Paper published in Elsevier Ad Hoc Networks https://doi.org/10.1016/j.adhoc.2015.04.006

Internet of Multimedia Things: Vision and Challenges

Sheeraz A. Alvi[1], Bilal Afzal[1], Ghalib A. Shah*,[1], Luigi Atzori[2] and Waqar Mahmood[1]

[1]Al-Khawarizmi Institute of Computer Science,

University of Engineering and Technology, Lahore 54000, Pakistan.

[2]Department of Electrical and Electronic Engineering,

University of Cagliari, 09123 Cagliari, Italy.

E-mail: {sheeraz.akhtar,bilal.afzal,ghalib}@kics.edu.pk, l.atzori@diee.unica.it, director@kics.edu.pk

Abstract

Internet of Things (IoT) systems cannot successfully realize the notion of ubiquitous connectivity of everything if they are not capable to truly include 'multimedia things'. However, the current research and development activities in the field do not mandate the features of multimedia objects, thus leaving a gap to benefit from multimedia content based services and applications. In this paper, we analyze this issue by contemplating the concept of IoT and drawing an inspiration towards the perspective vision of 'Internet of Multimedia Things' (IoMT). Therein, we introduce IoMT as a novel paradigm in which smart heterogeneous multimedia things can interact and cooperate with one another and with other things connected to the Internet to facilitate multimedia based services and applications that are globally available to the users. Some applications and use-cases for IoMT are presented to reflect the possibilities enabled by this new paradigm. An IoMT architecture is presented which is segregated into four distinct stages; (i) Multimedia sensing, (ii) Reporting and addressability, (iii) Multimedia-aware cloud, and (iv) Multi-agent systems. Moreover, a survey of already existing technologies is done for each of the individual stages, providing a synthesis for the realization of the vision of IoMT. Subsequently, various requirements and challenges as well as the feasibility of existing solutions for each stage of proposed IoMT architecture are comprehensively discussed.

Index Terms

internet of things, internet of multimedia things, wireless communication, architecture, multimedia processing.

I. Introduction

THE ever-increasing services and applications offered by the Internet have explosively widen the span of the global inter-network. Currently, over 9 billion network devices are connected to the Internet facilitating more than 2.5 billion people around the globe for communication (emails, social networks, chatrooms, blogs, forums, etc.), Leisure and Entertainment (games, books, music, videos, shopping, etc.), sharing knowledge (education, geographical information, encyclopedias, etc.), among others services. The recent advancements in designing low-cost small scaled devices, enabled by technologies such as Micro Electro Mechanical Systems (MEMS), have harbingered a monumental surge in the number of Internet-enabled devices [1], [2], [3]. An explosive growth in the number of devices is forecasted over the next decade. Therefore, in addition to traditional machines, e.g. desktop computers, laptops, mobile phones, etc., the physical objects or things around us will be getting the capability to communicate with each other [4], [5].

* corresponding author

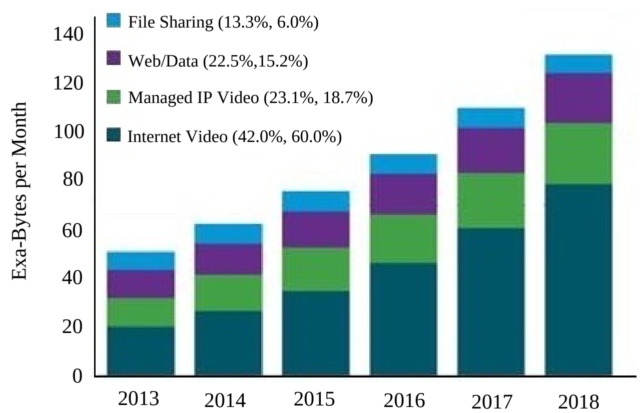

Fig. 1. Global Internet Traffic Usage and Forecast [11].

Smart things equipped with the capability to observe and/or interact with physical environment and the ability to communicate with other things, are extending the Internet towards the so called 'Internet of Things' (IoT) [2], [3]. IoT has the potential to significantly influence our lives and the way we interact with the devices such as sensors, actuators, mobile phones, home automation devices, smart grid devices, etc. [1]. It has promoted concepts of flexible designs, visions and enormous applications, some of them are depicted in [6], [7], [8] and [9]. For individual users, IoT brings useful applications in home automation, security, automated devices monitoring and management of daily tasks. For professionals, automated applications provide useful contextual information frequently to help on their works and decision making. For Industrialists, Internet enabled sensors and actuators operations can be rapid, efficient and more economic. Managers who need to keep eye on many things can automate tasks by connecting digital and physical objects together.

Existing surveys on IoT define it in terms of the things' sensing and actuating capabilities, networking, and web technologies, and market potentials; these also, contemplate the hurdles towards standardization and assess its wider implications for society [1], [2], [3], [5]. However, these research studies do not consider the requirements/challenges posed by multimedia devices or the transportation of multimedia traffic over the network along with other scalar data. Inherent characteristics of multimedia information impose a number of restrictions on the design of IoT, in addition to the challenges imposed by other heterogeneous devices which are part of the IoT. To meet given Quality of Service (QoS) requirements, the network characteristics defined in terms of end-to-end delay, jitter and error rate, among others, are required to be regulated to ensure acceptable delivery of the multimedia content.

There have been enormous growths in multimedia traffic on the global inter-network, due to the huge interest in development and usage of multimedia based applications and services. Real-time multimedia applications, services, and solutions such as video conferencing, remote video-on demand, telepresence, real-time content delivery, online-gaming, etc. have contributed to the exponential growth of the Internet multimedia traffic. The existing balance between non-multimedia data traffic and multimedia-traffic is now shifting away towards an increase in multimedia content specifically in terms of video content. Recent studies on trends and forecast of global Internet traffic [10] have suggested a definite boost in multimedia traffic flow in next five years. In this regard, Cisco carried out an initiative [11] to forecast the trends of the visual networking applications in the global IP traffic on the Internet. Fig. 1 depicts one of the key findings of this report, in which it is clearly shown that the multimedia (video) traffic will significantly dominate the IP traffic on the Internet. This phenomenon is expected to quickly characterize the word of Internet of Things applications as well.

Multimedia content, e.g. audio, video, etc, acquired from the physical environment possess distinct characteristics as compared to the scalar data acquired by typical IoT devices. On the contrary, the multimedia devices require higher processing and memory resources to process the acquired multimedia information. Moreover, the multimedia transmission is more bandwidth hungry as compared to the conventional scalar data traffic in IoT. Introducing multimedia objects fosters a wide array of applications in both commercial and military domains. Some examples are: real-time multimedia based security/monitory systems in smart homes, remote patients monitored with multimedia based telemedicine services in smart hospitals, intelligent multimedia surveillance systems deployed in smart cities, transportation management optimized using smart video cameras, remote multimedia based monitoring of an ecological system, etc. However, augmenting IoT systems with multimedia devices and content is not straightforward and requires the introduction of additional functionalities and the revision of existing ones, bringing to a specialized subset of IoT, which we refer to with ‘Internet of Multimedia Things’ (IoMT).

The physical environment information acquired by the traditional wireless sensor devices in IoT may include statistics about light, temperature, pressure, etc and the things reporting their conditions/states such as the water level in a water-dispenser, battery status or fault reporting for predictive maintenance, etc. The nature of this sensed data is periodic and requires less memory and computational resources. Thus, these applications demand simple processing capability and lower data rates at the sensing device. On the contrary, the multimedia data in IoMT is bulky in nature and specifically for real-time communication higher processing and memory resources are required. Therefore, the multimedia acquisition and communication by current IoT devices is not realizable.

In IoMT, the delivery of multimedia data should be within the bound of QoS constraints (i.e. delay, jitter) which obligate higher bandwidth and efficient communication mechanisms. The routing protocol RPL in current IoT communication stack is flexible and adaptive to operate in energy efficient way as per the application requirements. Yet the current MAC and PHY layer (e.g. ZigBee) proposed for IoT only supports a theoretical data rate of 250 kbps, which is far less than a typical multimedia based application’s requirement. These protocols are adopted for IoT communication stack due to their energy efficient operation. Recently, low-power IEEE 802.11 devices supporting much higher data rates are being designed, for example Qualcomm’s QCA-4002 and QCA-4004-Qualcomm [12], MICROCHIP’s RN171 [13].

In an IoT based system intelligence and action triggering capability is embedded in the devices with the help of sensors and actuators, respectively. Similarly, a cloud enables the capability to develop, maintain, and run, different services by providing scalable computing and storage resources. Thereby, the users are allowed to monitor and control the devices from anywhere and any time. However, the cloud services provided by various cloud servers, such as ThingSpeak [14], xively [15], ioBridge [16], Carriots [17], Axeda [18], among others, is limited to a simple device management system and sensor data logging. Existing cloud servers deployed for the IoT support scaler data processing which is mostly periodic in nature. On the other hand, the multimedia data is continuous in nature and mandates high processing capability. Additionally, the multimedia data requires high bandwidth and storage capacity, specifically the multimedia data processing is highly complex and challenging when supporting real-time video streaming.

Multimedia communication over wireless networks have been addressed in many prior studies, but with a limited scope that is device specific [19], application specific [20], content specific [21], among others. However, none of these studies focus on the system architecture of IoMT which involves heterogeneity of multimedia devices, communication bandwidth issues, complex multimedia processing, cloud services, and other issues. Thus, there is a need to develop a whole new architecture for IoMT paradigm that focuses on these issues collectively.

The goal of this paper is to articulate the perspective vision of IoMT inspired by the concepts of IoT. In addition, the characteristics of the IoMT are thoroughly analyzed and compared with the existing relevant systems. Thereby, distinct requirements and challenges posed by the IoMT paradigm are identified which make it a special subset of IoT. Moreover, some applications and use-cases are presented to reflect the possibilities enabled by the IoMT. Instead of discussing a particular implementation of an IoMT system, we present a potential IoMT architecture which segregates the operation of an IoMT system into four distinct stages; (i) Multimedia sensing, (ii) Reporting and addressability, (iii) Multimedia-aware cloud, and (iv) Multi-agent systems. Moreover, a survey of already existing technologies is done which provides a synthesis and guidance for the realization of IoMT based systems, by considering various requirements and challenges as well as the feasibility of existing solutions for IoMT.

The main contributions of this paper can be summarized as follows:

- To the best of our knowledge this is the first paper to present the concept and vision of the IoMT.
- The distinct architectural design and characteristics of IoMT as compared to the existing multimedia systems are comprehensively discussed.
- The technical specifications and requirements posed by the IoMT systems are identified and discussed.
- The communication protocols designed for IoT are discussed and their feasibility for IoMT is analyzed.
- The potential multimedia processing technologies are presented that can facilitate efficient multimedia communication, specifically via wireless multimedia device.
- The solutions to the processing/computational issues are provided by introducing the notion of multimedia-aware cloud combined with multi-agent systems in IoMT architecture.

The rest of the paper is organized as follows. In Section II, we present the perspective vision of IoMT along with the proposed IoMT architecture. In Section III, we present multiple use-cases to drive the definition of important requirements and necessary changes when considering the IoMT paradigm. Section IV focuses on multimedia sensing and the required coding of the multimedia data. In Section V, IoT communication stack is revisited and its infeasibility to realize IoMT vision is investigated along with highlighting the new requirements and challenges considering multimedia data stream communication. Section VI focuses on multimedia-aware cloud, its essentials and technical specifications for IoMT. Section VII discusses cloud aware multi-agents for service provisioning and composition to broaden the scope of available multimedia data utilization. Finally, Section VIII concludes the survey.

II. IoMT Vision

To better define the role and the major features of the IoMT, we start analyzing the current characteristics of two systems which are currently used to deploy services in major IoT multimedia applications: the Wireless Multimedia Systems, which are used to monitor the environment and where the sensing devices can receive feedback to control the acquisition process; and the Wireless Multimedia Sensor Networks (WMSN), where the multimedia devices have no (or limited) capability to receive feedback.

WMS have been implemented to provide services and applications in several fields, such as surveillance, transportation, telemedicine. Generally, the design of WMS is the one depicted in Fig. 2, where the multimedia sources send the data to Internet through a gateway, where the multimedia content is processed and stored for either synchronous or asynchronous access of the users and the administrators. If we consider a wireless multimedia based surveillance and monitoring system, the multimedia devices can be camera nodes, harvesting the multimedia information from the environment and reporting back to the control center using the underlying wireless technology. Various projects have been carried out such as transportation monitoring and management systems deployed in the city of Irving [22], at the University of Minnesota [23] and at the University of

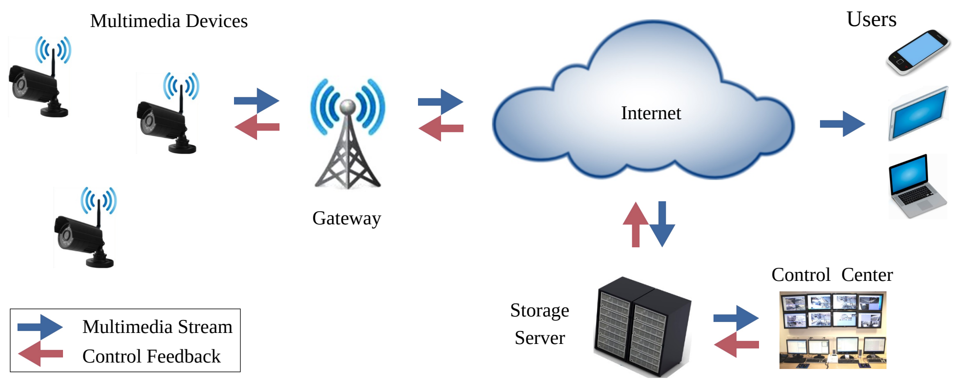

Fig. 2. Typical Wireless Multimedia System architecture

North Texas [24]. In some wireless video surveillance systems, the camera nodes may be reporting the multimedia content to a control center or towards a cloud server via the Internet [25], [26]. The control center gives feedback to alter the camera state (switch on or off) or change the camera position to change the view of interest. In these projects, the multimedia devices are the closed-circuit television (CCTV) cameras with fixed [23], [25] or pan-tilt-zoom (PTZ) movement capability [22], [24], [26], and distinct wireless technologies such as IEEE 802.11 or IEEE 802.16 are employed.

