Monshizadeh, Mehrnoosh; Khatri, Vikramajeet; Varfan, Mohammadali; Kantola, Raimo

LiaaS: Lawful Interception as a Service

Published in:
2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)

DOI:
10.23919/SOFTCOM.2018.8555753

Published: 01/01/2018

Document Version
Peer reviewed version

Please cite the original version:
Monshizadeh, M., Khatri, V., Varfan, M., & Kantola, R. (2018). LiaaS: Lawful Interception as a Service. In *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)* (pp. 268-273). (International Conference on Software, Telecomm unications and Computer Networks). IEEE. https://doi.org/10.23919/SOFTCOM.2018.8555753

LiaaS: Lawful Interception as a Service

Mehrnoosh Monshizadeh*†, Vikramajeet Khatri*, Mohammadali Varfan[Ψ], Raimo Kantola†

*Nokia Bell Labs, Finland

† Department of Comnet, Aalto University, Espoo, Finland

[Ψ]Bonn-Rhein-Sieg University of Applied Science, Sankt Augustin, Germany

mehrnoosh.monshizadeh@nokia-bell-labs.com*, mehrnoosh.monshizadeh@aalto.fi†, vikramajeet.khatri@nokia-bell-labs.com*, ali.varfan@smail.inf.h-brs.de[Ψ], raimo.kantola@aalto.fi†

Abstract – **Machine learning techniques are the key to success for big data analytics in forthcoming 5G and cloud networks. Internet Service Providers (ISPs) and mobile networks are still relying on traditional Lawful Interception (LI) mechanisms that use error prone meta data and are vulnerable to cyber-attacks. While new identity methods are used to monitor suspected end users, the major challenge is the amount of data that needs to be monitored to find the traffic of interest related to the specific targets. On the other hand, for a conversation (audio or video) between two or multiple attendees, such as a conference call or interview, extracting, briefing and classifying important information can be prone to errors and exhaustion of resources if it is done by humans. This paper proposes an intelligent, secure, fast and reliable platform called Lawful interception as a Service (LiaaS) to detect, analyze and intercept content from different media such as voice and video call. The proposed platform also extracts the minutes of conversation and the most important information from the media (audio or video) so any desired detail can be searched from it.**

Keywords – Lawful Interception; Machine Learning; Automated Minutes; Automated Audio Analysis; Automated Video Analysis.

I. INTRODUCTION

The communication providers have an obligation to ensure and maintain Lawful Interception (LI) capabilities. They must be able to intercept all applicable communications of a certain target without any gaps in coverage, and they must provide a secure and reliable network to transmit the intercepted information to the police and other Law Enforcement Agencies (LEAs). We should also distinguish between different applications of LI even though they share a great deal in terms of the challenges. We can consider LI as applied to criminal investigations and then LI as applied to national security. The requirements and specifications will differ, and some may require more advanced techniques such as "keyword detection". In fact, regarding to national security, the technology is very flexible and has also been adapted to support cyber security requirements that protect many types of critical networks from attack, such as essential computerized monitoring and management services for electrical power, water supply, etc. [1-2].

LEAs must consider a fast and reliable mechanism for detecting, analyzing and intercepting content from the different media such as voice and video calls. While new identity methods are used to monitor suspected end users, the major challenge is the amount of data that needs to be monitored in order to find the traffic of interest applicable to the specific targets. Machine learning techniques are the key to success for analysis of big data in forthcoming 5G and cloud networks.

However, mobile networks and Internet Service Providers (ISPs) are still relying on traditional LI mechanisms that use error prone meta data and are vulnerable to cyber-attacks. An example is Greek wiretapping case of 2004-2005, also known as Greek Watergate. In the mentioned case, the LI system in one of the big mobile operators had been penetrated and more than 100 mobile phones of high ranking civil servants had been illegally tapped.

On the other hand, for a conversation (audio or video) between two or multiple attendees, such as a conference call or interview, extracting, briefing and classifying important information can be prone to error and resource exhaustion if it is done by humans. Therefore, this paper proposes an intelligent and secure lawful interception platform based on automated minutes from the call that may reside in cloud and be offered as a service. The proposed platform also extracts the minutes of conversation and the most important information from the media (audio or video) so any desired detail can be searched from it.

The rest of the paper is organized as follows. Section 2 briefly reviews related work. Section 3 discusses the literature review and research motivation. Section 4 discusses our proposed architecture for Lawful interception as a Service (LiaaS) that analyzes audio and video content and delivers alerts and meeting minutes. In Section 5, planned performance evaluation is briefly discussed. Finally, conclusion is presented in the last section.

II. RELATED WORK

Berna et al. [3] propose a meeting recorder that captures multimodal information of a meeting. Subsequent analysis of the information produces scores indicative of visually and aurally significant events that can help identify segments of the meeting recording. Textual analysis can enhance searching for

meeting segments and otherwise enhance the presentation of the meeting segments. This study does not do any important information extraction neither it is applied for voice or image recognition.

Zili et al. [4] propose a conference recorder with a speaker speech extracting function, which would provide meeting transcript (for each attendee attending the conference). The proposed system consists of main control module, a recording and playback module, a removable storage module, an interaction and display module and speech processing module. The speech processing module consists of speaker segmenting module and speaker clustering module. The speaker segmenting module detects speaker changing points in speech stream and divides speech stream into multiple speech sections in accordance with changing points. The speaker clustering module performs spectral clustering and joins sections of same speaker together in a sequence and determines number of speakers and speech for each speaker, which is delivered as output of system (transcript for each attendee in a conference). Their proposed conference recorder extracts a meeting transcript only based on each attendee.

In a study by Itay and Itai [5], the information elements such as IMSI, IMEI, IP address, location coordinates, name, address and any other useful information, are extracted from captured data. The extracted information elements are tagged with a semantic role and stored into database. The information exchanged between a source and a destination in a network is captured by a tap. In case, if the extracted information cannot be tagged with a known semantic rule, the learning process starts with training data, learns and formulates a rule with certainty. The study mentions using supervised, unsupervised and semi-supervised machine learning algorithms but its implementation details are not mentioned. The study mentions the use case for lawful interception, but only extracts useful information from data packets (rather than voice and video information). In addition, the study doesn't offer any specific mechanism (such as big data techniques) for voice or video calls.

Evan et al. [6] propose a distributed audio streaming and speech recognition method that is implemented on client and server sides. The processes on client and server side communicate over the network. The client process monitors audio input and detects whether it contains speech or not using voice activity detection techniques. If it contains speech, then it looks for "wakeword". Upon detection of "wakeword", the client process signals the server side and starts streaming audio over the network. The server process upon receiving audio, processes it and transcribes the audio if needed. While processing audio, if the server determines disconnect criterion (no speech, noise, low energy, end of interaction indication, silence etc.), the server signals the client to stop streaming audio and therefore audio streaming is stopped.

In a study by Shirin et al. [7], a speech interactive command system is proposed with two modes of operation: a stationary base device and a handheld remote device. The speech is originated from one or several devices. The user either speaks a keyword or uses Push To Talk (PTT) button to interact with device. The speech service then performs automatic speech recognition and Natural Language Understanding (NLU) to understand user intent. A follow-up dialog is conducted with user in multiple turns. As an example, a user delivers speech "Set a timer", and the system asks, "For how long?" to understand it completely.

III. Literature Review and Research Motivation

While LEAs are still applying traditional mechanisms for LI, earlier studies in the field of automated meeting minutes and analysis of audio and video content, mainly focus on the speech detection and conversion of speech to text and search specific keywords in the speech. Therefore, in this study we apply data analytics and machine learning techniques to deliver an automated, efficient and reliable architecture for lawful interception called LiaaS.

As shown in Fig. 1, the proposed LiaaS, compromises of four main modules: Anomaly Detection, Automated Information Extraction, Data Mining (DM) Based Content of Communication (CC) Analyzer and Minutes Taking. The anomaly detection module is used to secure LiaaS from cyber-attacks such as Denial of Service (DoS). The Automated Information Extraction module extracts text and important information from call. The DM based CC Analyzer module classifies received information, labels and provides results to LEA.

In the first step, LiaaS checks the media for its data type (audio or video call). If it is audio call, the data will be forwarded to a scenario that is responsible for audio analysis. If it is video call, the audio and image will be extracted from video call and will be forwarded to relevant scenarios. In the case of image, it will be forwarded to another scenario that is responsible for image analysis.

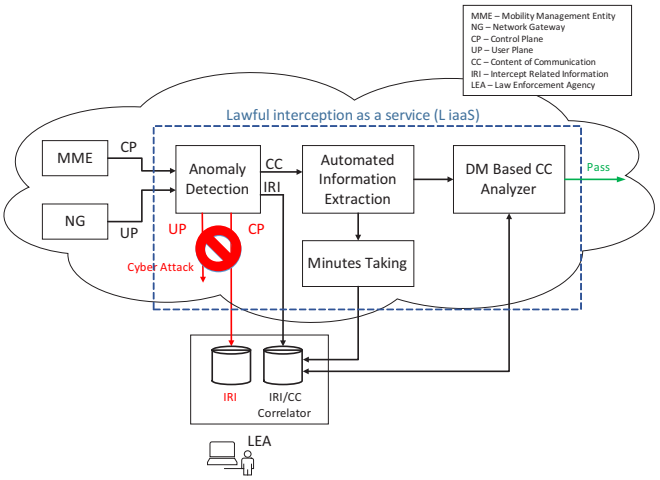

Fig. 1. A secure and reliable platform for lawful interception

In the next step, LiaaS uses pre-trained DM algorithms for classification, link analysis and clustering of extracted information. This step labels the suspicious parties and objects and forwards the results to LEA for further investigations. Intercept Related Information (IRI)/CC correlator will forward the location information and other useful information (mentioned in 3GPP TS 33.108) for further analysis in case of mobile call to DM based CC Analyzer, as shown in Fig. 1. In

the proposed platform, the minutes taking module is considered to provide a brief of the most important content of call. The database contains a copy of transcript, image and other most important content which, can be helpful to LEA for further investigations.

IV. ARCHITECTURE

The LiaaS platform as illustrated in Fig. 1, compromises of four modules:

- Anomaly Detection
- Automated Information Extraction
- Data Mining (DM) Based Content of Communication (CC) Analyzer
- Minutes Taking

A. Anomaly Detection

The anomaly detection module has been illustrated in Fig. 2.

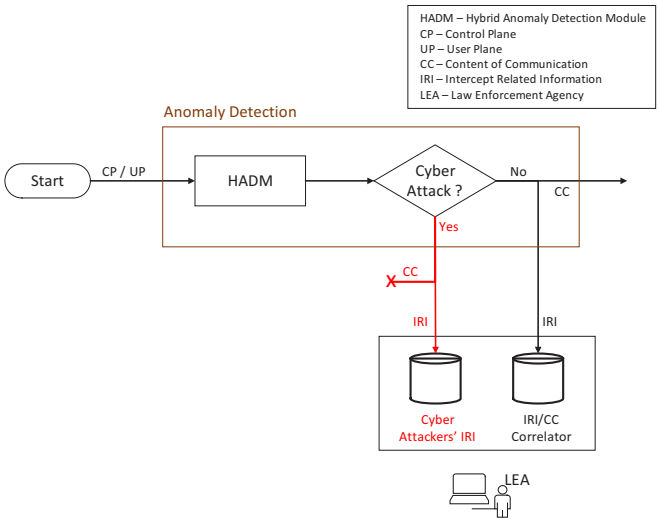

Fig. 2. Anomaly Detection module for proposed LiaaS

For security purpose of LiaaS, Control Plane (CP) and User Plane (UP) traffic will be analyzed first in Hybrid Anomaly Detection Module (HADM). A decision module will check whether the input traffic is safe or not. If the input traffic is detected as attack, the UP, known as CC for LI, will be dropped; but the CP, known as Intercept Related Information (IRI), will be sent to LEA for future investigations since the LEA may be interested to identify the attacker intensions. If the input traffic is detected as safe, the CC will be forwarded further; whereas IRI is sent to LEA's IRI/CC correlator for future analysis. The IRI/CC correlator at LEA will correlate the information in CC with related IRI in order to identify to whom CC belongs to. The CC correlator identifier will be preserved during the process.

It should be considered that not all data will be forwarded to LEA from core network that may reside in the cloud. Only the data of the LEA's specific targets, which is filtered by using some identifier such as IMEI, IMSI, IP address etc. will be forwarded to HADM. The filtered data will be then further analyzed by HADM as mentioned above. Our analysis is per call, so there can be several parties in a call (similar to conference call) and IMSI, MSISDN etc. for all parties in the call should be forwarded to LiaaS.

B. Automated Information Extraction

The Automated Information Extraction module extracts text and important information from the call and has been illustrated in Fig. 3.

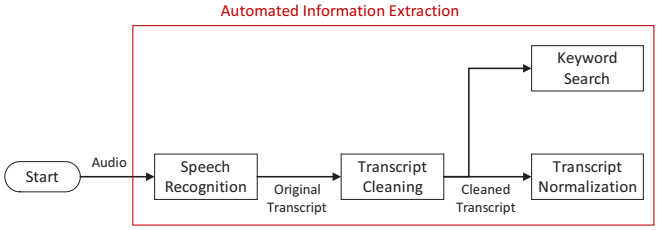

Fig. 3. Automated Information Extraction for LiaaS

This module takes the audio speech as input and performs speech recognition on audio which converts the audio into text and produces the original transcript of conversation. The speech recognition uses pre-trained machine learning algorithms such as WaveNet algorithm, which is based on convolution neural network and specially designed for speech recognition. The speech recognition is trained with different datasets such as TIMIT, free VTCK, LibriSpeech, TED-LIUM, etc.

The original transcript of conversation is passed to transcript cleaning process, which produces the cleaned transcript of conversation. The transcript cleaning uses automatic methods for finding and solving errors and unnecessary words and characters in the produced original transcript. Some of the errors include apostrophe lookup, removal of stop words, removal of expressions such as laughing, slang lookup, standardizing words such as "Hellllloooo" to "Hello", grammar checking and spelling correction etc.

A copy of cleaned transcript is passed to keyword search process, which offers LEAs to search for keywords in it and displays search results. The transcript is searched for keywords using the sentence tokenization methods.

A copy of cleaned transcript is also passed to transcript normalization, so it can be analyzed by the Natural Language Processing (NLP) algorithms. Text normalization includes various methods such as case conversion (changing all characters to upper or lower case), stemming (creating the root stem of words), lemmatization (creating the root word of words), tokenization (separating the text to words or sentences), etc.

C. Data Mining (DM) Based Content of Communication (CC) Analyzer

The DM based CC Analyzer module classifies received information, labels and provides results to LEA. The module can be seen in Fig. 4.

This module consists of pre-trained classification, link analysis and clustering algorithms. The classification algorithm

includes classes of either important keywords pre-defined by LEA. A decision module checks whether the received information matches with any of those classes. If it matches, a copy of information is sent to LEA for further investigations. In addition, the information is analyzed in link analysis algorithm in order to find the link with suspected destination, parties, locations etc. The results of link analysis are forwarded to an un-supervised clustering algorithm in order to cluster them. The cluster information will be forwarded to LEA for further investigations.

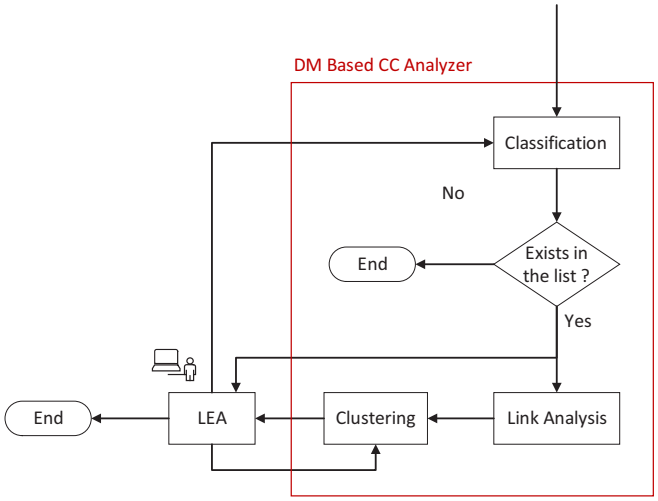

Fig. 4. DM Based CC Analyzer for LiaaS

In the case of analyzing images, image recognition is responsible for a wide range of processes such as data extraction, location detection, path detection, behavior recognition, object recognition and so on. All the extracted information will be used later by classification algorithm. For example, object recognition defines suspicious objects such as specific flag, firearm or any other dangerous equipment. The classification algorithm includes pre-defined classes by LEA. A decision module checks whether the received image matches with any of those classes or not. If it matches, a copy of image is sent to LEA for further investigations. In addition, the information is analyzed in link analysis algorithm in order to find the link with suspected destination parties. The results of link analysis are forwarded to an unsupervised clustering algorithm in order to cluster them. The cluster information will be forwarded to LEA for further investigations.

A feedback from LEA is sent to clustering algorithm in order to remove unwanted clusters or rectify clusters. Another feedback from LEA is sent to classification algorithm in order to add new classes or rectify the existing classes based on the received information and final investigations made by LEA.

D. *Minutes Taking*

Minutes taking module consists of one or more pre-trained machine learning algorithms for text summarization of the normalized transcript of the conversation. There are two methods for text summarization as illustrated in Fig. 5: extraction based and abstraction based. In extraction based approach, the important sentences from the normalized transcript are taken and the final summary is generated using those sentences. In abstraction based approach, instead of copying the important sentences from the normalized transcript, the important concepts are found in the normalized transcript and then new sentences are generated for creating a final summary. As show in Fig. 6, the minutes can be separated for different parties in the conversation.

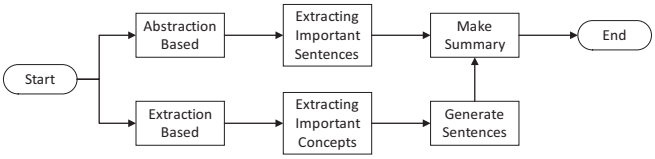

Fig. 5. Minutes taking methods

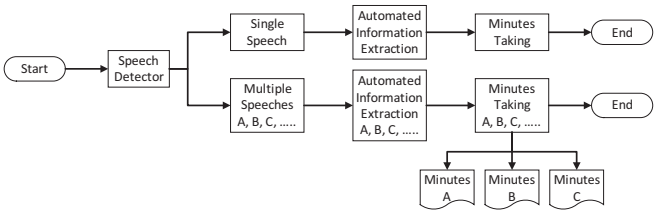

Fig. 6. Minutes taking for single and multiple parties

The minutes taking module may use more than one machine learning algorithm, cover all the categories and have an extractive and abstractive summary of the conversation. The applied machine learning algorithms include recurrent Long Short-Term Memory (LSTM) neural network and phrase selection and merging. These algorithms have shown promising results in text summarization and the latter is designed mainly for abstractive based summarization. These algorithms can find the important information and concepts of a text (conversation) and they can also be implemented to summarize multiple texts. For training these algorithms, different datasets such as Text Analysis Conference (TAC), Gigaword, Reuters Corpuses, Washington Post Corpus etc. The minutes taking module generates minutes of the normalized transcript.

E. *Overall Functionality*

The overall functionality of LiaaS has been illustrated in Fig. 7. A decision module checks whether the call type is audio or video. In the case of an audio call, approach 1 is applied; otherwise approach 2 is applied.

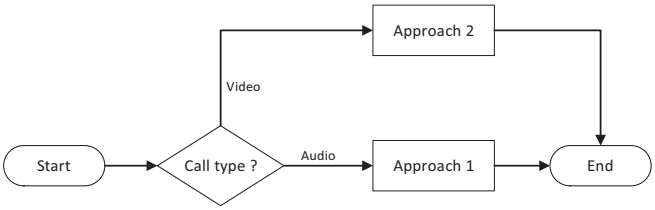

Fig. 7. LiaaS overall functionality based on call type

1) *Approach 1*

Approach 1 for processing audio call can be seen in Fig. 8. The input audio will be divided into main audio of conversation and background audio. The main audio is forwarded to speech processing module for further analysis. The background audio will be divided into speech and non-speech audio.

The speech audio is forwarded to speech processing module for further analysis. From non-speech audio, important information is extracted. The extracted information is stored into database also sent to Link Analysis module to extract other useful information.

Link Analysis module analyzes the received information to find the link with suspected destination, parties, locations etc. The results of link analysis are forwarded to Clustering module. Clustering module clusters the received results using an unsupervised clustering algorithm. The clustered information is forwarded to LEA for further investigations.

LEA analyses the received results and sends feedback to Clustering module to remove unwanted clusters or rectify clusters. LEA also queries the database to acquire a copy of extracted information if needed in investigations.

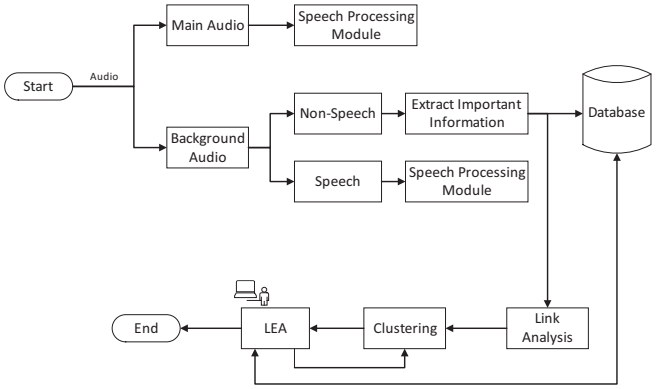

Fig. 8. Approach 1 for processing audio call

2) *Speech Processing Module*

The module has been illustrated in Fig. 9 and processes the input audio, which is fed to Speech Recognition module. Speech Recognition module converts audio into text and produces the original transcript of conversation using pre-trained machine learning algorithms. The produced original transcript is forwarded to Transcript Cleaning module and a copy is saved into database.

Transcript Cleaning module produces cleaned transcript of conversation. The cleaned transcript is forwarded to Transcript Normalization module and Keyword Search module. In addition, a copy is saved into database. Keyword Search module searches and displays the keywords that are pre-defined by LEA. The search results are saved into database.

Transcript Normalization module analyses cleaned transcript using NLP algorithms and produces normalized transcript. The produced normalized transcript is forwarded to Minutes Taking module and Classification module. In addition, a copy is saved into database. Minutes Taking module produces summary / minutes of the normalized transcript of the conversation using extraction and abstraction based mechanisms. The extracted minutes of the conversation are stored into database and forwarded to LEA.

Classification module classifies the received information from transcript normalization module. The classes are pre-defined by LEA. A decision module checks whether the received information matches with any of those classes. In case it matches, it is forwarded to Link Analysis module and LEA. Link Analysis module analyzes the received information to find the link with suspected destination, parties, locations etc. The results of link analysis are forwarded to Clustering module.

Clustering module clusters the received results using an unsupervised clustering algorithm. The clustered information is forwarded to LEA for further investigations. LEA analyses the received results and sends feedback to Clustering and Classification modules to make changes based on investigations. LEA also queries the database to acquire a copy of extracted information if needed in investigations.

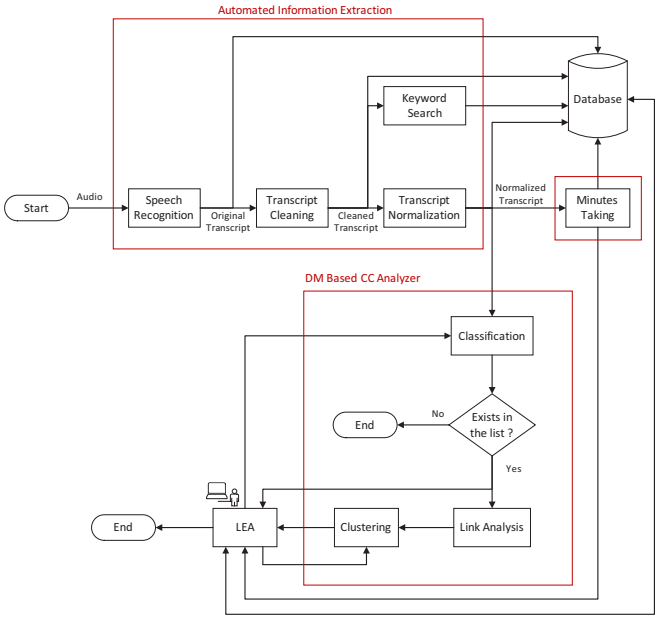

Fig. 9. Speech processing module

3) *Approach 2*

Approach 2 for processing video can be seen in Fig. 10. Initially, Extract Audio module extracts the audio from the input video, which is forwarded to Approach 1 for further analysis. Extract Image module extracts images from the input video, which are forwarded to Image Recognition module and stored into database.

Image recognition is responsible for a wide range of processes such as data extraction, location detection, path detection, behavior recognition, object recognition and so on. All the extracted information will be used later by classification algorithm. Image Recognition module forwards the information to Classification module.

Classification module classifies the received information into classes pre-defined by LEA. A decision module checks whether the received information matches with any of those classes. In case it does, it is forwarded to LEA and to Link Analysis module. Link Analysis module analyzes the received information to find the link with suspected destination parties. The results are forwarded to Clustering module.

Clustering module clusters the received results using an unsupervised clustering algorithm. The clustered information is

forwarded to LEA for further investigations. LEA analyses the received results and sends feedback to Clustering and Classification modules to make changes based on investigations. LEA also queries the database to acquire a copy of extracted information if needed in investigations.

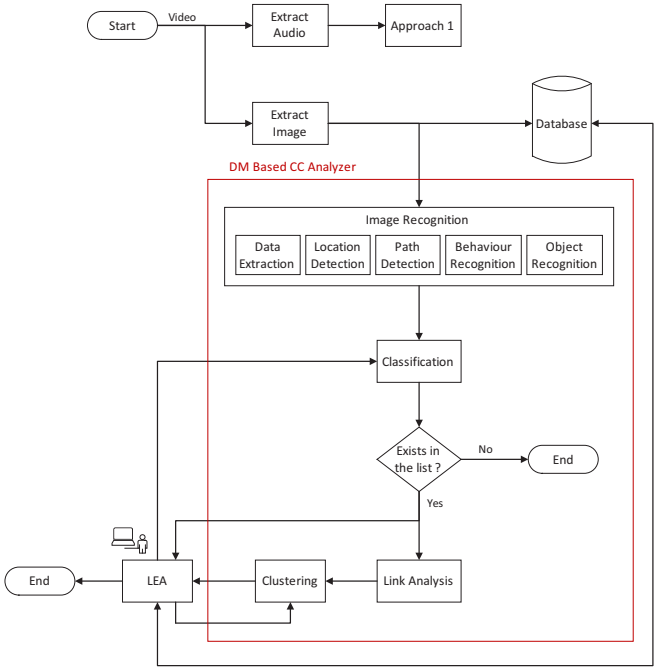

Fig. 10. Approach 2 for processing video call

V. Performance Evaluation

In order to evaluate LiaaS performance, applied algorithms in different modules must be tested against different datasets and efficiency must be measured based on the metrics such as accuracy, precision, recall and computation time.

The implementation of the proposed platform is not presented in this paper due to space limitations. However, in a separate study, we applied similar DM techniques to evaluate the HADM module efficiency, robustness and scalability. In the mentioned paper, network traffic flow control in combination with DM techniques are proposed. Different feature selection methods including F-Score, Chi2, Recursive Feature Elimination (RFE) and SVMonline were applied to find the best features. The feature selection methods had been selected based on the algorithms computation time and detection rates. Different classification and clustering algorithms including Extreme Learning Machine (ELM), MultiLayer Perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Decision Tree (DT) and Logistic Regression (LR) were applied. In order to evaluate the HADM module robustness and scalability, different datasets had been used. The best algorithms were then selected through a benchmark on applied datasets and based on the metrics such as accuracy, precision, recall and computation time. The result of proposed approach showed tremendous growth in accuracy and reduced computation time for applied algorithms and feature selection methods in the mentioned study. All the experiments were carried out on a workstation with Intel core i5 Quad @2.6GHz, 16 GB RAM and 500GB HDD. The scripts were developed in Python in a Linux environment [8].

The similar test environment or a server with higher computation capacity can be utilized to evaluate LiaaS performance. However, the whole functionality of LiaaS can be deployed on several Virtual Machines (VMs) for load balancing and decentralized monitoring purposes in order to handle large amount of traffic at core network in the cloud.

VI. Conclusion

Law Enforcement Agencies (LEAs) must consider a fast and reliable mechanism for detecting, analyzing and intercepting content from the different media such as voice and video calls. The current Lawful Interception (LI) architecture is used to monitor suspected end users with traditional methods rather than using machine learning. The major challenge is the amount of data that needs to be monitored to find the traffic of interest related to specific targets. While data analytic and machine learning techniques are the keys to success in big data analysis in forthcoming 5G and cloud networks, mobile networks and Internet Service Providers (ISPs) are still relying on traditional LI mechanisms that use error prone meta data. In this study we applied data analytic and machine learning techniques to deliver an automated, efficient and reliable architecture for lawful interception called Lawful interception as a Service (LiaaS). This platform analyzes audio and video content and delivers the alerts and meeting minute and can be integrated to 5G and other technologies like Internet of Things (IoT) and cloud as intelligent LI functionality.

References

[1] K. Wale, "Modern Day Challenges for Lawful Intercept", radisys, http://go.radisys.com/rs/radisys/images/paper-dpi-modern-day.pdf

[2] Lawful Interception in the Digital Age: Vital Elements of an Effective Solution, Ultimaco LIMS, https://lims.utimaco.com/products/lawful-interception-management-system/

[3] E. Berna, G. Jamey, H. J. Jonathan, and L. Dar-Shyange, "Multimodal access of meeting recordings", US Patent 7606444 (B1), 20 Oct. 2009.

[4] W. Zili, L. Yanxiong, and L. Guanglong, "Conference recorder with speech extracting function and speech extracting method", CN Patent 103530432 (A), 22 Jan. 2014.

[5] M. Itay, and Z. Itai, "System and method for learning semantic roles of information elements", US Patent 2017293595 (A1), 12 Oct. 2017.

[6] S. H. Evan, B. K. John, S. Nikko, and T. R. Pauls, "Distributed endpointing for speech recognition", US Patent 9818407 (B1), 14 Nov. 2017.

[7] S. Shirin, P. A. Therese, T. Marcello. S. Shamitha, and P. K. Wesleys, "Multiple-source speech dialog input", US Patent 9792901 (B1), 17 Oct. 2017.

[8] M. Monshizadeh, V. Khatri, A. Buse and R. Kantola, "An Intelligent Defense and Filtration Platform for Network Traffic," *IFIP 16th International Conference on Wired/Wireless Internet Communications (WWIC)*, June 18-20, 2018, Boston, USA.