Paper

Introduction to Big Data Management Based on Agent Oriented Cyber Security

Jamal Raiyn

Computer Science Department, Al Qasmi Academic College, Baqa Al Gharbiah, Israel

Abstract—This paper deals with information security and safety issues in public open spaces. Public open spaces include high streets, street markets, shopping centers, community gardens, parks, and playgrounds, each of which plays a vital role in the social, cultural and economic life of a community. Those outdoor public places are mashed up with various ICT tools, such as video surveillance, smartphone apps, Internet of Things (IoT) technologies, and biometric big data (called Cyber Parks). Security and safety in public places may include video surveillance of movement and the securing of personalized information and location-based services. The article introduces technologies used in Cyber Parks to achieve information security in big data era.

Keywords—big data, cyber security, information security.

1. Introduction

The data volume used in Internet technologies is rising rapidly. This huge amount is known as big data [1] and is characterized by three aspects according to Madden [2]:

- the data are numerous,
- the data cannot be categorized into regular relational databases,
- the data are generated, captured, and processed very quickly.

Big data has generated significant interest in various fields, including the manufacturing of healthcare machines, banking transactions, social media, and satellite imaging. Big data challenges have been described by Michael and Miller [3], such as rapid data growth, transfer speeds, the diversity of data, and security issues. Big data is still in its infancy stage and has not been reviewed in general. Hence, this study comprehensively surveys and classifies its various attributes, i.e. volume, management, analysis, security, nature, definitions, and rapid growth rate. The development of new IT technologies has rapidly increased the volume of information, which cannot be processed using existing technologies and methods [4]–[6]. In computational sciences, big data presents critical problems that require serious attention [7]. In the IT industry as a whole, the rapid rise of big data has generated new challenges with respect to data management and analysis. According to Khan *et al*. [8], five common issues involve: volume, variety, velocity, value, and complexity. Madden [2] note additional issues such as the fast growth of volume, variety, value, management, security, and efficiency. In some fields, data have grown rapidly. However, the type of data that increases most rapidly is unstructured data. This type is characterized by "human information" such as high-definition videos, movies, photos, scientific simulations, financial transactions, phone records, genomic datasets, seismic images, geospatial maps, e-mails, tweets, website data, call-center conversations, mobile phone calls, documents, sensor data, telemetry information, medical records and images, climatology and weather records, log files, and text. According to Khan *et al*. [8], unstructured information may account for more than 70% to 80% of all data in organizations. Currently, 84% of IT managers process unstructured data, and this percentage is expected to drop by 44% in the near future [9]. Most unstructured data are not modeled, are random, and are difficult to analyze.

Big data technology aims to minimize hardware and processing costs and to verify the value of information before committing significant company resources. Properly managed big data are accessible, reliable, secure, and manageable. Hence, such applications can be applied in various complex scientific disciplines (either single or interdisciplinary), including atmospheric science, astronomy, medicine, biology, genomics, and biogeochemistry. Khan *et al*. [8] have proposed a new data life cycle that uses the technologies and terminologies of big data. This new approach to data management and handling required in e-science is reflected in the scientific data life cycle management (SDLM) model. With this model, existing practices are analyzed in different scientific communities. The generic life cycle of scientific data is composed of sequential stages, including experiment planning (for research projects), data collection and processing, discussion, feedback, and archiving. The proposed data life cycle consists of the following stages: collection, filtering and classification, data analysis, storing, sharing and publishing, data retrieval and discovery.

In processing big data, users face several challenges [10]. Applications requires a huge storage capacity, rapidly search engines, sharing and analysis capabilities, and in some areas data visualization. These and others challenges

need to overcome to maximize big data. Currently, various techniques and technologies are used, such as SAS, R, machine learning platforms and Matlab to handle extensive data analysis. However, the proposed schemes are limited in managing big data effectively and are still lacking. According to Khan *et al*. [8], others challenges to big data analysis include data inconsistency and incompleteness, scalability, timeliness, and security.
This paper introduces a new scheme for big data management based on agent oriented cyber security in public spaces.

2. IoT Big Data Generation

In the IoT various area devices with enormous of sensors networks are used in different fields, such as, security and privacy, social network, transportation, medical care, industry, traffic, and public department. IoT devices are grown up quickly and collect the most important part of big data. IoT is considered an important source of big data.