Despite the success of traditional wireless multimedia systems, there are significant limiting factors which restrict the ubiquitous adaption of these systems. Firstly, the scope of these systems is strictly limited to the deployment scenario as per a fixed architecture with restrictive mobility, pre-defined set of multimedia devices that can be adopted, and pre-defined set of functionalities. Secondly, the multimedia devices are generally powered by main energy source. Thus, there is no restriction on energy usage so that the deployed solutions are not energy-efficient. Thirdly, the multimedia devices possessing similar communication stacks are not meant to communicate with other network devices performing different tasks. For example, consider a scenario where cameras are required to start recording when a particular signal is received from an object detector sensor. Therein, first the detector value is sent to control center by the detector and then the response is generated by the control center towards the cameras to start recording. An efficient way to achieve this objective is to enable direct communication between these two devices which may also enhance the scope of possible applications. Fourthly, in cloud based multimedia systems the multimedia content is globally available to users for streaming or processing. Yet, users cannot address the individual multimedia device or trigger different operation on network multimedia devices, since these devices were not designed for two-way communication architecture. As a last point, the cost of multimedia devices is still very high, restricting there large scale deployment and widespread usage in everyday life.

Fig. 3 shows the reference architecture for Wireless Multimedia Sensor Networks (WMSNs), where the multimedia sources have usually limited functionalities and send the content through a WSN (Wireless Sensor Network). The scope of WMSNs is strictly limited to the deployment scenario, in which the characteristics of the network devices are known at the deployment time, thus the network operation is predictable as well as the QoS requirements at the individual device or at the network level are pre-determined. Due to the fixed architecture of WMSNs [19], the individual multimedia devices are not addressable nor they are equipped with any context-awareness or application specific intelligence, that is why a WMSN behave like a single entity to the system. Moreover, the multimedia devices in WMSNs lack true heterogeneity in terms of their resources and capabilities. Thus, the network operations and QoS obligations are not adaptive to current network scenarios and application requirements. The cooperating, communicating and interacting devices are desired to be radically distinct in terms of their

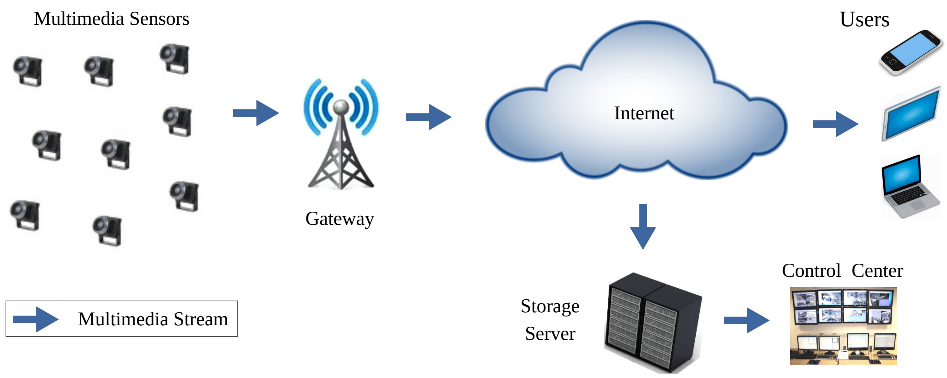

Fig. 3. Typical Wireless Multimedia Sensor Network architecture

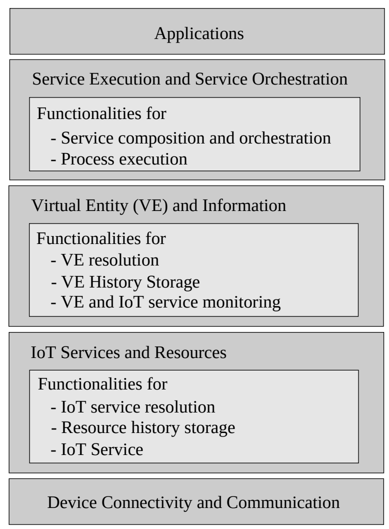

Fig. 4. Representative IoT architecture

resources, communication capabilities, as well as the multimedia content acquisition and processing capabilities.

If we consider the IoT systems that have being recently designed and deployed, we see that several of the mentioned features in WMS and WMSN have to be revised in order to fully integrate multimedia devices in the IoT world. Fig. 4 shows a reference architecture in this regards, as also adopted by the IoT-A European research project [27]. The services and resources layer is meant to provide the discovery and search functionalities, so that things are not restricted to provide data to a single specific vertical deployment but are available to external systems for the benefit of the connected things ecosystem. Virtual entities are introduced with the objective of having a virtual counterpart of the physical objects, which augment the things with additional functionalities implemented by the cloud. As an example, annotations are added in the description of the objects' functionalities as well as past generated data is stored and can be retrieved when needed. Additionally, the object services can be combined with others coming from the Internet of services at the fourth layer, so that different applications can be built on the fly.

As a results of this analysis, Table I shows a comparison of the distinctive characteristics of WMS, WMSN and IoT. This also lists the features of the envisioned IoMT, which are expected to be equal to those of IoT systems (apart from QoS and required bandwidth); however, this does not mean that these features are directly available when introducing multimedia content in IoT. Similar to traditional multimedia networks, the applications based on IoMT require stringent QoS requirements to provide the satisfactory user-experience. Some multimedia applications may be loss-tolerant and yet some other may require

TABLE I
COMPARISON OF EXISTING SOLUTIONS FOR WMS, WMSN, IOT AND IOMT

Parameters	**WMS**	**WMSN**	**IoT**	**IoMT**
Scope of applications	limited	deployment dependent	dynamic & flexible	dynamic & flexible
Node operation	predefined	predefined	adaptive	adaptive
Node resources & capabilities	heterogeneous	homogeneous	limited heterogeneity	heterogeneity
Energy efficiency	not considered	available	available	available
Scalability	low	high	high	high
Interoperability	low	moderate	high	high
Deployment cost	very high	low cost	low cost	low cost
Topology	fixed	limited	ad hoc & dynamic	ad hoc & dynamic
Bandwidth capacity	high	moderate	low	low
Node operation	predefined	adaptive	predefined	adaptive
IP connectivity	limited	limited	uniquely addressable	uniquely addressable
QoS multimedia	available	limited availability	unavailable	available

delay-sensitive communication. The multimedia applications can be categorized into three classes, (i) streaming of stored multimedia content i.e. audio, video, etc, (ii) live streaming of multimedia content, and (iii) real-time interactive multimedia communication. Nevertheless, today's TCP/IP Internet provides 'Best Effort Service' irrespective of the type of the content being communicated and therefore the QoS support is provided (as much as possible) by application-level techniques. Additionally, the bandwidth capacity of IoT systems is usually low as the things are expected to either provide scalar measures about the monitored environment or to receive small commands' packets. Differently, in IoMT the amount of data generated may vary from few kbps to several mbps so that the relevant systems should be adaptive in the offered bandwidth.

Recently, the availability of low-cost low-power multimedia devices, i.e. Complementary Metal Oxide Semiconductor (CMOS) cameras, CMOS-MEMS microphones, etc, have gained lots of attraction in wireless multimedia networks. As the saying goes, 'a picture is worth a thousand words', the multimedia data provides very comprehensive information which can be adequately standardized in appropriate formats, models, and semantic description based on the application context to realize it as an useful information. Consequently, it is expected that IoMT has the potential of enormous number of applications and it will be an essential part of the IoT.

In conventional multimedia networks, the multimedia devices only report the acquired multimedia information from their vicinity. This multimedia content may be streamed to the user or stored at the cloud for processing and on-demand retrieval [19], [28], [29]. Generally, these multimedia devices are designed to be able to communicate to other multimedia devices possessing similar characteristics (homogenous devices) i.e. similar communication stacks, similar resources, etc. Conversely, IoMT envision enabling multimedia devices, such as cameras or microphones, to be globally accessible by a unique IP address with the same spirit as of computers and other networking devices connected to the Internet. Moreover, some computing capability is embedded in the multimedia devices to make them smart enough to perceive the system and service requirements and to trigger actions on their own. Thereby, heterogeneous multimedia devices acquiring multimedia content from the physical environment, i.e. audio, video, images, etc, can communicate and interact with each other as well as to other smart 'things' connected via global network cloud (Internet). This approach enables a wide range of applications in the areas of home and building automation, factory monitoring, smart cities, transportation, smart grid and energy management [30].

As a result of the progressive reduction in the size and cost of the multimedia sensor devices, IoMT based networks are expected to be deployed at large scales where the multimedia devices and/or other things are capable of forming ad hoc links

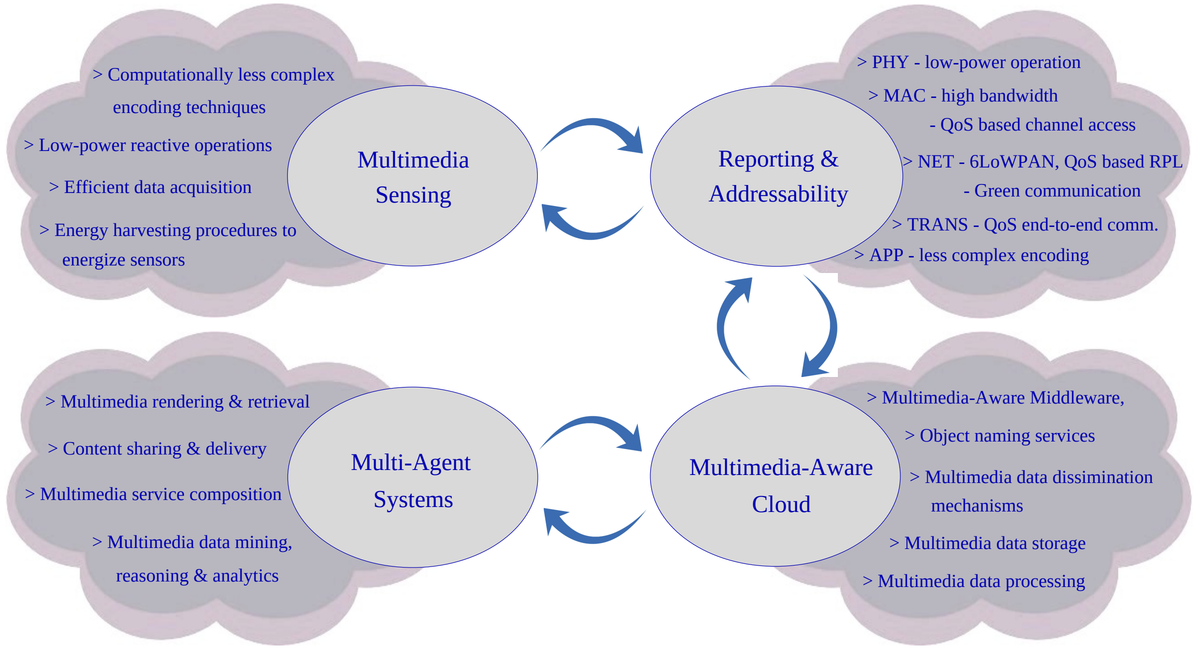

Fig. 5. Proposed IoMT Architecture

with neighbouring things as they move around in the global network. Enabling autonomous organization and self-management of these highly dynamic networks questions the presently available solutions that were designed and developed for predictable network patterns. Similarly, the multimedia data from the various smart things exhibiting heterogeneous functionalities and reasonably different resources needs to be given standardized representation to logically combine and compose the acquired information into the context-aware services for end-users. Nevertheless, the realization and adoption of IoMT or generally the IoT will be reasonably incremental, since its concept and development requires considerable upgradation in both hardware designs as well as the software solutions.

From the above discussion, we can define the IoMT as *the global network of interconnected multimedia things which are uniquely identifiable and addressable to acquire sensed multimedia data or trigger actions as well as possessing capability to interact and communicate with other multimedia and not multimedia devices and services, with or without direct human intervention*.

The incorporation of radically heterogeneous multimedia devices forming a highly dynamic and large network and allowing non-multimedia devices as well, necessitate to reinvestigate the currently employed architectural designs and communication procedures. The operation of an IoMT based service can be described with the help of a four-stage IoMT architecture. Wherein, each distinct stage includes a definite set of processes executed at each step of a service formulation. The proposed IoMT architecture, illustrating various aspects, constraints and challenges at each of these stages to realize multimedia communication in IoMT, is shown in Fig. 5. The functionalities of each of these individual stages are extensively discussed in the following sections of this paper. Firstly, the multimedia data acquisition and encoding methodologies are discussed. Specifically, the operation of multimedia sensing devices and their characteristics are analyzed. Secondly, after the data acquisition process is completed, the next step is to report the multimedia content to the cloud by incorporating efficient communication and addressing techniques. Subsequently, the multimedia content is stored, processed and disseminated as per end-user demand at the cloud. Lastly, the computational and post-processing tasks may be carried out at the cloud according to the application/service requirements articulated by multi-agent systems.

III. Multimedia Sensing in IoMT

The IoMT is an extension to the IoT, where one of the prime objectives is to enable video streaming as part of the realization of IoT. In IoMT, resource constrained low-cost low-power heterogeneous multimedia devices can interact with each other and globally accessible by unique IP addresses with the same spirit as of the computers and other networking devices connected via the Internet. The challenges posed by IoMT are similar to IoT such as dealing with large amounts of information, queries, and computation as well as some distinct requirements. In IoMT based wireless multimedia networks, the multimedia devices are supposed to be small sized objects equipped with a limited amount of power resources, which they have to utilize efficiently to increase network life time. Therefore, energy efficient methods are needed to be devised for network administrative procedures. Similarly, multimedia devices should be embedded with application and context aware intelligence, so that the multimedia content from the physical environment is only acquired when needed, thus minimizing redundant information acquisition.

Unlike the scalar data acquisition by the sensor nodes, the multimedia data acquisition from the physical environment is bulky in nature. Tiny multimedia devices have limited memory resources, thus the acquired multimedia data needs to be quickly processed and transmitted in the air to vacant the space in the memory for the incoming data, as most of the multimedia sources are continuous in nature. In traditional WMS and WMSN, the sensors are constrained devices in terms of their energy, processing and computational resources. These sensor devices are supposed to report the acquired data to the actors, which are intelligent and resource rich devices possessing high energy, processing and communicating capabilities [31].

The actors generate reactive procedures in response to the event information received from the sensor nodes. The actor nodes are much lower in number as compared to the sensor nodes; thereby a large number of sensor nodes are dependent on a single actor node even for low level instructions such as pan-tilt-zoom setting for camera. Therefore, in case of dynamic network environment an actor nodes' failure results in collapse of whole network communication and operation, due to the disintegration of the network into many sub-WSNs [32]. For this reason, in IoMT architecture it is desirable that the multimedia nodes are equipped with some intelligence, thus the local responses can be generated at the sensor node level. Unlike WMS/WMSN actor nodes, the multimedia sensor nodes in IoMT are deficient in terms of energy, processing and memory resources, so the event driven responsive procedures should be computationally less complex.

