] Magnusson, J., & Kvernvik, T., Subscriber classification within telecom networks utilizing big data technologies and machine learning. Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining Algorithms, Systems, Programming Models and Applications - BigMine ’12. Presented at the 1st International Workshop, ACM Press, 77-84. doi: 10.1145/2351316.2351327, 2012

[8] Khan, S., Ansari, F., Dhalvelkar, H. A., & Computer, S.. Criminal investigation using Call Data Records (CDR) through Big Data technology. 2017 International Conference on Nascent Technologies in Engineering (ICNTE). Presented at the 2017 International Conference on Nascent Technologies in Engineering (ICNTE), IEEE, 1-5. doi: 10.1109/icnte.2017.7947942, 2017

[9] Ju-Chi Tseng et al., "A successful application of big data storage techniques implemented to criminal investigation for telecom," 2013 15th Asia-Pacific Network Operations and Management Symposium (APNOMS), Hiroshima, 2013, pp. 1-3., 2013

[10] Zhu, M., Liu, C., Wang, J., Wang, X., & Han, Y., A Service-Friendly Approach to Discover Traveling Companions Based on ANPR Data Stream. 2016 IEEE International Conference on Services Computing (SCC). Presented at the 2016 IEEE International Conference on Services Computing (SCC), IEEE, 171-178. doi :10.1109/scc.2016.29, 2016

[11] Sun, Y., Zhou, X., Sun, L., & Chen, S., Vehicle Activity Analysis Based on ANPR System. 2014 12th IEEE International Conference on Embedded and Ubiquitous Computing. Presented at the 2014 12th IEEE International Conference on Embedded and Ubiquitous Computing (EUC), IEEE, 89-96. doi: 10.1109/euc.2014.22, 2014

[12] B., C., K.V., K., D., R., & Sandeep, R., Monitoring Traffic Signal Violations using ANPR and GSM. 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC). Presented at the 2017 International Conference on Current Trends in Computer, Electrical,

Electronics and Communication (CTCEEC), IEEE, 341-346. doi: 10.1109/ctceec.2017.8455045, 2017

[13] Mottini, A., & Acuna-Agost, R., Relative Label Encoding for the Prediction of Airline Passenger Nationality. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). Presented at the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), IEEE, 671-676. doi: 10.1109/icdmw.2016.0100, 2016

[14] Chen, S., Zhu, J., Xie, Q., Huang, W., & Huang, Y., Understanding Airline Passenger Behavior through PNR, SOW and Webtrends Data Analysis. 2015 IEEE First International Conference on Big Data Computing Service and Applications. Presented at the 2015 IEEE First International Conference on Big Data Computing Service and Applications (BigDataService), IEEE, 323-328. doi: 10.1109/bigdataservice.2015.48, 2015

[15] Homayounfar, A., Ho, A. T. S., Zhu, N., Head, G., & Palmer, P., Multi-vehicle convoy analysis based on ANPR data. 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011). Presented at the 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), IET, 1-5. doi: 10.1049/ic.2011.0135, 2011

[16] *Uyuşturucu tacirlerinin öncü araçlı önlemini polis bozdu*. Haberturk.com, Retrieved May 8, 2019, from https://www.haberturk.com/denizli-haberleri/67522009-uyusturucu-tacirlerinin-oncu-aracli-onlemini-polis-bozduuyusturucuyu-otomobilin-motoruna, 2019

[17] United Nations Office on Drugs and Crime Vienna. *Investigation, Prosecution and Adjudication of Foreign Terrorist Fighter Cases for South and South-East Asia*, Retrieved May 7, 2019 from https://www.unodc.org/documents/terrorism/Publications/FTF%20SSEA/Foreign_Terrorist_Fighters_Asia_Ebook.pdf, 2018

[18] Kumar, M., Hanumanthappa, M., & Kumar, T. V. S., Crime investigation and criminal network analysis using archive call detail records. 2016 Eighth

International Conference on Advanced Computing (ICoAC). Presented at the 2016 Eighth International Conference on Advanced Computing (ICoAC), IEEE, 46-50. doi: 10.1109/icoac.2017.7951743, 2017

[19] American Based Biker Gangs: International Organized Crime. American Journal of Criminal Justice, 36(3), Springer Nature, 207–215. doi: 10.1007/s12103-011-9104-8, 2011

[20] Klein, M. W., Weerman, F. M., & Thornberry, T. P., Street Gang Violence in Europe. European Journal of Criminology, 3(4), SAGE Publications, 413–437. doi: 10.1177/1477370806067911, 2006

CURRICULUM VITAE

PERSONAL INFORMATION

Name Surname : Hüsrev Abdulcelil KARACABEY
Date of Birth : 11/08/1993
Phone : 5053905022
E-mail : husrevkaracabey@gmail.com

EDUCATION

High School : Adıyaman Fatih Anadolu Lisesi
Bachelor : Ankara Yıldırım Beyazıt University

WORK EXPERIENCE

Deputy Inspector : Turkish National Police

TOPICS OF INTEREST

- Suspicion Detection
- Data analysis
- Policing Software