2.1. Security and Privacy Indoor

Video surveillance system is the most important issue in homeland security field because of its ability to track and to detect a particular person. To overcome the lack of the conventional video surveillance system that is based on human perception this paper introduces a novel cognitive video surveillance system (CVS) that is based on mobile agents. CVS offers important attributes such as suspect objects detection, smart camera cooperation for person tracking. According to many studies, an agent-based approach is appropriate for distributed systems, since mobile agents can transfer copies of themselves to other servers in the system.
Various numbers of papers in the literature have been proposed and focused on computer vision problems in the context of multi-camera surveillance systems. The main problems highlighted in these papers are object detection and tracking and site-wide, multi-target, multi-camera tracking. The importance of accurate detection and tracking is obvious, since the extracted tracking, information can be directly used for site activity/event detection. Furthermore, tracking data is needed as a first step toward controlling a set of security cameras to acquire high-quality images, and toward, for example, building biometric signatures of the tracked targets automatically. The security camera is controlled to track and capture one target at a time, with the next target chosen as the nearest one to the current target. These heuristics-based algorithms provide a simple way of computing. Here the scenario is considered that the smart camera captures two similar objects (e.g. twins), then each object selects different path. The tracking process will become confused. Furthermore, the smart camera is limited to cover certain zone in public place (indoor).
The suggested solutions to improve the conventional video surveillance system are extended in various ways. A part of the approaches was to use an active camera to track a person automatically, thus the security camera moves in a synchronized motion along with the projected movement of the targeted person. These approaches are capable of locating and tracking small number of people. Another common approach was to position the camera at strategic surveillance locations. This is not possible in some situations due to the number of cameras that would be necessary for full coverage, and in such cases, this approach is not feasible due to limited resources. A third approach is to identify and track numerous targeted people at the same time involves image processing and installation of video cameras at any designated location. Such image processing increases server load.
The limitation of human perception system in conventional video surveillance system increases the demand to develop cognitive surveillance application. Many of the proposed video surveillance system are expensive and lack the capability of cognitive monitoring system (such as no image analysis) and ability to send warning signal autonomous in real-time and before the incidents happen. Furthermore, it is difficult and might take a long time for the human to locate the suspects in the video after the incidents did happen. The problem may get more completely in the larger scale surveillance system.
The next generation video surveillance systems expected not only to solve the issues of detection and tracking but also to solve the issue of human body analysis. In the literature, it can be found many references in development. In such area, the CVS aims to offer meaningful characteristics like automatic, autonomy, real-time surveillance such as face recognition, suspects object, target detection, and tracking using cooperative smart cameras. Many face recognition systems have a video sequence as the input. Those systems may require being capable of not only detecting but also tracking faces. Face tracking is essentially a motion estimation problem. Face tracking can be performed using many different methods, e.g., head tracking, feature tracking, image-based tracking, model-based tracking. These are different ways to classify these algorithms.

2.2. Model of CVS System

In this section we introduce the system model of the video surveillance system. Video surveillance system has been used for monitoring, real-time image capturing, processing, and surveillance information analyzing. The infrastructure of the system model is divided in three main layers: mobile agents that are used to track suspect objects, cognitive video surveillance management (CVS), and protocol for communication as shown in Fig. 1. Each end device, smart camera, covers a certain zone or cell. Smart camera used for collecting parameters of human face.
In the system model has been introduced two communication protocols. The first protocol is used for agent-to-agent communication protocol. The protocol is based on messages exchange as shown in Fig. 1. The goal is to update the agents. The second protocol is used for communication between CVS and mobile agent.

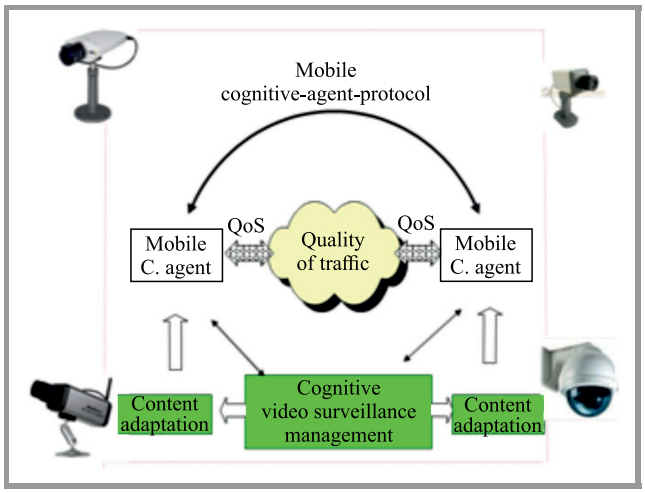

Fig. 1. CVS system model concept.