The acquired huge amount of multimedia data undergoes various pre-transmission processing procedures at the multimedia device i.e. transformation, quantization, estimation, entropy coding, etc, so that it can be compressed to reduce bandwidth requirements while transmission. These processes are computationally complex and consume significant amount of energy. However, various promising solutions proposed for efficient multimedia communication like compressive sensing [33], distributed video coding [34], etc can be utilized to facilitate multimedia acquisition and processing at the multimedia device in IoMT. Compressing video to enable a certain level of quality over a low capacity channel (offering low data rate), increases encoder complexity and energy consumption. On one side, energy constrained IoMT devices exhibit limited bandwidth capacity, but enabling a good video quality requires high compression that is infeasible due to high energy consumption. Thus, there is a tradeoff between the achievable compression and the energy utilization for a specific level of user experience restriction.

Mostly, the wireless multimedia devices are expected to be battery powered. Since, the multimedia acquisition and its processing are very power consuming procedures. Thus, it is deemed to make multimedia sensor networks capable of harvesting as much energy as they can from the environment. Consequently, in IoMT paradigm self-powered multimedia sensor devices are reckoned to harvest energy from different energy sources in the network vicinity. For this reason, in [35] integrating

TABLE II

TYPICAL ENERGY HARVESTING TECHNIQUE.

Source	**Source of energy**	**Source power**	**Harvested power**	**Efficiency**	**Issues**
***RF**-GSM 900MHz*	RF harvester in urban environment	0.3 to 0.03uW/cm^2	0.1uW/cm^2	∼50%	Coupling and rectification
***RF**-GSM 1800MHz*		0.1 to 0.01uW/cm^2			
Thermal	Human body heat dissipation in air	20mW/cm^2	25uW/cm^2	∼0.1 to 3%	Small thermal gradients; efficient heat sinking
Vibration	Human walking with harvester on the floor	0.5+1m/s@1+50Hz	4uW/cm^2	∼25 to 50%	Variability of vibrational frequency
***Photovoltaic**-Indoor*	Harvester inside a building environment	0.1mW/cm^2	10uW/cm^2	∼10 to 24%	Conform to small surface area; wide input voltage range
***Photovoltaic**-Outdoor*	Harvester open air in a sunny day at noon	100mW/cm^2			

multiple energy sources such as the solar cells, piezoelectric, thermoelectric and radio waves recharging devices, for energy harvesting from photovoltaic, vibration, thermal, radio frequencies, etc, is promoted. The potential harvested energy from these sources along with their respective efficiencies are presented in Table II. To prolong network lifetime, in addition to batteries the multimedia devices should be equipped with multiple energy sources i.e. solar cells, thermoelectric harvesters, etc. This will result in a substantial increase in usability of IoMT in harsh environments where battery replacement is not practical.

One of the primary aspects of the multimedia data acquisition is to encode the sensed raw multimedia information. However, contemporary video coding frameworks have been driven by a broadcast scenario in which the video is to be compressed only once and decoded each time it is played back, mandating complex encoder and simple decoders. Conversely, IoMT applications are more akin to ‘uplink’ video transmission and pose a new set of stringent requirements on video codecs as given below:

- *Low-complexity encoding*: Motion vector calculation in predictive encoding is a computationally intensive task. IoMT based multimedia devices have limited energy and processing capability, so the complexity should be shifted to the cloud.
- *Resilience to transmission errors*: Traditional video encoders are highly susceptible to transmission errors, which tend to spread over several frames due to their predictive nature. Thus in order to achieve good reconstruction quality, error resilience is imperative.
- *High data rate with low-power*: ZigBee has been adopted for IoT due to its energy efficient operation. However, even for transmitting a low resolution CIF video, which is 352×288 pixels per frame at 12 frames per second, the required bit rate well exceeds the 70 kbps [36] offered by ZigBee. MAC protocols such as IEEE 802.11, EV-DO, LTE do offer high data rates for video transmission but not in a power efficient manner.
- *Delay bound for multimedia streaming*: Multipath fading corrupts a number of adjacent bits due to recurring deep fades. At transmitter, interleaving is applied so that errors are spread apart while doing de-interleaving at the receiver. However, interleaving distances at the transmitter and reconstruction algorithms at the receiver limits real-time delay bound.

Considering these set of requirements, the video coding techniques are categorized into three broad classes (i) conventional video coding, (ii) distributed video coding, and (iii) compressive sensing.

A. Conventional Video Coding

Current video compression/encoding standards such as H.26x and VPx are primary choice for broadcast scenarios. Compression in these standards is achieved by exploiting both the spatial and temporal redundancies inherent in video frames at

TABLE III
MULTIMEDIA COMPRESSION TECHNIQUES.

Parameters	**H.264/AVC**	**H.265/HAVC**	**VP9**	**Daala**
Compression Ratio	50:1	100:1	60:1	30:1
Complexity	High	Very High	Medium	Medium
Transform	DCT	DCT 8x8 and 4x4	DCT 32x32 to 4x4 + DST Luma intra 4x4	Lapped Transform
Intra prediction	Up to 9	35	10	Frequency domain predictors
Color Spaces	YUV 4:2:0 -4:4:4, 8-14 bits	YUV, RGB, YCbCr, 8-12 bits	YUV 4:2:0, 8 bits	YUV 4:4:4, 4:2:2, 4:2:0 sub-samplings, 8-bit, 10bit

the encoder. The H.264/AVC is a well-known video encoding standard developed in 2003. H.264 mainly relies on a hybrid of block based transform to exploit spatial redundancy and inter-frame motion compensated predictive coding. However, the H.265 standard designed in 2013 for higher resolution video coding and achieves better rate-distortion performance as compared to H.264 [37]. It is also based on block based transform coding. Major improvements include an increase in maximum block size from 16×16 to 64×64, quad-tree partitioning with adaptive block sub splitting and adaptive prediction units and transform units partitioning. Intra-frame coding efficiency is improved by allowing up to 35 intra prediction directions.

Another popular encoding standard VP9 is developed by Google for the video streaming applications. VP9 shares a number of features with H.264. Major improvements include the increase of largest block size to 64×64 with adaptive sub-splitting and enhanced motion vector calculation. The intra prediction in VP9 still proceeds in 10 directions. Performance comparisons of these three video compression schemes are referenced therein [38], [39], [40]. Authors in [38] presented the comparative analysis of these encoding techniques, using x264 [41] reference software for H.264, HM [42] for H.265 and VP9 [43] encoder by Google. Results show that H.265 provides average bit-rate savings of 43.3% and 39.3% over VP9 and H.264, respectively. It was also reported that VP9 encoder produces an average bit-rate overhead of 8.4% at the same objective quality over the x264 encoder. Meanwhile the average encoding times for VP9 are greater than x264 encoding times by a factor of over 100 and around 7.35 times lower than HM.

DAALA is another video encoder which uses different encoding techniques as compared to aforementioned encoders. In DAALA, instead of block-based Discrete Fourier Transform (DFT), lapped transforms are used while pre and post filtering instead of deblock filtering. Other distinctive features include frequency domain intra prediction and time/frequency resolution among others. Experimental results in [40] depicts that the rate distortion performance of DAALA encoder is significantly inferior to those of HM, VP9 and JM reference software for H.264. A feature wise comparison for these three encoder is summarized in Table III.

B. Distributed Video Coding

Distributed video coding (DVC) is primarily designed for an uplink scenario and thus offers the converse complexity distribution. It is based on two information theoretic theorems: Slepian-Wolf [44] and Wyner-Ziv [45]. A well-known result from Shannon's work states that the lossless rate for two correlated encoders X and Y is lower bounded by their joint entropy H(X,Y). Slepian and Wolf established that this rate is still achievable with an arbitrarily small error if the two sources are encoded separately as long as they are jointly decoded and the correlation information is known both at the encoder and the decoder. Wyner and Ziv in [45] extended this result for lossy coding. It is shown that for two correlated Gaussian sources X and Y, and for a mean square error distortion criterion, there is no loss in conditional rate even when the side information Y is available only at the decoder. These results lead to the proposal of a number of video coding architectures in which video frames are encoded jointly but decoding is conditioned on the side information available at the decoder.

An early implementations for DVC is *PRISM* [46]. PRISM combines the features of intraframe coding with inter frame coding compression efficiency. In this implementation DCT transform is applied to macro-blocks. The blocks are classified based on their temporal correlations and subsequently either intracoded, syndrome coded or skipped. At the decoder end, the frame blocks in skip class can be reconstructed by the collocated blocks in the previously reconstructed frame. The frame blocks in the intra coding class are reconstructed by the traditional decoder. Syndrome encoded blocks are decoded by performing motion estimation by using CRC bits. The previously decoded sequences and multiple candidates' side information generation are used for calculating motion estimation.

Stanford architecture was another early proposal proposed around the same time as PRISM [47], [48]. In this architecture the video sequence is divided in Group of Pictures (GOP) and the first GOP is encoded with H.264 intraframe mode. The rest are encoded using WZ encoding and subsequently passed through a turbo encoder. At the decoder side, information is generated using the motion-compensated interpolation of previously decoded frames and used in turbo decoding. Another proposed architecture, DISCOVER builds on the Stanford architecture and introduces several enhancements [49], [34]. Notable enhancements include motion vector smoothing for SI generation, rate estimation, virtual channel model and soft input calculation. Performance of this codec is extensively profiled and available in [50].

C. Compressive Sensing

Compressive sensing (CS) theory suggests that a sparse or compressible signal can be recovered from a small number of linear and incoherent measurements [51], [52], [53]. Numerous methods for video encoding have been proposed in the literature which are based on compressive sensing. The simplest method encodes videos as a series of compressed sensed images. Each frame is acquired separately at the encoder and subsequently reconstructed individually at the sink. These techniques however, do not exploit the temporal correlations between adjacent frames. Various other techniques have been proposed which make use of the temporal correlation either for greater compression [54], [55] or enhanced reconstruction [56].

For enhanced reconstruction, the encoder is kept as simple as possible while the temporal correlation is exploited at the receiver during reconstruction. In [56], this is achieved through reconstruction of the frames in the KLT basis. Correlation matrix for a macro-block in a given frame is reconstructed heuristically from a few previously decoded frames adjacent to the given frame. Other methods which exploit the temporal correlations during the decoding process only, typically bank on total variation minimization [57], [58], [59]. Methods which exploit temporal correlation at the encoder are hybrids of CS based and traditional video coding schemes. Their complexity at the transmitting node therefore is typically more than those based on compressive sensing only. DISCOS presented in [54] divides a video into I-frames and CS-frames. The scheme uses standard video compression (such as MPEG/H.26x intra-encoding) on I-frames. The CS-frames, however, are encoded using CS through a combination of both frame-based calculations and block-based calculations.

The comparison of above mentioned encoding techniques is as follows. DVC implementation i.e. DISCOVER, outperforms H.264/AVC intracoding, except for scenes with complex motion. Performance margin is higher for scenes with low motion. For scenes with simpler motion structures, DISCOVER matches the performance of H.264/AVC [34], [50]. Compressive video sensing (CVS) techniques provide inferior coding efficiency as compared to other coding schemes. CVS schemes provide the least complex encoders. Mostly, computations at the encoder involve measurement vector generation through measurement matrix. DVC schemes also provide low-complexity encoders as compared to conventional coding schemes, which employ complex motion estimation at the encoder. Decoders for CVS and DVC schemes entail most of the complexity. In CVS, iterative algorithms are used to find minimum values for non-linear objective functions. DVC decoders need to generate side information and perform FEC decoding. A simple CVS based encoder provides several advantages for IoMT based energy contained environment. As mentioned earlier, existing MAC layer protocols support high data rate for video streaming but consume a lot of energy. Within an energy budget, a simpler encoder may leave aside enough energy for transmission protocols with high energy demands. Conventionally, there is a trade-off between error resilience and low encoder complexity (manipulated by moving to different rate-distortion points). Fortunately, compressive sensing based techniques provide both these desired features simultaneously.

D. Open issues

- IoMT multimedia devices are expected to be low-cost and tiny, possessing limited memory resources. Thus, efficient data acquisition methodologies should be developed which can alleviate the burden on memory resources.
- Conventionally, the multimedia devices are considered dumb and all the intelligence is kept on the control centers. However, embedding low level local intelligence on small multimedia devices can enable enormous application in automation and robust system operation.
- The processing power for IoMT multimedia devices is curbed, therefore the operation of existing multimedia encoding schemes on these devices is infeasible. Consequently, computationally less complex encoding techniques are required, in this matter compressive sensing based encoding techniques seem like a potential technology.
- Higher compression of acquired multimedia raw data needs higher processing resources as a result the transmission bandwidth requirement is relaxed and vice versa. Therefore, it is a trade off between the level of compression achieved and the bandwidth requirement. An interesting issue is to minimizing the processing overhead which is directly proportional to energy consumption for a given set of bandwidth and compression requirements.
- Multimedia sensing demands high processing and continuous data acquisition, which results in higher energy consumption. Since, the IoMT devices are expected to be operated on batteries that may not last longer due to demanding nature of multimedia data. Thus, efficient energy harvesting procedures need to be devised to energize sensors and prolong the network lifetime.

IV. Reporting and Addressability

The transmission of the multimedia content from the sensor to the application server, typically located in the Internet cloud, imposes stringent traffic management requirements as compared to the gathering of data from scalar sources, especially in terms of the bandwidth, and reliability [28], [29]. Specifically, these requirements are critical in case of real-time multimedia streaming, which is continuous in nature and it may or may not be delay tolerant depending upon the given application requirement, e.g., in smart business management or automatic public security systems.

IoMT paradigm requires communication standards similar to HTTP, IP, TCP, etc. However, their direct adaptation in IoMT is not straight forward, since these protocols are not designed for energy constrained devices. On one hand, IoMT multimedia devices are realized as low-cost, battery powered, low-complexity devices. On the other hand, existing protocols consume significant amount of energy due the redundant data transmission, protocol overheads, headers transmission, acknowledgement packets exchange for higher layers to ensure reliability, etc. Since, these communication protocols are not optimized for low-power communication, thus the adaptation of these mechanisms is infeasible in their current form and structure.

The current standardization activities to enable Internet-connectivity to ‘Things’ [60] are not focused to address the challenges provisioned by the multimedia communication over IoT. The main obstacles of realizing IoMT, enabling Internet access to wireless multimedia network devices, are limited available power, limited available capacity, and heterogeneity of multimedia devices. In this section, the IoT communication stack is analyzed and its potential to be adopted for IoMT is reviewed. Since, multimedia communication pose stringent requirements and challenges, that is why some of the protocols designed for IoT may not be feasible for IoMT. Thereby, we suggest some potential alternate techniques and protocols which may perform better in supporting multimedia communication over IoMT.

A. *IoT Communication Stack*

For low-power communication in IoT, IEEE 802.15.4 [61] is proposed which defines both Physical Layer (PHY) and Medium Access Control (MAC). The PHY/MAC layers defined by IEEE 802.15.4 formed the basis of ZigBee 1.0 and later for ZigBee 2006 [62]. The single-channel characteristic of IEEE 802.15.4 MAC is unreliable in multi-hop network scenarios. Therefore, Dust Networks [63] introduced a channel hopping protocol named Time Synchronized Mesh Protocol (TSMP) [64] to mitigate multi-path fading. TSMP inspired IEEE 802.15.4e working group to design a time synchronized channel hopping protocol, which later integrated into the IEEE 802.15.4 in 2010. To facilitate Internet connectivity for low-power devices IETF working groups proposed a set of low-power based communication protocols i.e. 6LoWPAN [60] as a convergence layer, ROLL RPL [65] as a routing protocol, and CoAP [66] as an application layer protocol. These protocols, summarized in Table IV, have been considered as key technologies in realization of IoT. In the later part of this section, the functionalities of each of these protocols are discussed.

1) Link Layer: The IEEE 802.15.4 [61] provides design specifications for a low-power Wireless Personal Area Networks (WPAN). IEEE 802.15.4 PHY layer uses Offset-Quadrature Phase-Shift Keying (O-QPSK) modulation with Direct Sequence Spread Spectrum (DSSS) to provide 250 kbps data rate. It also defines 2 MHz wide 16 non-interfering orthogonal channels, located every 5 MHz between 2.405 GHz and 2.480 GHz. The radio can subjectively send and receive on any of these channels and is able to switch channels in less than 192 μsecs. When a radio sends a packet, it starts by transmitting a physical preamble for 128 μsecs to allow the receiver to lock to its signal. Preamble is followed by a Start of Frame Delimiter (SFD) to indicate the start of the physical payload. The first byte of the physical payload indicates the length of the payload itself. Its maximum value is 127, which limits the maximum length of a packet to 128 bytes. Based on SFD and packet length, the receiver estimates the earliest time after which it can switch off its radio.

IEEE 802.15.4 MAC protocol proposes two network topologies i.e. star and mesh topologies. In star topology, dedicated relay nodes referred as Full Function Devices (FFDs) are required to forward data from Reduced Function Devices (RFDs) to Network Coordinator (NC) node. Whereas, in mesh topology each node is a FFD. However, in multi-hop networks the FFDs are required to maintain 100% duty cycle, resulting in lower node/network life time. IEEE 802.15.4 MAC is based on a Time Synchronized Channel Hopping (TSCH) [61] in which a node can transmit, or receive, or sleep, according to a known schedule in a slot-frame structure. A maximum length packet and an acknowledgement packet can be transmitted in a slot. While in a sleeping slot, the radio is switched off to save power. A simple backoff scheme is also devised to avoid collisions among channel competing nodes. Moreover, IEEE 802.15.4e TSCH MAC promotes channel hopping which provides frequency diversity, thereby the effects of interference and multipath fading can be mitigated. A ‘manager’ node is responsible

TABLE IV
TCP/IP PROTOCOL SUITE VS IPSO PROTOCOL SUITE.

IP Layers	**Internet Protocol TCP/IP protocol suite**	**IP for Smart Objects (IPSO) protocol suite**
Application Layer	HTTP, SMTP, FTP, etc	CoAP
Transport Layer	TCP, UDP	UDP
Network Layer	IPv4, IPv6	6LoWPAN
Link Layer	Wi-Fi, WiMAX, etc	ZigBee

for maintaining the network schedule for each node depending upon the demand of the nodes. Manger node also draws a connectivity graph using the neighbours information received from the network nodes.

2) Network Layer: For smart objects in IoT or IoMT, the IPv6 is recommended due to the exhaustion of IPv4 [30]. However, enabling IPv6 for resource constraint devices is not straight forward. For example, minimum MTU (maximum transmission unit) size for IPv6 is 1280 bytes, and without fragmentation it cannot be transmitted using 128 bytes maximum packet size of IEEE 802.15.4. Thus, the IETF IPv6 over low-power WPAN (6LoWPAN) working group has defined an adaptation layer in order to compress IPv6 headers and fragmentation of large packets [67]. The adaptation layer resides between IP and link layer, in order to suppress redundant headers as some of the headers information can be inferred from other layers and secondly to fragment large sized packets. The Internet Protocol Header Compression (IPHC) encoding technique [68] defined for LoWPAN can compress unique local, global, and multicast IPv6 addresses up to 2 bytes. Likewise IPv6 Next Header Compression (IPNHC) encoding technique is used to compress next headers.

For low-power and lossy link networks (LLNs), an effective IPv6 routing protocol for Low-Power Lossy Network (RPL) is proposed by IETF. In RPL, a root device collects data by coordinating with other devices through multi-hop routes. Root device creates a Destination Oriented Directed Acyclic Graph (DODAG) based on the link costs and other device attributes, which are weighted using an objective function according to the requirements. The RPL protocol determines both the downward and upward routes between root devices and sensor/actuator devices. The information of end devices about the Rank, Objective Function, IDs and so on, is embedded in DODAG Information Option (DIO) messages. These DIO messages are shared periodically by devices with their neighbors to create DODAG and route towards the root device. Using these DIOs, nodes select a parent device as per implementation requirements identified by the objective function. In addition, RPL exploits key network metrics such as: node energy, hop count, link throughput, latency, link reliability, and link color, to effectively adapt to the network channel conditions [69].

3) Transport Layer: The concept of IoT to enable IP connectivity for resource constraint devices is not realizable over the current higher layer communication protocols due to their high protocol overhead. Therefore, much work is going on to possibly suppress or at least compress the overheads of the transport and application layer data assuring same level of application experience as enjoyed by the current Internet devices.

At the transport layer, end-to-end protocol TCP ensures reliability by providing flow control and congestion control. However, it comes with additional overhead in terms of bandwidth efficiency, energy consumption and memory usage. Although, User Datagram Protocol (UDP) provides best effort transport with no guarantees of packet delivery, but retransmission control at the application layer combined with UDP can give an effective tradeoff between energy efficiency and reliability. Moreover, the UDP headers can be compressed by using LoWPAN NHC [67]. Since, UDP Length field can be omitted estimating the size from MAC layer or fragmentation headers, transmitter and receiver ports may also be omitted if they are the same, in

TABLE V

COMPARISON OF IEEE 802.11, LOW-POWER WI-FI, AND ZIGBEE [71]

Parameters	**Standard 802.11**	**Low-Power Wi-Fi**	**ZigBee**
Max Data Rate (Kbps)	54000	54000	250
Tx energy (nJ/bit)	6	4	300
Rx energy (nJ/bit)	20	15	280
Transmit power (dbm)	18	18	0
Sleep (mW)	400	15	5
Wake-up time (msecs)	500-1000	8-50	2

this way at least UDP header can be compressed up to 2 bytes. Thus, UDP is considered to be more feasible for IoT based on 6LoWPAN architecture over TCP.

4) Application Layer: The resource constraint LLN devices cannot support traditional client/server model defined by HTTP, due to the large pay load size which requires fragmentation for IEEE 802.15.4 devices. Therefore, limiting the packet size at the application layer is necessary to make its operation compliant with restrictive lower layers. Thus, the IETF CORE working group [70] has proposed the Constrained Application Protocol (CoAP) [66]. CoAP translates HTTP for resource constraint devices and this is achieved by multicast support and reduced overhead. The main features addressed by CoAP [66] are constrained web protocol, application layer reliable Unicast or best-effort multicast support, asynchronous message exchanges, low header overhead, content-type support, optional resource discovery, among others.

In the HTTP client/server based architecture the end points have definite roles. However, CoAP is an asynchronous request/response protocol based on a datagram oriented transport i.e. UDP, where both endpoints act as clients and servers. The architecture of CoAP is divided in two layers, a message layer and a request/response layer. The message layer controls message exchanges over UDP between two endpoints. ‘Requests’ and ‘Responses’ share a common message format and are used for reliability and to detect duplicates. In request/response layer, CoAP messages carry request and response semantics including a method code or response code, respectively. Since, CoAP is implemented over non-reliable transport layer, thus CoAP messages may arrive out of order, appear duplicated, or be lost without notice.

B. Infeasibility of IoT’s Communication Stack for IoMT

The present IoT protocol stack is defined considering low-power smart devices that are communicating and exchanging scalar data that is periodically acquired from the physical environment. However, the multimedia data specifically real-time continuous multimedia data is bursty in nature. Since, the communication requirements of IoMT have been over looked while designing present IoT protocol stack, therefore current IoT communication mechanisms are inadequate for IoMT traffic. These limitations of IoT communication protocols are discussed in the following part of this section.

The IEEE 802.15.4 based link layer specifies a well suited trade-off between energy-efficiency, transmission range, and data rate, which makes it particularly suitable for small to medium sized WPAN. However, the multimedia bandwidth requirements are significantly higher as compared to the offered data rate i.e. 250 kbps. In case of real-time continuous multimedia communication, the continuously generated data requires large packet size so that the bandwidth is utilized efficiently, maintaining the delay and jitter under the required thresholds. Thus, the IEEE 802.15.4 small MTU size, i.e. 128 Bytes, increases delay and jitter. It also requires fragmentation at higher layer which comes with additional protocol overhead and cost in terms of bandwidth and energy efficiency.

TABLE VI

ENERGY AND TIME COMPARISON OF 6LOWPAN VS LOW-POWER WI-FI [71]

Parameters	**6LoWPAN/ZigBee**	**Low-Power Wi-Fi**	
Data Rate (Kbps)	250	1000	54000
Packet size (bytes)	1000	1000	1000
Time (msecs)	23.61	25.82	16.58
Energy (mJ)	9.17	8.46	1.28

Furthermore, many off-the-shelf video devices based on IEEE 802.11 [72] standard have been widely used nowadays, due to the higher data rate support. These multimedia devices, largely deployed for video surveillance and monitoring applications, makes Wi-Fi a good candidate for IoMT if its energy efficiency mechanism is devised comparable to IEEE 802.15.4 [61]. Recently, there have been much interest in designing very low-power IEEE 802.11 devices, for example Qualcomm's QCA-4002 and QCA-4004-Qualcomm [12], MICROCHIP's RN171 [13], among others. In this low-power Wi-Fi the device's duty cycle is decreased to get longer battery lifetime which in-turn increases network lifetime. Similarly, the channel access mechanism is made efficient in terms of bandwidth utilization. In addition, the hardware designs are optimized to draw minimal current in sleep state.

A comparative study on low-power Wi-Fi, ZigBee and IEEE 802.11 standard is done in [71], some statistics are presented in Table V. IEEE 802.11 standard higher data rates support can effectively decrease the reception and transmission time of the packets providing a higher bits exchange per energy consumed. Thereby, efficient bandwidth utilization combined with the power management system can enable low-power Wi-Fi as a promising communication technology for IoMT. A comparison of low-power Wi-Fi with 6LoWPAN/ZigBee is shown in Table VI, for details and more comprehensive analysis reader is referred to [71]. However, for a detailed analysis on the energy efficiency and throughput tradeoff see [73].

The TSCH protocol [61] efficiently exploits multiple orthogonal channels to allow multiple active links simultaneously. When combined with slot-frame structure, this approach allows nodes to maintain lower duty cycle as low as 1%. The time slot structure can be inefficient in terms of bandwidth utilization, when nodes have variable traffic requirements. Therefore, a MAC protocol for IoMT is required which provides specific QoS depending upon the traffic requirements of the nodes. In addition, due to the continuous traffic nature of multimedia applications, maintaining a very low duty cycle is hard. Therefore, additional energy saving techniques should be adopted maintaining a trade-off between duty-cycling and energy efficiency while assuring the delay or jitter bounds.

The routing protocol, RPL, used at the network layer is proactive. The routes are maintained using DIO messages which are transmitted by each parent node in the network and the status of an already build link is known through keep alive messages. It seems that the availability of the routes can be supportive in providing lower latency or jitter for multimedia traffic. However, it has been investigated in [74] that the proactive routing protocols have sub-optimal performance in multimedia communication in terms of packet end-to-end delay, overhead per packet, packet delivery ratio. It is due to the unnecessary route establishment and maintenance overhead that wastes bandwidth and energy. An energy efficient RPL based routing scheme is proposed in [75] which considers energy efficiency as well as link quality for determining the routes. Moreover, a cluster based RPL routing protocol is proposed in [76], which is reported to reduce packet loss ratio and end-to-end delay. Similarly, such kind of routing scheme should be devised considering multimedia traffic requirements. Thus, existing operation of RPL is infeasible for multimedia traffic and a preferably reactive routing protocol is needed which considers network metrics to establish links and yet requires less routing overhead.

For 6LoWPANs UDP is promoted due to its lower overhead operation. Further overhead reduction is achieved using header compression techniques. For multimedia traffic the UDP is considered more suitable transport layer protocol particularly for real-time multimedia communication. To improve multimedia communication, other transport layer protocols designed on top of UDP could be used. For example, the Lightweight User Datagram Protocol (UDP Lite) [77] provides partial checksum as well as data checksum (only dropping the packet if header is erroneous). Similarly, the Datagram Congestion Control Protocol (DCCP) [78] which has slightly larger header size (12 to 16 Bytes) as compared to UDP (8 Bytes), nevertheless it is connection oriented like TCP supporting path MTU, data checksum, partial checksum, and congestion control, which makes it more suitable for timing constraint multimedia application. Multimedia transmission protocols, Real-time Transport Protocol (RTP) and Real-time Control Protocol (RTCP) [79] facilitate the real-time multimedia services. RTP provides functionalities for time constraint applications using time stamp, sequence numbering, and payload specifications. Whereas, RTCP provides feedback on the quality of multimedia content delivery. Therefore, real time streaming protocols are required to initialize sessions and retrieve multimedia content from servers till the end of the sessions, considering the processing capability and the buffer size of the user device.

At the application layer, one of the most important aspects for multimedia applications is decoding of the compressed multimedia data sent by the multimedia devices. There is a clear trade-off between the achievable compression and the encoder/decoder complexity. Detailed discussion on the operation and limitations of existing encoding techniques along with the potential solutions are already discussed in Section III. From the above discussion, it can be concluded that the currently employed communication stack for IoT is infeasible for multimedia communication in 6LoWPANs in terms of bandwidth utilization and energy efficiency. In IoMT therefore, energy efficient, high data rate supportive and low overhead communication protocols are required to facilitate multimedia traffic, considering acceptable level of jitter, end-to-end delay, latency, etc over resource constraint devices in LLNs.

C. Open issues

- One of the most critical challenge for the IoT communication stack to support multimedia communication is its higher data rate requirement. IEEE 802.15.4 was never meant to deal with multimedia traffic, therefore unlicensed wireless technology providing higher data rate such as Wi-Fi should be adopted for IoMT. However, energy efficiency along with QoS support by Wi-Fi that is devisable to ZigBee is an open challenge.
- Routing protocol is highly flexible and adaptive. However, proposed routing metrics for route selection are based on link quality, energy, or path length. For IoMT new routing metrics should be considered that may energy efficiently route the multimedia traffic for specific QoS bounds and requirements.
- In an IoMT system, the end-user devices multimedia streaming support may be radically different from a smart phones with a limited cellular bandwidth to a high-end desktop computer with a high speed broadband Internet connection. Therefore, real time streaming protocols are required to initialize sessions and retrieve multimedia content from servers, considering the processing capability and the Internet bandwidth of the user device.
- In IoMT system, the acquired multimedia video needs to be encoded using low-complexity encoders to alleviate the processing and energy consumption overheads from the multimedia sensing devices. Therein, high compression can be achieved but then complexity of the decoder at the end device is radically increases. Therefore, the trade-off between the compression and complexity at the sensing device and the complexity of the decoder needs to be investigated.

V. Multimedia-Aware Cloud

The IoMT based networks are expected to be deployed on large scales, thus the massive amount of multimedia information generated by these networks has to be managed and filtered efficiently in order to support various IoMT applications and services [80]. Traditionally, a 'multimedia-aware cloud' is intended to provide services for huge data storage and powerful distributed processing services to the end-user as if the cloud is a super-computer providing ubiquitous access to the enormous amount of multimedia content from sensor networks. In IoMT, however the robustness and fault tolerance, data integration from heterogeneous devices/networks is also of critical importance which needs to be investigated [81], [82]. For this reason, along with multimedia content dissemination, processing and storage, a multimedia aware-cloud has to incorporate heterogeneity at various levels including devices, networks, applications and services. In addition, the multimedia-aware cloud has to provide application/service specific QoS adaptation, in terms of bandwidth and delay while performing tasks such as storage, delivery, sharing, submission and retrieval, for enormous number of heterogeneous end-user devices [81], [82], [83]. Consequently, a 'multimedia-aware middleware' supporting distinct multimedia networks seems like a potential solution to address this heterogeneity issue.

The integration of cloud with smart things can provide enormous number of services. For example, the users can be enabled to ubiquitously access the sensor data from remote sensor devices as a Sensing as a Service (SaaS) [84], rule engines can be implemented to control the actuators operation automatically from the cloud as a Sensing and Actuation as a Service (SAaaS) [85], providing control to identity and policy management systems Identity and Policy Management as a Service (IPMaaS) [86], enabling access to video analyses and streaming of recorded video content in the cloud Video Surveillance as a Service (VSaaS) [87].

A. *Multimedia-Aware Middleware*

The integration of wireless heterogeneous networks, i.e. WiFi, WiMAX, sensor networks, etc, has attracted lot of attention of the research community. Supporting multimedia traffic over these networks mandates new standardization of architectures in the field of communication and networking. The multimedia-aware middleware has to be adaptive to enable service provisioning in heterogeneous multimedia networks for pervasive end-users. Moreover, the multimedia middleware should also render distinct multimedia services such as multimedia data aggregation, reasoning, learning, filtering, recommendation, adaptation, and delivery [88], [89]. Thus, the middleware should acquire multimedia information from the heterogeneous multimedia devices depending upon the application/service requirements and then multiple multimedia services procedures operated by the multi-agent system are initiated on the acquired multimedia content.

The multimedia service procedures enable distributed data processing on multimedia content coming from heterogeneous multimedia devices, providing the abstraction of multimedia devices to multi-agent system for composing distinct multimedia applications/services and distributed multimedia traffic flow control. Since, the multimedia communication and processing pose application specific QoS requirements [90]. Thus, the middleware is required to facilitate the management of network resources, e.g. efficient bandwidth utilization and computational capacity, based on the context information. Additionally, the multimedia middleware should not only adapt to the disparity in multimedia content and computational algorithms, but also deal with the data integration according to the dynamics of the system environment.

Similarly, multimedia-aware middleware should be adaptive to support different multimedia services such as video calling, multimedia streaming, object tracking, content sharing, data storage, etc, by enabling various service components like QoS, mobility, security, tracking, QoE, etc. Since, the end-user devices can be radically different in terms of multimedia processing

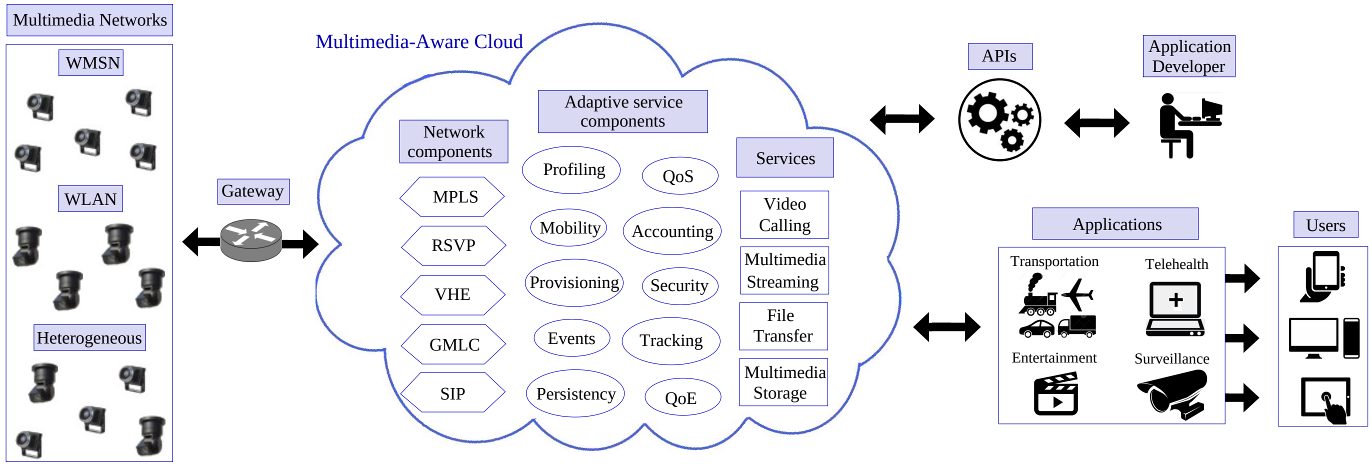

Fig. 6. Multimedia-Aware Cloud's Middleware Architecture

capability and service specific QoS requirements. Thus, middleware should adapt to support devices with heterogeneous resources (i.e. CPU, memory, power, etc) and enable capabilities like profiling, accounting, provisioning, etc for end-users as shown in Fig. 6. In order to facilitate the specific service requirements, the adaptive service components are translated to some implementation dependent network components i.e. VHE (virtual home environment), GMLC (gateway mobile location center), VHE (virtual home environment), SIP (Session Initiation Protocol), RSVP (Resource Reservation Protocol), MPLS (multiprotocol label switching), etc [88], [89]. The multi-agent systems provide large number of application and composite services; however an end-user may be interested in only a subset of these services. Thus, user personalization feature is an essential element of a middleware for fulfilling the end-user application requirements and improving the user experience. Middleware should also provide access of various applications and services to the mobile users connected via heterogeneous networks. In addition, efficient end-to-end translational protocols be devised to present a virtual abstraction of multimedia sensor devices, so that multi-agent systems may control these remote devices to develop services and applications based on their characteristics and states.

B. Object Naming Services

The multimedia devices need to be identifiable and addressable through the Internet by distinct IDs such that the IDs are unique, scalable, persistent, and reliable. In order to realize true IoMT architecture, the cloud should enable the ability to access and control large number of remote multimedia devices for end-users and multi-agent systems. Traditionally, IPv4 addressing has been used to identify group or network of sensor devices based on their geographical location, yet the individual identification (naming) and addressability of these devices was not realizable. Thus, naming scheme for multimedia devices is an elemental component in IoMT architecture. Multiple naming schemes have been proposed based on current Internet architecture i.e. IPv6, Domain Name System (DNS), uniform resource identifier (URI), Glowbal IP protocol, etc. Moreover, some schemes for future Internet design i.e. 6LoWPAN, GS1 Object Name Service (ONS), Sensor Web Enablement, MobilityFirst, etc, are proposed providing better solutions specifically related to security and mobility issues.

The location and identity of a device can be revealed by an IP address up to some extent. On the other hand, URI is more user friendly than IP address. Therein, a user can access a device globally using the DNS servers without the knowledge about its IP address. URI uses both universal resource locator (URL) as well as universal resource name (URN) for global accessibility and identification of a device or web resources' name. IPv6 supporting 10^{38} number of addresses [5] seems to meet the demand of the explosive growth rate of the Internet devices that is expected to reach 50 to 100 billion devices by 2020 [91]. To enable IPv6 for low-power PAN, 6LoWPAN is proposed by the IETF group which allows IPv6 based packet transmission over IEEE 802.15.4. The 6LoWPAN still comes with overhead problem and for this reason authors in [91] proposed a low overhead naming scheme Access Address/Identifier (AAID). AAID is used by Glowbal IP protocol to reduce the IPv6 protocol overhead by 22-35 bytes which increases data space for payload up to 40% to 60%. However, Glowbal IP protocol does not provide mobility support.

GS1 standardization organization has introduced a naming scheme, 'Identification Keys', which uses keys to globally access the status and location of devices by employing an Object Name Service (ONS). The ONS translates the keys into the format that can be understood by the DNS [92]. The ONS system is a feasible solution as it is globally available and can be integrated with a DNS system at a little cost. Sensor web enablement (SWE) standards developed a naming scheme, Sensor Model Language (SensorML), to discover, access and use the sensor devices via Internet [93]. It uses a special encoding scheme to access sensor metadata which contains information about several parameters i.e. name of the device. Moreover, another futuristic naming scheme referred as MobilityFirst is proposed in [94]. MobilityFirst totally replaces the traditional IP architecture to resolve issues posed by IoT. It uses Global unique identification (GUID) and Name Assignment Service (NAS) mechanisms to separate the device name from network address to provide mobility services for millions of devices. However, it is immature and requires more extensive techniques and mechanisms to succeed the current solutions for the Future Internet architecture.

C. Digital Content Naming

Along with the standardization efforts for naming of multimedia devices, another critical functionality of a multimedia-aware cloud is digital content naming in order to manage the huge amount of multimedia content. Digital content naming in general is addressed in various research studies since the last decade, and more specifically the naming of multimedia content has gained special attention such as [95], [96], among others. Digital object identifiers (DOI) are digital identifiers used to identify any digital content along with its metadata about that digital content which may include URL, referring to the location of that digital content. DOI is standardized by International DOI Foundation (IDF) in 2001. In past, similar standardization effort is carried out by ISO for unique identification of music videos and/or sound recording, by naming the multimedia content with an International Standard Recording Code (ISRC).

In case of multimedia content, MPEG-21 multimedia framework [97] provides a method for digital item identification (DII). The DII method uses existing identification schemes, such as DOI, ISRC, etc, for unique identification of multimedia content globally. MPEG-21 digital item comprises of an elemental piece of multimedia content which may range from a simple image or sound-track to a more complex high definition video content. It consists of multimedia content itself, known as a resource, the metadata about resource such as MPEG-7 which provides content management and description facilities, and a standard structure which provides relationship among various resources. The digital item declaration (DID) and digital item identification (DII) methods are considered as structured entities and these methods are developed considering their compatibility with existing digital data identification schemes. Therefore, they can be utilized for unique identification of multimedia content in an environment where heterogeneous devices and network are deployed. Moreover, to meet this challenge of seamless universal multimedia content (which may exist in multiple coding formats) across heterogeneous networks and/or devices, MPEG-21 framework also specifies the multimedia adaptation mechanism, referred as digital item adaptation (DIA) [98]. DIA provides description tools and formats for the adaptation of digital items in terms of QoS management, bitstream adaptation, resource adaptation, etc so as to meet the end user requirements and to enable inter-operability among various vendors.

D. Data storage

The IoMT paradigm mandates the need of scalable storage infrastructure development and deployment, whereby massive amount of data, its context and meta-data can be stored along with the analytical tools required by end-users for comprehension and cognition of this overall data. Thereby, a multimedia-aware cloud should support different type of QoS requirements as per user and various types of multimedia content. The storage centers should not only provide scalability and accessibility but also provide means of indexing and categorization of massive data as per the end-users' requirements. Moreover, a key challenge is the integration of this storage infrastructure with other cloud components such as computational infrastructure, middleware, multi-agent systems, etc [5]. This issue is further extended by the growing security and privacy issues with the ubiquitous increase in sensor devices, diversity of applied services/applications and user profiling.

The fundamental task for data storage in multimedia-aware cloud is data reliability and security. Data security has various elemental challenges such as data privacy, confidentiality and authenticity, to mitigate increasing security threats such as identity theft, identity defamation, security breach, propaganda, data forgetting and infringement, etc [99]. The pervasive nature of wireless technology has further multiplexed and diversified the security vulnerabilities. The notion of IoT and ubiquitous presence of multimedia devices (i.e. surveillance cameras or sensors) and cloud computing have strengthen the significance of data security. Consequently, in recent years numerous solutions have been proposed [100], [101] to mitigate these challenges. However, the diverse nature of cloud computing and wireless communication still needs to be investigated. The key challenges include scalability, energy consumption and standardization of current solutions together with their enforcement methods in heterogeneous environment, re-definition of privacy policies and access rights.

Authorization of stored multimedia data, its corresponding access and privacy policies mandates the need of object tracking along with redefinition of constraints to give data rights to specific set of users. Data protection methods may include various strategies such as randomization and suppression of data [102] to guarantee anonymity of information while storing it at cloud; and additionally to keep the secure identification of various records. A typical example include surveillance and monitoring cameras are capturing a traffic violation or a robbery event and immediately criminal records of the accused person is required to be cross-examined at cloud based servers. Since, the criminal record stored at multimedia aware-cloud is sensitive data, the situations demands data storage algorithms and security methodologies to work concurrently for the efficient and reliable retrieval of information out of data streams. Similarly, designation of trust authority has become increasingly challenges with the growing trends of keeping the sensitive data in cloud based servers. Device-to-device communication and device-to-human interaction entails the need to fulfil interoperability requirements along with the standardization of these procedures. In this regard, the inter-device trust relationship establishment and its exchange among humans or devices is another pre-requisite, especially when these devices are assumed to take decisions themselves. Consequently, a new security framework is therefore essential that will be able to leverage all the aforementioned security issues and challenges.

E. Data processing

With the advancement of Web 2.0 and 3.0, the multimedia rich content, i.e. audio and/or video or images, is increased radically and poses new challenges to provide capability to search, edit and process multimedia content. To carry out such computational intensive applications and services to millions of end-users simultaneously, massive amount of storage as well as computational resources are required. In most of the scenarios in IoMT, the multimedia sensor devices are expected to be low powered, with limited memory and processing capability. Thus, in order to mitigate the overhead of installation and maintenance of multimedia softwares, the multimedia-aware cloud should enable users to carry out storage and processing tasks in a distributed manner and eliminate the requirement of complete software installation on end-user devices [81].

In traditional multimedia networks, the multimedia raw content such as video is compressed once while recording and decompressed every time it is played back, mandating complex encoders and simpler decoders. However, we have to alleviate both the processing overhead at sensing devices and decoding overhead at low power mobile users connected via wireless gadgets, by shifting it towards the multimedia-aware cloud. This can be achieved by simplifying the encoders by making them computationally less complex. Consequently, the complexity will be shifted towards the decoder present at the cloud which has powerful computation resources. Further, to facilitate mobile and low power end-user devices, the multimedia content which is intended to be delivered should be re-encoded at cloud in such a way that the decoder at end-user devices can operate with limited utilization of resources (i.e. complex encoder and simple decode technique). In prior research work, server-based computing such as content delivery networks (CDN) are proposed which address the issues like latency and bandwidth constraints to provide QoS to the end-users [103]. Some popular CDN include; Youtube Akamai's CDN, Limelight Networks, among others. Similarly, a peer to peer (P2P) multimedia computing architecture enables workload distribution among peers in a distributed manner.

The proposed multimedia-aware cloud architecture aspires to capitalize on traditional CDN infrastructure, whereby servers are deployed at the edge of cloud, not only for multimedia content delivery but also for processing of data in order to reduce latency and delay along with the bandwidth maximization. High performance computing is achieved by accomplishing processing of multimedia content of similar types and related to particular user in a distributed parallel manner. Thus, to reduce computational cost and workload, various tasks of similar types are carried out at single edge-server while tasks of different users can be allocated to several servers in a parallel fashion. Therefore, the multimedia-aware cloud truly envisions both distributed and parallel computing methodologies to realize IoMT paradigm.

F. Open issues

The services provided by existing cloud servers are either limited to simple device management system and sensor data logging, or massive data storage and processing services. However, the cloud services required for IoMT are not yet addressed in the prior studies. A short summary of the major requirements expected from a IoMT supportive cloud are as follows:

- Enable service provisioning and render distinct multimedia services and applications.
- Provide the abstraction of heterogenous multimedia devices to the multi-agent system in order to compose distinct multimedia applications/services.
- Provide context information based management of network resources, e.g. efficient bandwidth utilization and computational capacity.
- Provide user personalization feature to fulfill end-user application requirements and improved user experience.
- Object naming services to uniquely identify and address massive number of devices. This service should be scalable, persistent, and reliable.
- Owing to the dynamic change in access and authorization roles at run-time, which are in turn being employed on variable and dynamic multimedia data streams, the existing customary solutions lack the rigorous requirements posed by multimedia data retrieval in real-time and in a flexible manner.

VI. Multi-Agent Systems

In previous section, we discussed the resources and functionalities of a multimedia-aware cloud to facilitate the QoS requirements of multimedia based services and applications. Subsequently, these resources should be utilized in optimal manner to achieve the objectives of the distributed, multimedia IoMT applications at the desired overall QoE while composing and integrating heterogeneous services. This foster the need of ‘cloud-aware multimedia systems’ that can dynamically perform functionalities such as service composition, data mining and analytics, content sharing and delivery, rendering and retrieval, etc. To optimally utilize, combine, map, schedule, coordinate various resources provided by the cloud, researchers have recently proposed agent-based cloud computing [104], [105].

Agents are autonomous computer programs/software working independently to accomplish a specific objective while fulfilling the definite design constraints and requirements. For example, two or more agents may cooperate to achieve a common objective of composing a service while satisfying the constraint of utility maximization [106]. A multi-agent system consists of number of independent, cooperating, and communicating agents that respond to some event(s) according to the user demands. In a multi-agent system, the agents organize, optimize, and configure themselves with the help of interacting, negotiating, and collaborating with each other. Multi-agents are suitable software components for independently handling the cloud functionalities. For example, the multi-agent system makes decisions according to the appropriate set of demands by users, depending upon the available resources and current service obligations [107]. In addition, multiple offered services can be combined to form a unified virtual service for the users. Similarly, multi-agent systems can also compose services upon request, therein user have to provide the detail about the functionality of the service and the user device characteristics. Therefore, it is asserted that the multimedia-aware multi-agent system can efficiently anticipate the cloud-aware multimedia perspective. It is noteworthy that although the multi-agent systems are discussed separately, yet they are an integral part of the multimedia-aware cloud. In the specific context of IoMT, the multi-agent systems should be deployed to implement the following functionalities: service composition, data mining and analytics, content sharing and delivery. Table VII shows the major requirements that result from the use-cases described in Section 3, which then become objectives of the three identified functionalities.

A. *Service composition*

In recent research studies, the multi-agent systems are used for dynamic and automated service composition at the cloud. Since, multi-agents can conveniently manage cloud resources, i.e. storage and computation resources, thus service composition tasks, for example integration of various homogeneous and heterogeneous services, can be easily regulated and taken care of. The critical issue for multimedia based services is to compose and customize services according to the user preferences, device capabilities, QoS, and QoE requirements of the consumers. This involves both the division of complex services and aggregation of simpler ones into a composite service. In addition, multimedia service composition in dynamic environments demands context awareness at multiple levels including user-context, device-context, cloud resource-context, network-context and service-context. Consequently, lot of services are built on the multimedia data coming from heterogeneous multimedia networks and these services are provided to the users with distinct devices connected via heterogeneous networks according to the dissimilar QoS and QoE requirements. While considering heterogeneity at various levels, multiple services can be built on the basis of the same or different multimedia content [108].

Multimedia information as well as the scaler sensed data can be utilized to compose customized services and applications. This allows the reusability of the functionalities provided by the existing sensor infrastructures. The process of selection, filtration, allocation, and composition of the services is dependent upon the capacity of user device and network characteristics.

TABLE VII
MULTI-AGENT SYSTEMS REQUIREMENTS FOR USE-CASES CONSIDERED IN SECTION VII

Use-cases	**Service composition**	**Multimedia data mining & analytics**	**Content sharing and delivery**
Ecosystem of everyday things	Combination of more flows, Transcoding flows from different sources	Analysis of traffic conditions	N/A
Smart business management and marketing	Mapping of customer traffic flows, Scheduling alerts based on multimedia flows	Identification of people actions	Delivery of alerts/ multimedia data to interested customers
Behavioral interpretation systems	Flow coordination for object tracking, Transcoding flows from different sources	Understanding people actions, Identification of objects	Streaming of video content to monitoring centers
Telemedicine	Coordination of multimedia flows, Interaction of flows with sensor data	Understanding people actions	Sharing of medical record (x-ray images, ultrasound, etc) or live monitoring
Smart societal interaction	Integration of flows with social websites, Composition of various web services at cloud	Understanding people interaction, Analysis and integration of social web data	Sharing of multimedia content of user interest
Automated public security	Combination and coordination of flows, Transcoding flows from different sources	Identification of dangerous situations	Streaming of video content to monitoring centers, hospitals, etc

Multi-agents carry out these tasks while considering the available resources of specific user and also schedule, allocate, and manage the cloud resources efficiently to support millions of users accessing the cloud at the same time. The concepts of self-service provisioning, on demand rental services, quantitative and qualitative services are considered to be essential features of cloud which provide a consolidated platform for multimedia content based applications.

The multimedia content in IoMT application is used to present a geographical area that can be both outdoor and indoor environments. Depending upon the application requirements the sensors, including both multimedia and non-multimedia, can be placed in the vicinity of the area of interest in order to acquire visual and scaler information. As shown in Table VII, the information sensed from the environment can undergo different procedures like it can be presented to the humans/users; different multimedia flows may also be combined to present the best view of an area; the multimedia data arriving at the cloud can be in different formats and thus it should be transcoded appropriately; the multimedia data coming from heterogeneous devices may also be merged and application specific analytics can be applied; the information form scaler sensor nodes can be used to trigger the operation of multimedia devices. These are some of the examples where different multimedia services can be composed by the managing the scaler as well as the multimedia data in the IoT. For example, a gun shot sensor device may trigger high quality video recording in a vicinity and based on the notification alerts generated by this sensor various analytical processes can be initiated on the acquired visual information from the multimedia sensors. Similarly, the humans can be notified about the events of interest by simply forwarding the sensing data (video stream) or they may be inquired about any response strategy requiring an approval (reporting the event to competitive authority). Thus, the multi-agents need to compose and devise an efficient, scalable, and flexible service orientation framework which can encounter all these aspects.

B. Multimedia data mining and analytics

Multimedia data mining relates to the procedures that are used to extract concealed and beneficial information patterns from a large group of multimedia data, to give it an understandable context based representation. Due to the unstructured nature of multimedia content, the useful semantic information extraction and knowledge discovery from large multimedia databases is difficult. Considering different multimedia data types, i.e. image, text, audio, video, graphics, etc, the analysis and transformation of multimedia data, in order to compare and extract meaningful links, is a challenging task [109]. Moreover, the spatial and temporal diversity in various multimedia data types mandates the need of coordination among multiple independent entities (agents) to employ sophisticated artificial intelligence based techniques and algorithms. Therein, the multi-agents have to interact and apply different strategies to satisfy certain application specific objective functions to make sense of the collected information and learned knowledge by multimedia data mining.

Once the desired multimedia content is collected, the cloud based multimedia data analytics processes are applied on it by using various centralized and distributed algorithms. It carries multiple aspects such as multimedia content characteristics, processes applied on the multimedia content (i.e. image processing, facial recognition, object detection), information about multimedia sensor devices, analytical models and multimedia content sharing, etc. The multimedia content coming from devices from different networks has to be aggregated and condensed in a beneficial manner to efficiently utilize the power and storage resources. In order to facilitate the end-users, multi-agents use data analytics tools and algorithms to analyze the acquired/stored multimedia content, to make sense of this data and to take intelligent decisions as per the services and applications requirements. Number of applications in IoMT paradigm demand distinct QoS performance, thus the flexible and dynamic data analytics tools and platforms are foreseen in near future.

Analysing the use-cases presented in Section VII, we can highlight different tasks that need effective multimedia data analytics tools. These are mostly related to the identification and understanding of people actions. These tasks should be performed both online (for a proper reaction) and offline (to prepare the systems for a more rapid online reaction).

C. Content sharing and delivery

Multimedia content sharing and its delivery is one of the salient features of the cloud. The traditional data sharing mechanism involves both the clients to be in an active state during the process. In addition, the multimedia content imposes strict bandwidth and QoS restrictions. However, the cloud enables one-to-many data sharing capability by allowing the clients to upload data to the cloud while the receiving clients can simply access it anytime and anywhere, as the cloud is active all the time. Due to the availability of storage clusters deployed in different geographical locations, the cloud has the potential to provide better bandwidth utilization, QoS and QoE. Moreover, to maximize the cloud resource utilization, the agents in the cloud can cooperate and negotiate with each other in order to maintain load-balancing and multimedia service operations among multiple servers. To manage huge multimedia content coming from large number of multimedia acquisition devices, two important functions at the cloud are data authoring and data mashup. As the name suggests, data authoring involves modification and editing of different multimedia objects (segments) of the acquired multimedia content. Whereas, the mashup procedure combines the acquired multimedia content in the form of objects from various sources into a unitary entity. These functions operate according to the resource management obligations or in order to facilitate specific services and applications.

The multimedia data may be destined either to data analytics tools (as described in the previous section) or to various kinds of user devices, such as smart phones, TVs, and laptops, that may be connected to the cloud via heterogeneous networks with distinct communication conditions in terms of bandwidth, jitter, delay, etc. In the later case, multi-agent systems should also employ multimedia content adaptation [110] techniques while delivering data as per user obligations. For multimedia adaptation, the offline transcoding at the sensor devices is not feasible in real-time multimedia traffic, e.g. telemedicine and public security. Therefore, multimedia adaptation is proposed to be carried out at the resource rich cloud, an example is the Microsoft Azure platform's 'cloudcoder'. However, the multimedia adaptation comes with increased storage requirements and adaptability issues for dynamic network conditions to improve QoE. For example, in a single-layer and multi-layer video content, the adaptation operation is significantly distinct based on user device and network characteristics. A single-layer video may undergo changes in its frame rate, bit rate, resolution. Whereas, in multi-layer video the adaptation process may truncate the high definition layers to facilitate better QoE to the user depending upon its network conditions. This process is also known as scalable video coding, in which video quality is varied as per the user device and network scenarios.

D. Open issues

- Previously, multi-agent systems have only been considered for IoT systems. However, multimedia based services require critical composition and customization as per the user preferences, device capabilities, QoS, and QoE.
- Formation of context aware multimedia service composition methods is also of great importance. Thereby, the context awareness at multiple levels including user-context, device-context, cloud resource-context, network-context and service-context, can be categorized.
- Another open issue for multi-agent systems is to design self-service provisioning, on demand rental services, quantitative and qualitative services multimedia content based applications.
- Multi-agents should be able to compare and extract meaningful information from different types of multimedia content and apply different strategies to satisfy certain application specific objective functions.
- Multimedia adaptation mechanisms need to be devised which can adapt to the distinct user device and network characteristics.

VII. IoMT Applications and Use-Cases

Heterogeneity of multimedia sensors, ubiquitous connectivity, QoS and QoE requirements, cloud based analysis and autonomous reaction approaches constitute the key characteristics of IoMT. Leveraging cloud based resources to process video analytics, cameras can be the most intelligent IoT sensors. IoMT-enabled state of the art analytics and video surveillance systems can be transformed from a cost center for businesses to actually driving operational savings and delivering real return on investments besides providing superior security control. Deployment of cameras in an IoT setup would fulfil its full potential when accompanied with the ability to automatically glean actionable information from the visual streams, which could then be exchanged with other things when appropriate, eventually leading to some action taken by the things autonomously with rare human involvement. The bottom line is keeping human involvement in the entire process down to a viable minimum. Since, the multimedia content especially video data is extremely rich in information. Thus, advanced video analytics enables extraction of various kinds of information from a single video stream which would otherwise require a plurality of sensors. Object and motion detection, facial recognition, number plate identification, situational and acoustic awareness, video search and forensic analysis, counting applications, intrusion and virtual line/gate crossing detection, theft detection are some of the features that can be extracted from video streams. A few possible use-cases of IoMT are as follows.

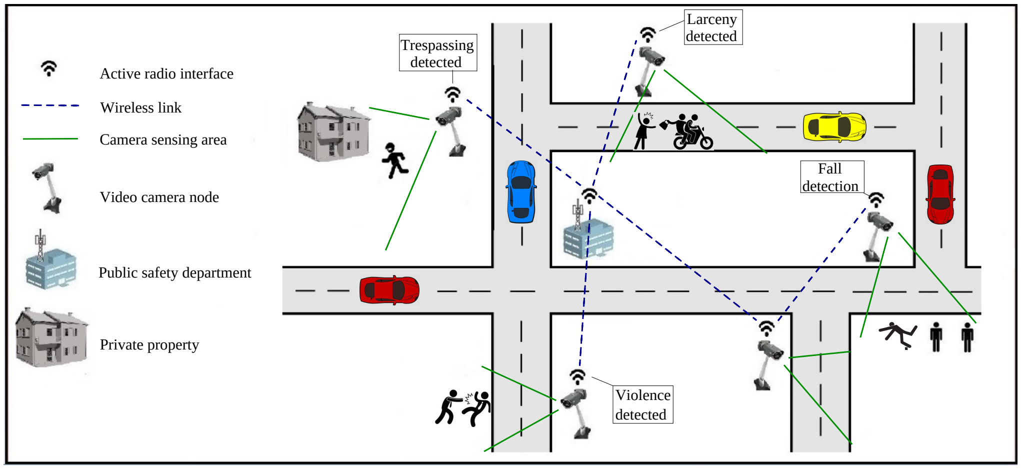

Fig. 7. Behavioral interpretation based IoMT application scenarios

A. *Ecosystem of everyday things*

A scenario of IoT with 'multimedia things' in everyday routine can be along these lines. You were scheduled to attend a meeting early morning. The meeting got delayed by 20 minutes. The scheduler (a simple computing device with Internet access) at the office informs your scheduler at home about the delay. Your scheduler also communicates with the car's navigation system, which in turn sends the route information to the multimedia-aware cloud for traffic information. With the help of multiple video sensors along the route, the multi-agent system analyzes the traffic and route statistics. It is informed that it has been raining all night and one of the main roads along the route is inundated causing a traffic jam. The navigation system recalculates the traveling time along a different route and informs the scheduler that you should leave 15 minutes earlier than usual. The scheduler thus delays the alarm by 5 minutes and signals the coffee maker to turn on 5 minutes later as well. It also signals the car to start in 10 minutes in order to warm up the engine beforehand in the cold or damp weather.

B. *Behavioral interpretation systems*

Recently, behavioural analytics based on multimedia content has emerged as an advantageous technique. Unusual behaviour in the video streams can be automatically determined and autonomous reactive procedures, i.e. warnings are issued. For example, a truck of disallowed dimensions or a vehicle traveling in the wrong direction can be identified and alerts can be generated. Behavioural analytics has many applications, from industrial manufacturing where a tank containing toxic substance when starts vibrating unusually can be detected or smoke/fire can be visualized especially in open areas such as parking lots, where smoke detectors are ineffective. Video information can be utilized for traffic management as well. It can help in evaluating traffic patterns, making prediction and hence better congestion control. People dropping off suspicious items in crowded places or trying to abandon their cars near security checkpoints instead of un-boarding normally can be distinguished. Videos can be integrated with other sensors for monitoring and tracking erosion of sand banks over seasons. IoMT can tremendously enhance the capabilities of the traditional surveillance and security systems. Object detection, facial recognition and theft detection or object displacement detection are some of the features useful for superior security. A few application scenarios are depicted in Fig. 7. Traditionally, the events are identified by streaming video content to a control center where a person is required to monitor any unusual behaviour. Conversely, in smart multimedia networks the behavioural analytics can be done at smart camera nodes to interpret the event and appropriate reactive approaches can be executed such as reporting to a public safety department, triggering alarms to notify the responsible personnel, calling for medical or other types of assistance, etc.

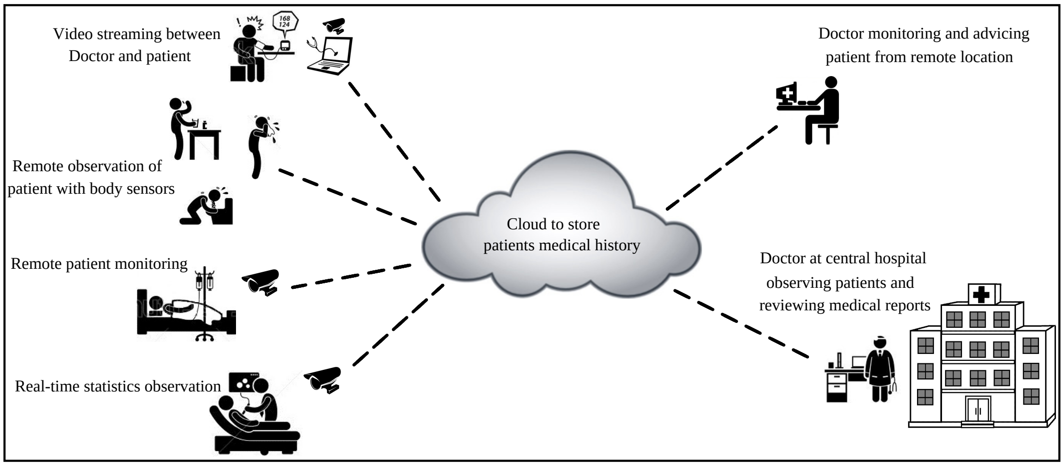

Fig. 8. IoMT application scenarios in Telemedicine

C. Telemedicine

IoMT paradigm can enhance the applicability of telemedicine and bring valuable advancement in the tele-healthcare technologies especially for rural areas. The technological solutions envisioned within IoMT can be specifically more beneficial for under-developed countries. As a case study, real-time monitoring of patients living in a remote town is realizable using smart multimedia devices. Doctors can be provided with an early and easy access to remote patient's medical history stored at the multimedia-aware cloud. On the other hand, patients can get the access to specialized doctors of particular treatments in a rapid and cost effective manner. Sensing devices can be employed to detect pollution level or a contagious disease by retrieving information from a remote environment or habitat. Sensors technologies which have already enabled the remote monitoring of patients can be complimented by multimedia camera nodes to deliver a real-time map of patients' everyday routine and proactive feedback and reminders can be sent. In case of any indescribable health problem, a disable or elderly patients can be kept under eye and accordingly be advised to take necessary precautionary measures. Similarly, in another scenario, the patients under observation in a hospital can be examined by doctors from a remote location via multimedia devices. Data collected from these patients can enable doctors to identify any indication of upcoming severe event so that necessary action may be taken pre-determinedly. It also facilitates the process of decision making such as a patient with higher risk should be given more priority while scheduling a treatment-list by the doctors. A few application scenarios are depicted in Fig. 8.

D. Smart Societal Interaction

The ideas of smart communities and social interactions, a hot topic in current research, can be given a new dimension in the presence of IoMT multimedia devices. The societal and communal impact of the multimedia devices is huge and the increase in offered functionalities enabled by multimedia sensor devices is reshaping people-to-people interaction, people-to-device interaction and inter-devices interaction. A possible scenario of IoMT in the context of social community is interaction of multimedia devices that capture two persons who are traveling in a close near-by location and may be interested in having a lunch at some fast food place. Smart multimedia devices, while interacting with social media and networking websites, intelligently learn that these two people are friends and thus inform them on their smart phones that they are in near-by location and potentially hangout with each other. Furthermore, multimedia multi-agents can collect and integrate social media information about these two persons and suggest them about some social event, near-by shopping mall, a cinema theater or any other activities of their interest.

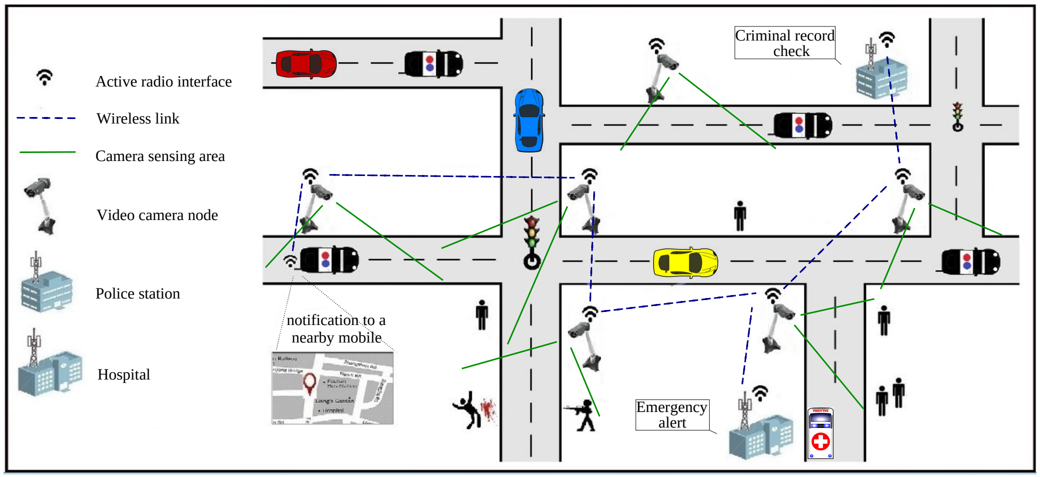

Fig. 9. Use-case: Automated public security

E. Smart business management and marketing

In business management and marketing, the decision making can be made smarter and timely with the help of modern video analysis techniques. When events with operational implications occur, the businesses can be alerted. These alerts can be delivered on smart phones of the relevant personnel or to the operation center making timely reactions possible. For instance, consider a retailer's place where customers are entertained based on their profile to offer intelligent services. Customers can be identified by their cars‘ registration number with the help of smart cameras. When a loyal customer with some disability arrives in the parking lot, the nearest staff member can be located and an alert can be dispatched to her smartphone so that she can welcome and assist the customer. Similarly, some other relevant implications are as follows: (i) Service personnel can be instantly alerted when queues are formed at the checkout, (ii) Customer statistics can be gathered and their behaviours can be gauged, (iii) The number of customers can be counted, average amounts of time spent looking at a certain product can be understood and traffic flow maps within a store can be generated. These patterns can be used to determine the most sought out products and efficacy of product displays. Over time this information can be collected for more informed resource allocation, product displays and promotions, (iv) Crowd fluctuations on daily, weekly and monthly basis can be monitored and staff duties can accordingly be assigned. Discrepancies can therefore be automatically identified and promptly reported.

F. Automated public security

In the IoMT world video streams can enable the public safety and one of the possible scenario is depicted in Fig. 9. A shot has been fired by a person in a crowded street. Audio analysis of the sound by the cameras in the vicinity of the shooter would enable the detection of gunshots. To provide rapid medical assistance to the victim, the smart multimedia devices deployed in the vicinity can detect the happening of a unusual event. The incident information is reported to a nearby hospital which in turn dispatches the ambulance right away. The doctors are facilitated to visualize the real-time video streaming from the incident, hence improving the medical response time. Simultaneously, the shooter would be identified amongst the crowd and his motion would be predicted. Based on the direction of motion of the shooter, camera controller will inform the next cameras in the path of the shooter to adjust their resolution and orientation. The controller will contact the nearest law enforcement patrol and continuously update the location of the shooter on the patrol's navigation. Meanwhile, the controller will transmit the facial information to identity database, police and medical databases so that his identification, past criminal record and medical history can be checked. The navigation on the patrol will communicate with the traffic signals to control the flow of traffic to avoid any interruptions in their movement.

VIII. Conclusion

Internet of Multimedia Things represents a specialized subset of Internet of Things, enabling the integration and cooperation among heterogeneous multimedia devices with distinct sensing, computational, and communication capabilities and resources. A multimedia-aware cloud combined with the multi-agent systems can help design prodigious multimedia services and applications. As compared to IoT, the realization of IoMT rather has some additional challenges and stringent requirements. The current task group activities for IoT do not mandate the features of multimedia things, thus leaving a gap to benefit from IoMT paradigm. This survey presents the vision of IoMT, highlight its unique characteristics in contrast to IoT, and identify the major challenges and requirements in order to realize the IoMT. Considering these technical specifications of IoMT, the feasibility of the already proposed communication stack for IoT is discussed. In addition, the multimedia content processing at resource constraint multimedia devices in IoMT is analyzed. The solutions to the processing/computational issues are provided by introducing the notion of multimedia-aware cloud combined with multi-agent systems for IoMT architecture. Lastly, some applications and use-cases of IoMT are discussed to further clarify and help make understand the research community about the true potential of IoMT. This is the first paper to introduce IoMT paradigm and architecture to identify its issues and obligations. We believe that substantial standardization efforts are needed to address the open issues in order to realize true potential of IoMT.

Acknowledgement

This work is supported by the National ICT R&D Fund, Government of Pakistan under Grant no. ICTRDF/TR&D/2013/04.

References

[1] Charu C Aggarwal, Naveen Ashish, and Amit Sheth. The internet of things: A survey from the data-centric perspective. In *Managing and mining sensor data*, pages 383–428. Springer, 2013.

[2] Lu Tan and Neng Wang. Future internet: The internet of things. In *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on*, volume 5, pages V5–376. IEEE, 2010.

[3] Hua-Dong Ma. Internet of things: Objectives and scientific challenges. *Journal of Computer science and Technology*, 26(6):919–924, 2011.

[4] P. Friess S. Woelffl H. Sundmaeker, P. Guillemin. *Internet of Things, CERP-IoT - Cluster of European Research Projects on the Internet of Things*. 2010.

[5] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer networks*, 54(15):2787–2805, 2010.

[6] Michele Zorzi, Alexander Gluhak, Sebastian Lange, and Alessandro Bassi. From today's intranet of things to a future internet of things: a wireless-and mobility-related view. *Wireless Communications, IEEE*, 17(6):44–51, 2010.

[7] *European Research Cluster on Internet of Things (IERC)*. Internet of Things - Pan European Research and Innovation Vision, IERC, Available online: http://www.internet-of-things-research.eu/documents.html, 2011.

[8] Jean-Philippe Vasseur and Adam Dunkels. *Interconnecting smart objects with ip: The next internet*. Morgan Kaufmann, 2010.

[9] Olivier Hersent, David Boswarthick, and Omar Elloumi. *The Internet of Things: Key Applications and Protocols*. John Wiley & Sons, 2011.

[10] Sandivine. Sandvine global internet phenomena complete, 2014.

[11] Cisco VNI. The zettabyte era: Trends and analysis, june 2014.

[12] Qualcomm. www.eeworld.com.cn/zt/wireless/downloads/qca4002-4004fin.pdf.

[13] MICROCHIP. http://ww1.microchip.com/downloads/en/devicedoc/70005171a.pdf.

[14] Thingspeak online: https://thingspeak.com/.

[15] xively online: https://xively.com/.

[16] iobridge online: https://www.iobridge.com/.

[17] Carriots online: https://www.carriots.com/.

[18] Axeda online: http://www.axeda.com/.

[19] Ian F Akyildiz, Tommaso Melodia, and Kaushik R Chowdhury. A survey on wireless multimedia sensor networks. *Computer networks*, 51(4):921–960, 2007.

[20] Ian F Akyildiz, Tommaso Melodia, and Kaushik R Chowdhury. Wireless multimedia sensor networks: Applications and testbeds. *Proceedings of the IEEE*, 96(10):1588–1605, 2008.

[21] Islam T Almalkawi, Manel Guerrero Zapata, Jamal N Al-Karaki, and Julian Morillo-Pozo. Wireless multimedia sensor networks: current trends and future directions. *Sensors*, 10(7):6662–6717, 2010.

[22] Sheldon Leader. Telecommunications handbook for transportation professionals: The basics of telecommunications. Technical report, 2004.

[23] John Hourdakis, Ted Morris, Panos Michalopoulos, and Prasoon Sinha. Advanced portable wireless measurement and observation station. 2005.

[24] Ning Luo. *A Wireless Traffic Surveillance System Using Video Analytics*. PhD thesis, University of North Texas, 2011.

[25] Nan Li, Bo Yan, Guanling Chen, Prabhu Govindaswamy, and Jie Wang. Design and implementation of a sensor-based wireless camera system for continuous monitoring in assistive environments. *Personal and Ubiquitous Computing*, 14(6):499–510, 2010.

[26] Atsushi Kawamura, Yuji Yoshimitsu, Kouichiro Kajitani, Takeshi Naito, Kaichi Fujimura, and Shunsuke Kamijo. Smart camera network system for use in railway stations. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 85–90. IEEE, 2011.

[27] Internet-of-things architecture (iot-a) project deliverable d1.2 initial architectural reference model for iot.

[28] Aura Ganz, Zvi Ganz, and Kitti Wongthavarawat. *Multimedia Wireless Networks: Technologies, Standards and QoS*. Pearson Education, 2003.

[29] Mihaela van der Schaar and Philip A Chou. *Multimedia over IP and wireless networks: compression, networking, and systems*. Academic Press, 2011.

[30] Adam Dunkels and Jean-Philippe Vasseur. Ip for smart objects, september 2008. *IPSO Alliance White Paper*, 1.

[31] Ian F Akyildiz and Ismail H Kasimoglu. Wireless sensor and actor networks: research challenges. *Ad hoc networks*, 2(4):351–367, 2004.

[32] Tommaso Melodia, Dario Pompili, Vehbi C Gungor, and Ian F Akyildiz. Communication and coordination in wireless sensor and actor networks. *Mobile Computing, IEEE Transactions on*, 6(10):1116–1129, 2007.

[33] Vladimir Stankovic, Lina Stankovic, and Samuel Cheng. Compressive image sampling with side information. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3037–3040. IEEE, 2009.

[34] Frederic Dufaux, Wen Gao, Stefano Tubaro, and Anthony Vetro. Distributed video coding: trends and perspectives. *Journal on Image and Video Processing*, 2009:10, 2009.

[35] Geoff V Merrett, Neil M White, Nick R Harris, and Bashir M Al-Hashimi. Energy-aware simulation for wireless sensor networks. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON'09. 6th Annual IEEE Communications Society Conference on*, pages 1–8. IEEE, 2009.

[36] Gang Lu, Bhaskar Krishnamachari, and Cauligi S Raghavendra. Performance evaluation of the ieee 802.15. 4 mac for low-rate low-power wireless networks. In *Performance, Computing, and Communications, 2004 IEEE International Conference on*, pages 701–706. IEEE, 2004.

[37] J Ohm and Gary J Sullivan. High efficiency video coding: The next frontier in video compression [standards in a nutshell]. *Signal Processing Magazine, IEEE*, 30(1):152–158, 2013.

[38] Dan Grois, Detlev Marpe, Amit Mulayoff, Benaya Itzhaky, and Ofer Hadar. Performance comparison of h. 265/mpeg-hevc, vp9, and h. 264/mpeg-avc encoders. In *PCS*, pages 394–397, 2013.

[39] Maxim P Sharabayko, Oleg G Ponomarev, and Roman I Chernyak. Intra compression efficiency in vp9 and hevc. *Applied Mathematical Sciences*, 7(137):6803–6824, 2013.

[40] Sharabayko. Next genarattion video codecs hevc, vp9, and daala. *Applied Mathematical Sciences*, 7(137):6803–6824, 2013.

[41] Projects from VideoLAN, x264 software library and applications, http://www.videolan.org/developers/x264.html, online; accessed 21-july-2014.

[42] HM reference Software, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/, online; accessed 21-july-2014.

[43] Paul Wilkins. Google Groups 'WebM Discussion', https://groups.google.com/a/webmproject.org/forum/?fromgroups#!forum/webm-discuss, online; accessed 21-july-2014.

[44] David Slepian and Jack K Wolf. Noiseless coding of correlated information sources. *Information Theory, IEEE Transactions on*, 19(4):471–480, 1973.

[45] Aaron D Wyner and Jacob Ziv. The rate-distortion function for source coding with side information at the decoder. *Information Theory, IEEE Transactions on*, 22(1):1–10, 1976.

[46] Rohit Puri, Abhik Majumdar, and Kannan Ramchandran. Prism: A video coding paradigm with motion estimation at the decoder. *Image Processing, IEEE Transactions on*, 16(10):2436–2448, 2007.

[47] Anne Aaron, Rui Zhang, and Bernd Girod. Wyner-ziv coding of motion video. In *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, volume 1, pages 240–244. IEEE, 2002.

[48] Bernd Girod, Anne Margot Aaron, Shantanu Rane, and David Rebollo-Monedero. Distributed video coding. *Proceedings of the IEEE*, 93(1):71–83, 2005.

[49] Xavi Artigas, Joäo Ascenso, Marco Dalai, Sven Klomp, Denis Kubasov, and Mourad Ouaret. The discover codec: architecture, techniques and evaluation. In *Picture Coding Symposium*, volume 17, pages 1103–1120. Lisbon, Portugal, 2007.

[50] DISCOVER. Discover dvc final results, http://www.img.lx.it.pt/ discover/home.html.

[51] Emmanuel J Candès. Compressive sampling. In *Proceedings oh the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, pages 1433–1452, 2006.

[52] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[53] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.

[54] Thong T Do, Yi Chen, Dzung T Nguyen, Nam Nguyen, Lu Gan, and Trac D Tran. Distributed compressed video sensing. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1393–1396. IEEE, 2009.

[55] Scott Pudlewski, Arvind Prasanna, and Tommaso Melodia. Compressed-sensing-enabled video streaming for wireless multimedia sensor networks. *Mobile Computing, IEEE Transactions on*, 11(6):1060–1072, 2012.

[56] Ying Liu, Ming Li, and Dimitrios A Pados. Motion-aware decoding of compressed-sensed video. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(3):438–444, 2013.

[57] Deanna Needell and Rachel Ward. Near-optimal compressed sensing guarantees for total variation minimization. 2013.

[58] Ying Liu and Dimitris A. Pados. Decoding of framewise compressed-sensed video via interframe total variation minimization. *Journal of Electronic Imaging*, 22(2, Id: 021012-02101).

[59] Mahdi S Hosseini and Konstantinos N Plataniotis. High-accuracy total variation for compressed video sensing. *arXiv preprint arXiv:1309.0270*, 2013.

[60] G. Montenegro N. Kushalnagar and C. Schumacher. *IPv6 over Low- Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals, Internet Engineering Task Force RFC 4919*. August 2007.

[61] IEEE. *802.15.4e-2012: IEEE Standard for Local and Metropolitan Area Networks - Part 15.4: Low-Rate Wireless Personal Area Networks (LRWPANs) Amendment 1: MAC Sublayer*. April 2012.

[62] *ZigBee Alliance, Available online: www.zigbee.org*.

[63] *D. Networks, Available online: www.dustnetworks.com*.

[64] K Pister and Lance Doherty. Tsmp: Time synchronized mesh protocol. *IASTED Distributed Sensor Networks*, pages 391–398, 2008.

[65] A. Brandt J. Hui R. Kelsey P. Levis K. Pister R. Struik J. P. Vasseur T. Winter, P. Thubert and R. Alexander. *RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks, RFC 6550, Internet Engineering Task Force RFC 6550*. March 2012.

[66] C. Bormann Z. Shelby, K. Hartke and B. Frank. *Constrained Application Protocol (CoAP), IETF CoRE Working Group*. February 2011.

[67] J. Hui and P. Thubert. *Compression Format for IPv6 Datagrams over IEEE 802.15.4- Based Networks, RFC 6282, Internet Engineering Task Force RFC 6282*. September 2011.

[68] Mikael Degermark. *IP header compression, RFC 2507*. 1999.

[69] K. Pister N. Dejean J. P. Vasseur, M. Kim and D. Barthe. *Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks, RFC 6552, Internet Engineering Task Force RFC 6552*. March 2012.

[70] *Constrained RESTful Environments (core). IETF Working Group, Available online: http://www.ietf.org/dyn/wg/charter/core-charter.html*.

[71] Serbulent Tozlu. Feasibility of wi-fi enabled sensors for internet of things. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, pages 291–296. IEEE, 2011.

[72] *IEEE 802.11n-2009 Standard for Information technology, Local and metropolitan area networks, Specific requirements, Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput*. IEEE.

[73] Roy Friedman, Alex Kogan, and Yevgeny Krivolapov. On power and throughput tradeoffs of wifi and bluetooth in smartphones. *Mobile Computing, IEEE Transactions on*, 12(7):1363–1376, 2013.

[74] George Adam, Vaggelis Kapoulas, Christos Bouras, Georgios Kioumourtzis, Apostolos Gkamas, and Nikos Tavoularis. Performance evaluation of routing protocols for multimedia transmission over mobile ad hoc networks. In *Wireless and Mobile Networking Conference (WMNC), 2011 4th Joint IFIP*, pages 1–6. IEEE, 2011.

[75] Tsung-Han Lee, Xiang-Shen Xie, and Lin-Huang Chang. Rssi-based ipv6 routing metrics for rpl in low-power and lossy networks. In *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, pages 1714–1719. IEEE, 2014.

[76] Bogdan Pavkovic, Andrzej Duda, Won-Joo Hwang, and Fabrice Theoleyre. Efficient topology construction for rpl over ieee 802.15. 4 in wireless sensor networks. *Ad Hoc Networks*, 15:25–38, 2014.

[77] S. Pink L-E. Jonsson Ed L-A. Larzon, M. Degermark. *The Lightweight User Datagram Protocol (UDP-Lite), Available: http://tools.ietf.org/html/rfc3828*.

[78] S. Floyd E. Kohler, M. Handley. *Datagram Congestion Control Protocol (DCCP), Available: http://www.ietf.org/rfc/rfc4340.txt*.

[79] Casner S. Frederick R. Jacobson V. Schulzrinne, H. *RTP: A Transport Protocol for Real-Time Applications. InternetDraft, IETF, draft-ietfavt- rtp-new-09.txt, March 2001.*

[80] Iacopo Carreras, Imrich Chlamtac, Francesco De Pellegrini, and Daniele Miorandi. Bionets: Bio-inspired networking for pervasive communication environments. *Vehicular Technology, IEEE Transactions on*, 56(1):218–229, 2007.

[81] Wenwu Zhu, Chong Luo, Jianfeng Wang, and Shipeng Li. Multimedia cloud computing. *Signal Processing Magazine, IEEE*, 28(3):59–69, 2011.

[82] Mingfeng Tan and Xiao Su. Media cloud: When media revolution meets rise of cloud computing. In *Service Oriented System Engineering (SOSE), 2011 IEEE 6th International Symposium on*, pages 251–261. IEEE, 2011.

[83] Jian He, Di Wu, Yupeng Zeng, Xiaojun Hei, and Yonggang Wen. Toward optimal deployment of cloud-assisted video distribution services. *Circuits and Systems for Video Technology, IEEE Transactions on*, 23(10):1717–1728, 2013.

[84] Sanjit Kumar Dash, Subasish Mohapatra, and Prasant Kumar Pattnaik. A survey on applications of wireless sensor network using cloud computing. *International Journal of Computer science & Engineering Technologies (E-ISSN: 2044-6004)*, 1(4):50–55, 2010.

[85] BBP Rao, P Saluia, N Sharma, A Mittal, and SV Sharma. Cloud computing for internet of things & sensing based applications. In *Sensing Technology (ICST), 2012 Sixth International Conference on*, pages 374–380. IEEE, 2012.

[86] Arkady Zaslavsky, Charith Perera, and Dimitrios Georgakopoulos. Sensing as a service and big data. *arXiv preprint arXiv:1301.0159*, 2013.

[87] Andrea Prati, Roberto Vezzani, Michele Fornaciari, and Rita Cucchiara. Intelligent video surveillance as a service. In *Intelligent Multimedia Surveillance*, pages 1–16. Springer, 2013.

[88] Sascha Karlich, Theodore Zahariadis, Nikos Zervos, Nikos Nikolaou, Brendan Jennings, Vangelis Kollias, and Thomas Magedanz. A self-adaptive service provisioning framework for 3g+/4g mobile applications. *Wireless Communications, IEEE*, 11(5):48–56, 2004.

[89] Liang Zhou, Naixue Xiong, Lei Shu, