ICAD: An Intelligent Framework for Real-Time Criminal Analytics and Detection

Raed Abdallah[1], Hassan Harb[2], Yehia Taher[3(✉)], Salima Benbernou[1], and Rafiqul Haque[4]

[1] LIPADE, Université de Paris, Paris, France
{raed.abdallah,salima.benbernou}@u-paris.fr
[2] College of Engineering and Technology, American University of the Middle East, Kuwait City, Kuwait
hassan.harb@aum.edu.kw
[3] DAVID Lab, UVSQ - Université Paris-Saclay, Versailles, France
yehia.taher@uvsq.fr
[4] Intelligencia, R&D Department, Paris, France
rafiqul.haque@intelligencia.fr

Abstract. Criminal investigation plays a vital role nowadays where the law enforcement agencies (LEAs) carry out this critical mission thoroughly and competently. However, such complicated mission involves a broad spectrum of tasks including collecting evidences from various data sources, analyzing them, and eventually identifying the criminals. Particularly, data may be collected by LEAs from telecommunication companies, online money transfer agencies, social media networks, video surveillance systems, bank transactions, and airways companies. LEAs confront various challenges from different fronts regarding criminal investigation. Thus, handling such big and heterogeneous data coming from different sources and recognizing potential suspects in a real-time is becoming a major challenge for LEAs in criminal investigation. In this paper, we propose an end-to-end Intelligent framework, called as ICAD, to help LEAs in Criminal Analytics and Detection. Mainly, ICAD uses cutting-edge technologies (data science and big data tools) as well as ontological models and inference rules to automatically identify suspects and reduce the human intervention in the investigation process. Furthermore, ICAD consists of four phases. The data sources phase in which we take benefits of various data collection sources that are essential in the crime investigation process. The data acquisition phase where data are collected, preprocessed, and stored using data science tools. The model phase in which a criminal-based ontology is defined that semantically integrates and enriches real-time data into useful information. The last phase is the knowledge extraction where a set of inference and reasoning rules are defined and applied over the ontology to detect criminals according to their activities.

Keywords: Crime Investigation · Heterogeneous Data Sources · Data Science · Ontology · Inference Rules

F. Zhang et al. (Eds.): WISE 2023, LNCS 14306, pp. 300–315, 2023.
https://doi.org/10.1007/978-981-99-7254-8_24

1 Introduction

Criminal investigation has become increasingly crucial in light of the growing prevalence of criminal acts and the formation of extremist radicalization groups in contemporary times. Traditionally, investigation processes were primarily conducted manually by law enforcement agencies (LEAs). However, a significant transformation has taken place in recent years with the advent of advanced technologies such as Big Data, the Internet of Things (IoT), and Artificial Intelligence (AI). These cutting-edge technologies have empowered intelligence agencies to establish a smart ecosystem for criminal investigation. The Big Data-driven ecosystem allows LEAs to collect data from numerous sources of varying sizes and speeds, storing them in highly scalable data lakes, and processing and analyzing them in massively parallel computational environments.

Undoubtedly, the challenges associated with criminal investigations arise from the vast amount of data available in different formats and structures from various sources such as telecommunications, money transfers, banks, video systems, travel history, and social media. Managing and analyzing this data manually to extract the necessary information for the investigation process becomes increasingly difficult. However, dealing with such massive, heterogeneous, and rapidly changing data from diverse sources poses a challenge in providing semantically rich information. Combining separated data is often more beneficial than analyzing them in isolation. Consequently, there is a need to design intelligent frameworks that automate data analysis tasks and reduce human intervention, without compromising the quality of the analysis outcomes. Such frameworks are crucial for improving criminal investigations in the present day.

To address the diversity, heterogeneity, and velocity of data sources, we propose an intelligent framework called ICAD (Intelligent Framework for Real-Time Criminal Analytics and Detection). The ICAD framework collects data from potential crime sources, provides a standardized data storage structure, and applies reasoning rules generated by LEA experts to identify suspects and assist in solving crimes. The fundamental idea behind ICAD is to develop an automated end-to-end system that encompasses data acquisition, preprocessing, analysis, information extraction, and obtaining valuable insights while minimizing human effort and intervention. The proposed framework consists of four phases: data sources, data acquisition, data modeling, and knowledge extraction. It leverages various data science tools and data analytics models, particularly ontology and inference rules, to determine crime suspects and individuals involved in criminal activities.

The remainder of this paper is organized as follows: In Sect. 2, we provide an overview of different data sources and techniques found in the literature pertaining to criminal investigation. Section 3 introduces our intelligent framework, ICAD, while detailing the various tools and techniques employed in each phase. In Sect. 4, we discuss the implementation and evaluation of ICAD. Finally, we conclude the paper and provide some perspectives in Sect. 5.

2 Related Work

In recent times, the significance of crime investigation has gained considerable attention from governments and communities due to its profound impact on people's lives and society as a whole. In response, researchers have focused on developing systems that primarily rely on data science and data analytics to identify suspects and criminals by analyzing historical data from various sources. In this section, we will explore existing systems proposed in the literature, along with the data sources utilized and the techniques employed.

Several studies, such as those referenced in [5,9,10], have leveraged social media data analysis for crime investigation. For instance, in [5], the authors highlighted the importance of social media in detecting radicalization and employed semantic web and domain ontologies to automatically process messages and posts on these platforms. They introduced a radicalization-based ontology model that determines indicators of radicalization in social media users and employs inference rules to identify messages related to radicalization. Similarly, in [10], the authors used similar indicators to detect radicalized individuals based on their behavioral patterns. The patterns were extracted by analyzing the date and time of tweets written in multiple languages. For example, ISIS followers are proficient in several languages, primarily Arabic, and another secondary language such as Russian, Turkish, or English. The results showed that a significant proportion of tweets were published from accounts with time zones set to Riyadh and Pacific Time, which are geographical zones associated with ISIS.

In other studies, including those referenced in [4,11,13], researchers focused on analyzing video data obtained from surveillance systems used in crime investigations. In [4], a criminal detection system based on the Sombrero's Theory of Criminology was proposed. This theory aims to predict criminal tendencies based on facial traits. The system utilized deep learning facial recognition models applied to video surveillance data, along with coordinate measuring machine and support vector machine classifiers, to detect and predict whether a person is involved in criminal activities. In [13], the authors presented an approach that analyzed human positions and distances in video scenes, comparing them with trained data to detect criminal behaviors. The proposed approach combined machine learning techniques with coordinate-based methods to identify criminal activities based on various human poses and postures.

Furthermore, other studies, as mentioned in [7,8,12], focused on criminal documents and text analysis provided by law enforcement agencies (LEAs). In [12], AI techniques were integrated with manual criminal investigation methods to aid in the analysis and identification of different types of crimes. The authors introduced a Cognitive Computing enabled Convolution Neural Network (CC-CNN) approach, combined with various learning algorithms, to recognize crime types extracted from unstructured textual data. In [8], a graph-based clustering approach was proposed to extract relationships from criminal data documents. The approach employed natural language processing (NLP) techniques on crime data collected by LEAs to extract and represent crime-related information in a graph using named entities and relations. The graph was built based on a calculated

similarity score, and an unsupervised clustering model was utilized to identify relations between criminological data and uncover criminal patterns. Similarly, in [7], a framework was introduced for the automatic extraction of information from criminal documents. The framework facilitated the discovery, extraction, and classification of reports and documents into named entities, which were then represented in a graph database tailored for the Portuguese language. The framework automated document processing and supported data representation in graphs.

Lastly, researchers in [1,3,6] utilized telecommunications data, specifically call detail records (CDR), to assist in criminal detection. In [3], an unsupervised data procedure was introduced to build a model that employed Neo4j for analyzing user behavior and identifying potential suspects. In [1], a method for tracking calls across multiple mobile networks was proposed to study, analyze, and link suspects to forensic crimes. The method took advantage of the spatio-temporal nature of CDR data to aid in forensic analysis. In [6], the authors employed a Bayesian Classifier to analyze CDR data collected from telecommunication companies in Pakistan and detect criminal or terrorist activities.

These studies demonstrate the diverse approaches employed in crime analytics and detection, utilizing social media data, video analysis, text mining, and telecommunications data. Each approach offers valuable insights and contributes to the overall objective of improving crime investigation techniques.

3 ICAD Framework

Despite the considerable efforts invested in criminal investigation, many of the techniques proposed suffer from several drawbacks. Firstly, they are often not fully automated systems and still require the intervention of law enforcement agencies (LEAs) to identify suspects recognized by the system. Secondly, these techniques typically focus on a single data source or multiple sources with similar data structures when investigating crimes. This limited scope restricts their ability to leverage diverse and heterogeneous data sources. Lastly, the real-time aspect and challenges posed by big data collection, including velocity, volume, and veracity, are not adequately addressed in most existing techniques.

To overcome these limitations, we propose the Intelligent Framework for Real-Time Criminal Analytics and Detection (ICAD). This framework aims to address the aforementioned challenges by enabling real-time data collection from various heterogeneous criminal data sources and identifying individuals potentially involved in criminal, radicalized, or terrorist activities. The general architecture of the ICAD framework is depicted in Fig. 1, which illustrates four distinct phases. Each phase utilizes a range of data science tools and techniques to achieve its objectives.

In the subsequent sections, we will present a motivation scenario that highlights the practical application of the ICAD framework. Additionally, we will provide detailed explanations of each proposed phase, outlining the methodologies and processes employed to enhance criminal analytics and detection capabilities.

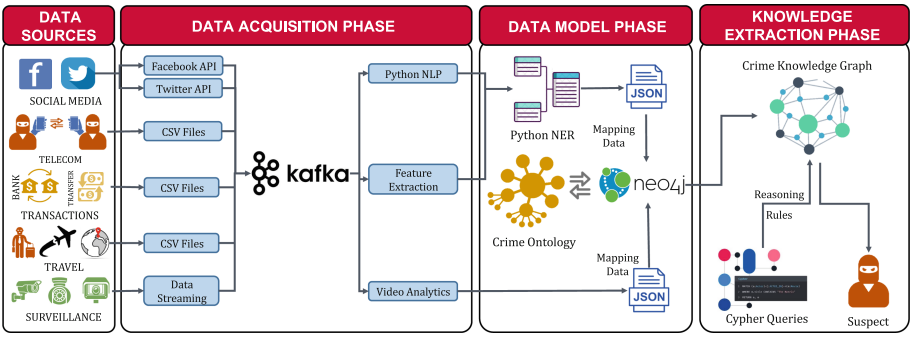

Fig. 1. ICAD architecture.

3.1 Motivating Scenario

Fig. 2 depicts a compelling scenario where our proposed framework can be effectively applied. Let's consider a typical individual named "Tom Harry", who, like many people, maintains multiple social media profiles across various platforms and frequently shares thoughts and ideas. However, authorities have recently discovered something alarming on one of Tom's accounts. Law enforcement agencies (LEAs) have come across posts that endorse the actions of "Al-Baghdadi" and express support for the ideologies of "Daech" (ISIS). This raises suspicions, prompting LEA analysts to delve deeper into Tom's activities to determine if any of them are connected to criminal, radicalized, or terrorist acts.

In their investigation, the LEA analysts stumbled upon a wealth of data about Tom in the transactions dataset of a well-known money transfer company, referred to as OMT. The data revealed that Tom has engaged in substantial financial transactions, both sending and receiving large sums of money to and from Syria, involving multiple individuals. Additionally, through the analysis of video data obtained from various surveillance systems strategically deployed in critical zones within Lebanon, Tom was identified in a region located between Lebanon and Syria. The video footage captured him holding a weapon and engaged in conversation with a suspected individual affiliated with a terrorist organization.

Furthermore, an examination of travel databases unveiled an extensive history of Tom's frequent trips to Syria, often for short periods of time. These recurring visits raised further suspicions and intensified the LEA's focus on investigating Tom. As part of their efforts, the LEA accessed Tom's telecommunications data and uncovered a substantial amount of call detail records (CDRs) indicating prolonged conversations with individuals residing in Syria.

The aforementioned findings have prompted the LEA to continue their comprehensive investigation into Tom's activities, using multiple sources of data such as social media, financial transactions, video surveillance, travel records, and telecommunications data. By meticulously analyzing these different data sets, the LEA aims to ascertain the extent of Tom's involvement and potential connections to criminal or terrorist activities, providing valuable insights for their ongoing investigation.

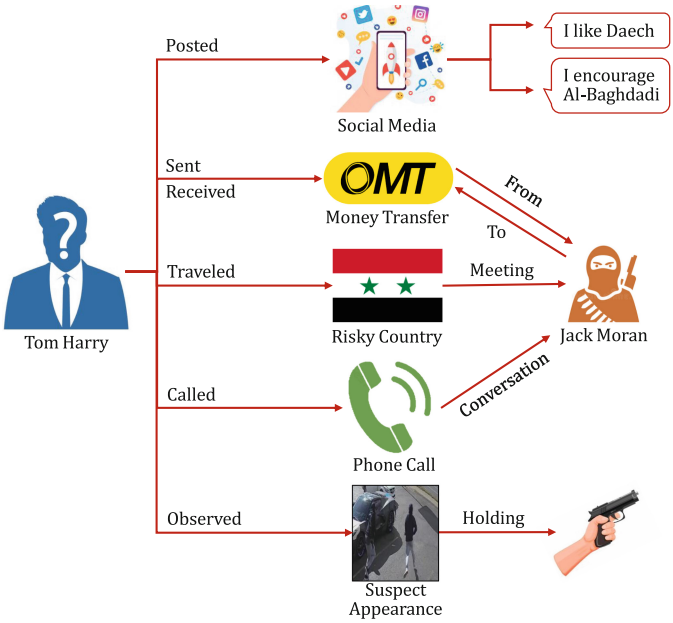

Fig. 2. Motivating scenario for ICAD application.

3.2 Data Sources Phase

ICAD offers several significant advantages, one of which is its comprehensive integration of various data sources in the crime investigation process. The primary objective of ICAD is to accurately identify suspects by tracking their activities in daily life. The following are the key data sources investigated in ICAD:

- **Social media:** In today's world, social media platforms have become crucial sources of data where individuals freely express their thoughts, ideas, locations, and more. As a result, social media platforms provide a wealth of information that can be invaluable in crime investigations.
- **Telecom:** Phone calls often serve as the initial means of communication among suspects involved in crimes, radicalized individuals, or members of terrorist organizations. Consequently, Law Enforcement Agencies (LEAs) can utilize Call Detail Records (CDR) to analyze the timing and content of calls, aiding in the crime analysis process.
- **Money transactions:** Money transactions, typically conducted through banks or global transfer companies, play a significant role in illegal activities. By collecting and analyzing data on financial transactions, LEAs can gain insight into the financial history of a person, detecting any suspicious or illicit activities.
- **Travel data:** Countries are often categorized as risky or non-risky, with the risky ones being prime targets for terrorists. By analyzing travel histories,

investigators can identify potential suspects who frequently travel to risky countries. This information becomes crucial in making informed decisions during the investigation. For example, in the aforementioned scenario, the extensive travel history of "Tom" to Syria to meet suspected individuals was instrumental for LEAs in their crime investigation process.
- **Surveillance systems:** Many cities have implemented surveillance systems for security purposes and to enhance various services such as traffic management and public health. The video data collected by these systems can assist LEAs in tracking suspect activities and movements. Additionally, surveillance footage can help identify the presence of weapons or other risky objects that may be relevant to the investigation.

By leveraging these diverse data sources, ICAD aims to provide a comprehensive and multi-dimensional approach to crime investigation, enhancing the ability of law enforcement agencies to effectively identify and apprehend suspects.

3.3 Data Acquisition Phase

The primary objective of this layer is to gather and consolidate semantically meaningful data from diverse data sources that may vary in format and structure. The purpose behind this data acquisition process is to provide law enforcement agencies (LEAs) with enhanced and valuable information. The data acquisition layer is carried out in two distinct stages, which are outlined below:

Data Collection. The data collection phase of the ICAD framework involves gathering data from various sources. The specific data collection methods employed depend on the format of the data stored in these sources. The following approaches are utilized:

- Facebook Graph and Twitter APIs: These APIs are used to collect data from social media platforms. This enables the retrieval of relevant information and activities from these platforms.
- CSV files: Data from telecommunication, transactions, and travel sources, which are typically structured, are gathered using CSV files. This allows for the extraction of data in a consistent and organized manner.
- Web Scraping: Streaming video data obtained from surveillance systems is handled using web scraping techniques. This enables the collection of real-time video data for further analysis.
- Textual data files: The content of police reports and other textual data is stored in text files. These files are accessed to retrieve and process relevant information.

Kafka, a messaging queue, is employed during this stage to store and manage the collected data from various sources. The data is produced into the Kafka messaging queue for subsequent analysis.

Data Analysis. Once the data is collected, the data analysis process begins. The Kafka consumer retrieves data from the message queue and performs the necessary analysis. The analysis techniques applied in ICAD can be categorized into three types:

1. Python NLP (Natural Language Processing): This technique is used to process and extract information from textual data obtained from social media and police reports. It enables the identification and extraction of relevant information from text-based sources.
2. Feature extraction: This technique involves selecting potential features (columns) from telecommunication, transactions, and travel databases. It helps identify relevant data points that are crucial for crime analysis.
3. Video analytics techniques: ICAD employs video analytics techniques, such as object detection (e.g., weapons) and human tracking, for analyzing video data obtained from surveillance systems. These techniques enable the detection and tracking of objects and individuals of interest.

ICAD has a flexible architecture that allows for the integration of existing or novel algorithms related to any type of data analysis. This ensures adaptability and the ability to incorporate new and advanced analysis techniques into the pipeline.

3.4 Data Model Phase

The Data Model Phase forms the core of the ICAD architecture and focuses on constructing a crime-based ontology. This ontology defines relevant concepts used by law enforcement agencies (LEAs) and establishes semantic relationships between them. The proposed ontology has the following characteristics:

1. Hybrid: The ontology takes into account information extracted from all data sources during the previous phase. It incorporates a comprehensive understanding of different aspects of crime analysis.
2. Scalability: The ontology is designed to accommodate the integration of new data sources and concepts developed by LEAs. This ensures that the ontology remains up to date and adaptable to emerging crime types.
3. Simplicity: The concepts and terminologies employed in the ontology are familiar to law enforcement investigators. This facilitates their ability to formulate queries and perform analyses without requiring extensive knowledge of the underlying data sources.

The ontology draws from existing ontologies in the literature and defines four main concepts: person, organization, location, and activity. Additionally, the following concepts and relationships are defined (Fig. 3):

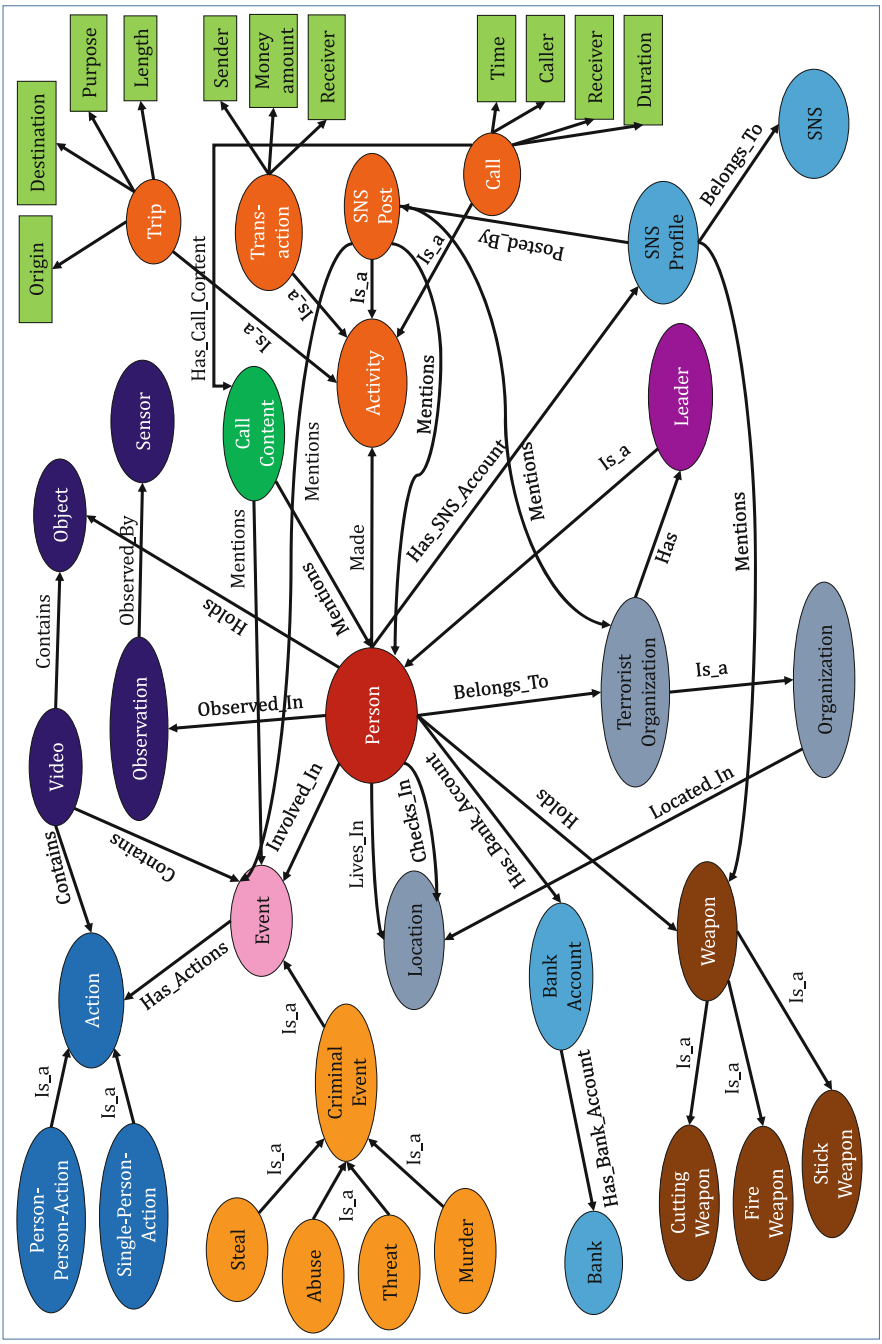

Fig. 3. Crime-domain ontology proposed in ICAD.

- A *person* can engage in various *activities*, such as traveling to another country, conducting financial *transactions*, posting on social media, or making *phone calls*.
- A person may be a member of a *terrorist organization* located in a particular country, which is managed by a *leader*.
- Individuals can be *observed* or detected by *sensors*, and the captured *video* footage may reveal *actions* that indicate a *criminal event*, such as *murder* or *threats*. The video may also provide evidence that the person *possesses* certain types of *weapons*.
- Regular *phone calls* between a person and members of an organization may indicate the occurrence of a *criminal event*.
- A person may have one or multiple *banking accounts* for performing various financial transactions.

3.5 Knowledge Extraction Phase

The process of extracting information from the data acquisition phase plays a vital role in generating a comprehensive knowledge graph by mapping the acquired information to various concepts defined within the ontology. This knowledge graph serves as a valuable resource for further analysis and decision-making. In the subsequent phase of knowledge extraction, semantic reasoning techniques are applied to the knowledge graph, enabling the retrieval of meaningful information and valuable insights.

The application of semantic reasoning adds an additional layer of intelligence to the system, empowering law enforcement agency (LEA) experts and analysts to define inference rules. These inference rules are designed to detect specific scenarios relevant to the criminal investigation process. By formulating these rules, LEA experts can uncover hidden connections, identify patterns, and make informed decisions based on the derived knowledge.

In the scope of this research work, several inference rules have been defined to enhance the effectiveness of the criminal investigation process. These rules are tailored to address distinct scenarios and contribute to the identification and understanding of complex criminal activities. The following inference rules have been established as part of this research:

- *Rule 1*: Person X posts about Terrorist Organization and have several calls to a Country C with a Person Y and X travels to C several times and made several transactions to C and has been observed holding weapon at Location L while meeting Person Y $\rightarrow$ X is a suspect.
- *Rule 2*: Person X calls several times Person Y and Y has been observed meeting X in videos holding weapons and (Y sends money to X and X send money to Y) and Person X is suspected $\rightarrow$ Y is a suspect.
- *Rule 3*: Person X is criminal and Person X supports Person X in a posts or call and Y posts about terrorist organization $\rightarrow$ Y is a suspect.
- *Rule 4*: Person X threats in a post or a call Person Y at Time T and Person Y lives in Place P and the Person X travels to Place P at Time T+1 and

[X observed by videos holding weapons at Place P and Time T+1] → X is a suspect.
- *Rule 5*: Person X threats in a post or call Person Y and been observed by video involved in a Criminal Event (for instance, Murder of Person Y) at Location L at Time T and holding weapon → X is a suspect.
- *Rule 6*: Person X threats by post or call Person Y at Time T and X sends money transaction to Person Z at Time T+1 then Person Z been observed meeting at same Location with Y at Time T+2 and Z was holding a weapon → X and Z are suspects.
- *Rule 7*: Person X threats by post or call Person Y and Person X checks-in at every location Y checked-in and X been observed holding weapons at the locations → X is a suspect.

4 Implementation and Evaluation

In this section, we provide a comprehensive overview of the implementation of each phase proposed within the ICAD framework.

4.1 Data Sources Phase Implementation

In our simulation phase, we employed a high-performance HPE ProLiant ML150 Gen9 Server equipped with a 64-bit 6-core Intel Xeon CPU operating at a clock speed of 1.7 GHz. The server boasted a substantial 64 GB RAM and had a storage capacity of 240 GB SSD (Solid State Drive) complemented by an 8 TB HDD (Hard Disk Drive). To support our simulation environment, we utilized the Windows Server 2012 R2 operating system.

To emulate the various data sources, we created dedicated virtual machines. Synthetic data was generated for three specific domains: telecom, transactions, and travel. For the telecom domain, we simulated telecommunications data reflecting call detail records (CDR), allowing us to analyze and investigate communication patterns. In the transactions domain, we generated artificial transactional data to mimic financial activities for investigative purposes. Lastly, in the travel domain, synthetic travel-related data was produced to simulate passenger itineraries and movements.

For the video surveillance aspect, we implemented a technique proposed in [2] that focused on the detection of knife-related crimes within public areas. This technique aimed to assist law enforcement agencies (LEAs) in minimizing the potential consequences of such incidents. By employing advanced algorithms and video analysis methods, the system could identify and alert authorities about suspicious activities involving knives captured by surveillance cameras.

By utilizing this robust server setup, virtual machines, and synthetic data generation techniques, we were able to simulate realistic scenarios and evaluate the effectiveness of our proposed approaches in crime analytics and detection.

4.2 Data Acquisition Phase Implementation

The implementation of this phase heavily relies on Apache Kafka and Python programming. Apache Kafka serves as the central component for data streaming, and Python is utilized for developing the necessary functionalities. Specifically, two Python-based Jupyter Notebooks are created: *DataProduce.ipynb* and *DataConsumer.ipynb*, each serving as a data producer and consumer, respectively.

In the data producer component, the Python code reads data from various data source files and sends it to the message queue implemented by Kafka. This enables the seamless flow of data from the different sources to the subsequent stages of processing.

On the other hand, the data consumer component receives the data from the message queue, leveraging Kafka's capabilities. Within the *DataConsumer.ipynb* Jupyter Notebook, different data analysis techniques are applied to the received data. These techniques include Natural Language Processing (NLP), feature extraction, and video analytics. All these data analysis techniques are implemented using Python libraries and frameworks.

By employing NLP, textual data can be processed and analyzed to extract meaningful insights and patterns. Feature extraction techniques enable the identification and extraction of relevant features from the data, enhancing subsequent analysis and modeling. Furthermore, video analytics techniques allow for the analysis and interpretation of video data, enabling the detection of specific events or actions of interest.

The combination of Apache Kafka and Python programming within these Jupyter Notebooks facilitates a smooth data flow, from data producers to data consumers, with intermediate processing and analysis steps. This enables efficient data processing, transformation, and extraction of valuable information before forwarding it to the next layer of the system for further processing or decision-making.

4.3 Data Model Phase Implementation

In this phase, the construction of the ontology involved utilizing the Neo4j Graph Database Desktop Application version 4.4.0. The data analysis process in the Kafka consumer generated two distinct types of data. Firstly, there were Named Entity Relationships (NER) derived from social media and CSV files as data sources. Secondly, there were objects detected through video analytics. To ensure compatibility and ease of integration, both types of data were initially converted into JSON format. Subsequently, a mapping process was performed, aligning the converted data with the various concepts defined within the ontology.

By leveraging the capabilities of the Neo4j Graph Database Desktop Application version 4.4.0, the ontology construction phase efficiently incorporated and processed the NER data originating from social media and CSV files, as well as the object data extracted from video analytics. This involved transforming the data into a JSON format, facilitating uniformity and standardization across

different data sources. The subsequent mapping process ensured a seamless integration of the converted data with the defined ontology concepts, enabling comprehensive and meaningful analysis within the ontology framework.

4.4 Knowledge Extraction Phase Implementation

During this phase, we employed the Cypher Query language, which is available in Neo4j, to implement the predefined inference rules. These rules were applied to the knowledge graph generated in the data modeling phase, allowing us to determine if there were any matches with the real-time collected data. The results of these queries were organized and presented in a table format. In cases where matches were found, immediate notifications were sent to law enforcement agencies (LEAs) to ensure timely action.

To showcase the effectiveness of our framework in detecting real-life scenarios, we implemented the same scenario described in section III.A and visualized in Fig. 2 using Neo4j. All the predefined inference rules were applied during this implementation. As a result, we obtained a knowledge graph, as depicted in Fig. 4. This graph represents the synthesized data, integrating information from diverse data sources and mapping it to the specified ontology. The graph serves as a comprehensive representation of the interconnected relationships within the data.

Furthermore, Fig. 5 illustrates the outcomes of applying the inference rules within Neo4j to the motivating scenario. The visualization clearly demonstrates that our proposed framework successfully identified Tom Harry as a criminal suspect. This determination was based on his activities gathered from various data sources, which were integrated and analyzed by our framework.

By leveraging the capabilities of Neo4j and implementing the defined inference rules, our framework proves its efficacy in detecting and identifying potential criminal suspects in real-life situations. The seamless integration of data from multiple sources and the application of inference rules enable efficient and accurate crime detection, empowering law enforcement agencies to take appropriate actions in a timely manner.

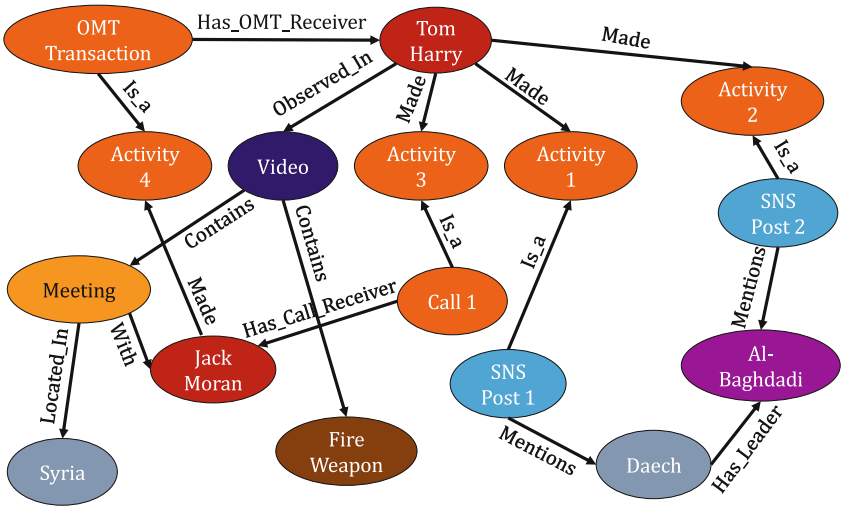

Fig. 4. Knowledge graph resulted from the motivating scenario.

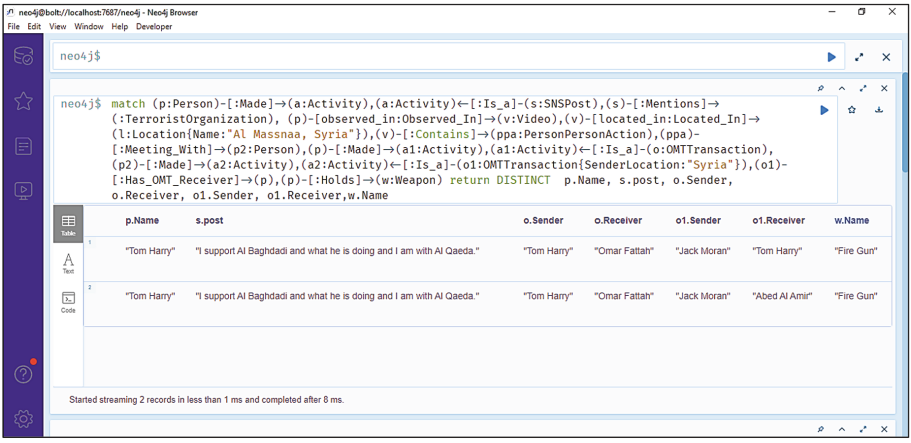

Fig. 5. Inference rules implemented in Neo4j.

5 Conclusion and Future Work

This research paper introduces the Intelligent Framework for Real-Time Criminal Analytics and Detection (ICAD), which offers a comprehensive solution for automating the crime investigation process and assisting law enforcement agencies (LEAs) in identifying suspects. The ICAD framework leverages ontological models and inference rules to enhance suspect recognition. To address various challenges associated with data collection and analysis, ICAD incorporates a range of data science tools and AI techniques across four key phases: data

sources, data acquisition, data modeling, and knowledge extraction. Through the implementation of this proposed architecture, the efficiency of ICAD in detecting criminals has been successfully demonstrated across diverse scenarios.

There are two potential directions for enhancing ICAD. Firstly, the integration of additional data sources can significantly improve the accuracy of suspect identification. For example, incorporating police reports and court documents would provide valuable insights into a suspect's history and help predict their future intentions. These additional sources of information contribute to a more comprehensive and reliable suspect recognition process.

Secondly, studying suspect behaviors and body language can further enhance the effectiveness of ICAD in crime investigations. Criminals often exhibit common traits and physical gestures that can serve as valuable indicators. By incorporating the analysis of these behavioral cues, ICAD can provide deeper insights into suspect profiles and contribute to a more robust investigative process.

By expanding the scope of data sources and incorporating the study of suspect behaviors, ICAD has the potential to become an even more powerful and accurate tool for real-time criminal analytics and detection, thereby strengthening the capabilities of law enforcement agencies in combating crime.

References

1. Abba, E., Aibinu, A., Alhassan, J.: Development of multiple mobile networks call detailed records and its forensic analysis. Digit. Commun. Netw. **5**(4), 256–265 (2019)
2. Abdallah, R., Benbernou, S., Taher, Y., Younas, M., Haque, R.: A smart video surveillance system for helping law enforcement agencies in detecting knife related crimes. In: Awan, I., Younas, M., Bentahar, J., Benbernou, S. (eds.) DBB 2022. LNNS, vol. 541, pp. 65–78. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16035-6_6
3. Abuhamoud, N., Geepalla, E.: A study of using big data and call detail records for criminal investigation. J. Pure Appl. Sci. **18**(4) (2019)
4. Amjad, K., Malik, A.A., Mehta, S.: A technique and architectural design for criminal detection based on Lombroso theory using deep learning. Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol. **4**(3), 47–63 (2020)
5. Barhamgi, M., Masmoudi, A., Lara-Cabrera, R., Camacho, D.: Social networks data analysis with semantics: application to the radicalization problem. J. Ambient Intell. Human. Comput. 1–15 (2018)
6. Burney, S.M.A., Arifeen, Q.U., Mahmood, N., Bari, S.A.K.: Suspicious call detection using Bayesian network approach. WSEAS Trans. Inf. Sci. Appl. 37–49
7. Carnaz, G., Nogueira, V.B., Antunes, M.: A graph database representation of Portuguese criminal-related documents. In: Informatics, vol. 8, p. 37. MDPI (2021)
8. Das, P., Das, A.K., Nayak, J., Pelusi, D., Ding, W.: A graph based clustering approach for relation extraction from crime data. IEEE Access **7**, 101269–101282 (2019)
9. Karpova, A., Savelev, A., Vilnin, A., Kuznetsov, S.: Method for detecting far-right extremist communities on social media. Soc. Sci. **11**(5), 200 (2022)

10. Lara-Cabrera, R., Gonzalez-Pardo, A., Barhamgi, M., Camacho, D.: Extracting radicalisation behavioural patterns from social network data. In: 2017 28th International Workshop on Database and Expert Systems Applications (DEXA), pp. 6–10. IEEE (2017)
11. Mushtaq, N., Ali, K., Moetesum, M., Siddiqi, I.: Impact of demographics on automated criminal tendency detection from facial images. In: 2022 International Conference on Frontiers of Information Technology (FIT), pp. 88–93. IEEE (2022)
12. Schiliro, F., Beheshti, A., Moustafa, N.: A novel cognitive computing technique using convolutional networks for automating the criminal investigation process in policing. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) IntelliSys 2020. AISC, vol. 1250, pp. 528–539. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-55180-3_39
13. Zaman, M., et al.: Execution of coordinate based classifier system to predict specific criminal behavior using regional multi person pose estimator. Ph.D. thesis, Brac University (2021)