Mobile agents are placed in smart camera stations and aims to track the suspect object from smart camera station to others. Mobile agent offers various characteristics, e.g. negotiation, making decision, roaming, and cloning.
CVS provide the mobile agent with information. Based on received information mobile agents make decision when and where to move to next smart camera station.
In order to track moving objects, two strategies are used. The first is based on messaging protocol (msg_protocol) informing the mobile agent about the position of the suspect objects. The second strategy uses the protocol to help the mobile agent to roaming from point to others.

2.3. CVS Methodology

CVS uses a database of images. Pixels are described by a set of binary sequences. Each sequence presents certain properties (color). The database is divided into two separate sets of pixels – the training set and the test set. In both there are pixels, which belong to a certain family of colors (attributes) and sequence TP, which do not belong TN:

$$TP = X = \{X_1, X_2, \ldots, X_n\},$$

$$TN = Y = \{Y_1, Y_2, \ldots, Y_n\}.$$

Each image is then divided into frames $X_1 \ldots X_n$, a frame being a subset of n pixel from the sequence. The number of pixel in each frame is a variable and is dynamically set to obtain optimal results:

$$\begin{aligned} X_1 &= \{x_1^1, x_2^1, \ldots, x_n^1\}, \\ X_2 &= \{x_1^2, x_2^2, \ldots, x_n^2\}, \\ &\vdots \\ X_n^m &= \{x_1^m, x_2^m, \ldots, x_n^m\}. \end{aligned}$$

For example if a certain frame is comprised of 200 segments, the frames might consist of pixels 1 to 10, 2 to 11, 3 to 12, etc. Statistical methods are then applied to find correlation between a certain properties of the frame.

The basic logic of statistical differentiation of pixel is known and widely used in many prediction systems.

$$J = X \oplus Y,$$

$$J = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{otherwise}. \end{cases}$$

A large number of correlating factors is defined by CVS and grouped in sets. A number is linked with each correlating factor. Each factor is then turned into a single number which represents the strength of the correlation factors for each frame with respect to the probability that this frame belongs to the certain family or not. As a result there are large number of frames, for each pair of a frame we have a number which is correlated to the probability that this frame belongs to a certain attribute (color similarity) or does not belong:

$$J = \{J_1^1, J_2^2, J_3^3, J_4^4 \ \ldots\}$$

or after optimization of J:

$$J_{Prediction}(J_1^*) - J_{demand} + k \cdot (\Delta J).$$

In addition to the statistical method logical XOR multiplication of matrices is applied to enrich the number of frames, which are potentially contributing to the prediction model. CVS can be implemented in a dynamic environment. When the training databases are modified, the prediction mechanism is modified as well with improved prediction capabilities.

2.4. Security and Privacy Outdoor

Modern cities offer various kinds of public places, and are created for different targets, i.e. public places for students and others on academic campuses, for visitors to historical sites, and for families and tourists. Public open spaces that are supported by various kinds of modern information communication technologies are called Cyber Parks [9]. Such places providing connectivity services to users on their personal computers, smart phones, tablets, and other mobile end-devices. Many users use Internet technologies for storing private data. Furthermore, Internet technologies are used for communication in business, the military, medicine, education, and government and public services. Over the last decade, as well, crime in virtual life has increased. Cyber attacks are performed through Internet networks that target individual machines, mobile devices, communications protocols, or smartphone application services. Cyber attacks are performed by spreading malware, by creating phishing websites, and by other means [6].
To implement information security policies and safety in Cyber Parks [10], security models are needed that lay out guidelines for securing information and communication. Cyber Park security models are based on formal models of access rights to smart phone applications and web services. In addition, an adaptive agent recognizes the applications that being used, and a mobile agent platform [11]–[15] creates mobile agents to serve the Cyber Park visitors. By

monitoring the behavior of users, detection systems ensure information privacy.
A mobile agent aims to fulfill user's preferences based on a dynamic environment. The mobile agent's structure is divided to three parts, as follows:

- Source code – the program consists of several classes to define the agent's behavior. In the source code, the backbone of the agent is created, which contains the basic rules. The agent then grows and develops itself according to the requirements of its environment;
- State – the agent's internal variables enable it to resume it is activities when it is found to be in one of the following states: offline (sleeping, in an evolution process), online (awake), busy, waiting (standby), or dead;
- Attributes – attributes consist of information describing the agent, its movement history, its resource requirements, and authentication keys.

In order to mediate useful tasks, a communication model to establish communication between mobile end users and the Cyber Park service provider is used. The agents in the system should be able to understand each other, and they should use the same message transport protocol. Messages are a data oriented communication mechanism, generally used to transfer data between processes. Communication is either asynchronous or synchronous.

2.5. Concept of Secured Information

Authentication refers to process of obtaining a confirmation that a person who is requesting a service, is a valid user. It is accomplished via the presentation of an identity and credentials, such as passwords, tokens, digital certificates, and phone numbers. To increase information security, users need a password to log in. The system starts the identification process and creates a mobile agent for each user, as shown in Fig. 2. The mobile agent is responsible for communication security in the system.

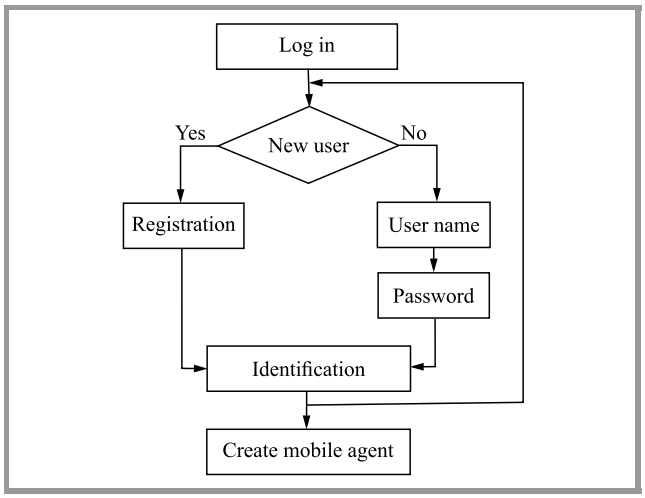

Fig. 2. Block diagram of authentication process.

Messages are a data oriented communication mechanism. Request-response mechanism is used to transfer data between a user end device and a service provider:

- inform message – includes the mobile device ID and the kind of information requested,
- re-inform message – includes information about Cyber Park resources,
- request message – includes the sender's name, a time stamp that indicates the time the request message was generated, the receiver's name and the requested resource,
- response message – includes the sender's name, a time stamp, and the requested resource.

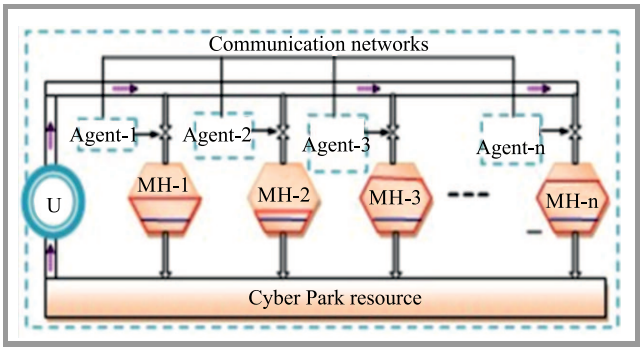

Fig. 3. Resources control.

Figure 3 illustrates the Cyber Park's resources and services. To increase information security communication between the mobile end user and service should be aware. It is necessary to checking the identity of the communication parties before establishing communication and allowing users access to information. Some users will follow a conventional scheme to access secure information. Namely, they access Cyber Park services with a password. For every user is allocated a mobile agent called a home agent. The home agent creates a new PIN number to access services. This number is a password shared between a user and a system that aims to authenticate the user.

3. Biometric Bid Data

Everyone uses several passwords to login to various systems and services. From simplicity and security viewpoint users demand new ways that replaces the passwords. Biometrics fills the user preferences and provide faster and easier. Biometrics is methods of recognizing a person based on physiological or behavioral characteristics. The recognizing process is based on measured features such as, face, fingerprint, hand geometric, iris, retinal, signature and voice. Authentication and security of biometrics data are very important issue. It refers to the collection of any kind of information about biological system, physiological or behavioral attributes. These data about humans is used to identify specific individual actions. Biometric security

is a security mechanism used to authenticate privacy data and to provide access to various IoT devices based on verification of an individual's physical characteristics. It is the strongest physical security technique used for identity verification. There are several algorithms, which convert the plain text data into cipher text data. These types of algorithms are known as encryption and decryption of data. It is used to protect data and it cannot be used for anyone except for the recipient.

3.1. Cyber Security and Privacy

It is important for Cyber Park visitors to keep their location secret. The privacy approach aims to protect private position information, as shown in Fig. 4. The mobile agent works to hide the identity of the user and his or her activity while the location for the user is visible. This prevents a cyber attacker from detecting the users location.

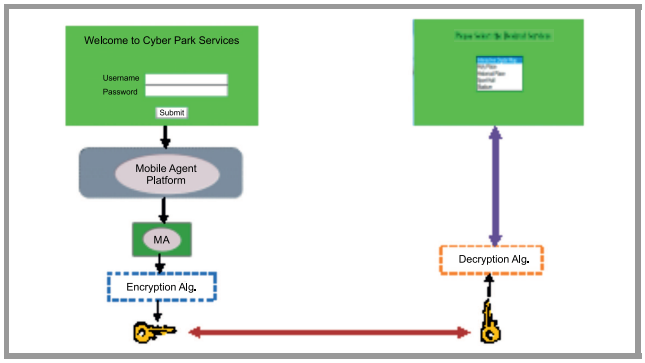

Fig. 4. Information security.

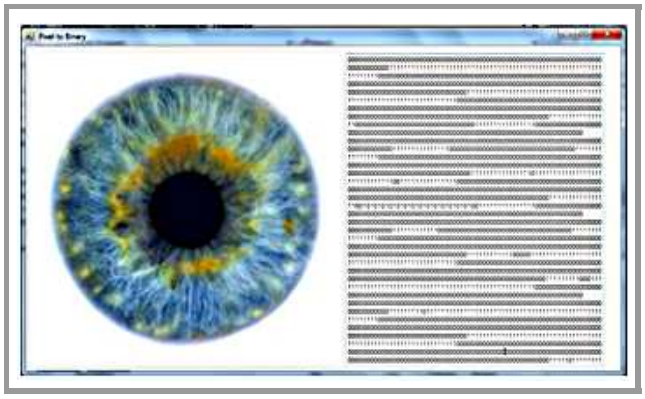

Fig. 5. Image representation in binary system.

The author has used the object oriented programming language C# to present the image in binary system as shown in Fig. 5. Hence, binary vectors are implemented in Waikato Environment for Knowledge Analysis (WEKA) platform, which implements many machine learning and data mining algorithms. As shown in Fig. 6 the image analysis in visual form is based on color classification. WEKA considers the color of the image. The colors are represented in binary system. WEKA clusters the binary vectors.

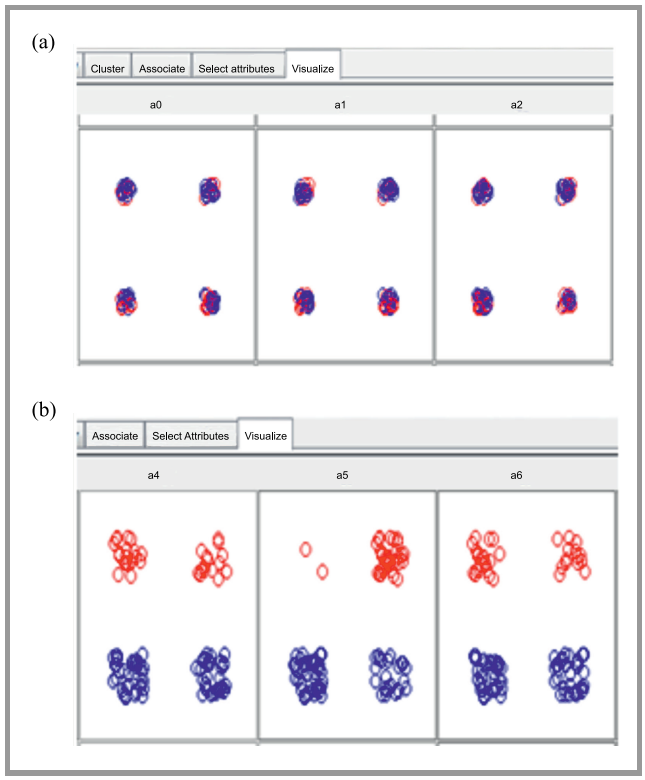

Fig. 6. WEKA platform: (a) image analyzis, (b) color classification.

4. Conclusion

The Internet and mobile technology are growing rapidly, and the data accumulated over twenty years have becomes big data. We have considered security big data indoor and outdoor, which is generated by IoT devices. The privacy approach aims to protect private position data. The mobile agent works to hide the identity of the user and his or her activity in Cyber Park (outdoor) services while the location for the user is visible. This prevents a cyber attacker from detecting the users location.

References

[1] D. Che, M. Safran, and Z. Peng, "From big data to big data mining: challenges, issues, and opportunities", in *Database Systems for Advanced Applications*, Berlin: Springer, 2013, pp. 1–15.

[2] S. Madden, "From databases to big data", *IEEE Internet Comput.*, vol. 16, no. 3, pp. 4–6, 2012 (doi: 10.1109/MIC.2012.50).

[3] K. Michael and K. W. Miller, "Big data: new opportunities and new challenges", *Computer*, vol. 46, no. 6, pp. 22–24, 2013 (doi: 10.1109/MC.2013.196).

[4] J. Raiyn, "Using cognitive radio scheme for big data traffic management in cellular systems", *Int. J. of Inform. Technol. & Manag.*, vol. 14, no. 2-3, 2015.

[5] J. Raiyn, "Toward developing real-time online course based interactive technology tools", *Adv. in Internet of Things*, vol. 4, no. 3, pp. 13–19, 2014 (doi: 10.4236/ait.2014.43003).

[6] J. Raiyn, "A Survey of cyber attack detection strategies", *Int. J. of Secur. & Its Appl.*, vol. 8, no. 1, pp. 247–256, 2014.

[7] C. A. Steed *et al.*, "Big data visual analytics for exploratory earth system simulation analysis", *Computers & Geosciences*, vol. 61, pp. 71–82, 2013 (doi: 10.1016/j.cageo.2013.07.025).

[8] N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Shiraz, and A. Gani, "Big data: Survey, technologies, opportunities, and challenges", *Scientif. World J.*, pp. 1–18, 2014 (doi: 10.1155/2014/712826).

[9] J. Raiyn, "Information security and safety in Cyberpark", *Global J. of Adv. Engin.*, vol. 2, no. 8, pp. 73–78, 2015.

[10] J. Raiyn, "Modern information and communication technology and their application in Cyberpark", *J. of Multidiscip. Sci. & Technol.*, vol. 2, no. 8, pp. 2178-2183, 2015.

[11] M. Wooldridge and N. R. Jennings, "Intelligent agents: Theory and practice", *The Knowl. Engin. Rev.*, vol. 10, no. 2, pp. 115–152, 1995.

[12] S. Russel and P. Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ: Prentice Hall, 1995.

[13] M. Luck, V. Marik, O. Stepankova and R. Trappl, Eds., *Multi-Agent Systems and Applications. 9th ECCAI Advanced Course ACAI 2001 and Agent Link's 3rd European Agent Systems Summer School, EASSS 2001, Prague, Czech Republic, July 2-13, 2001. Selected Tutorial Papers*, *LNAI*, vol. 2086. Springer, 2001.

[14] T. Springer, T. Ziegert, and A. Schill, "Mobile agents as an enabling technology for mobile computing applications", *Kuenstliche Intelligenz*, vol. 14, no. 4, pp. 55–61, 2000.

[15] Z. Lin and K. Carley, "Proactive or reactive: An analysis of the effect of agent style on organization decision-making performance", *Intell. System in Account. Finance and Manag.*, vol. 2, no. 4, pp. 271–287, 1993.

Jamal Raiyn received the M.Sc. degree in Mathematics and Computer Science from Hannover University in Germany, in 2000 and he finished the Ph.D. study at Leibniz University of Hannover in Germany. In 2010 he finished his Postdoctoral at the Technion in Israel. Since 2002 he is Assistant Professor in Computer Science Department at the Al Qasemi Academic College in Israel. Since 2010 he is the Head of Computer Science Department at Al Qasemi Academic College in Israel. Since 2014 he is the Dean of the Faculty of Science at Al Qasemi Academic College.

E-mail: raiyn@qsm.ac.il
Computer Science Department
Al Qasmi Academic College
Baqa Al Gharbiah, Israel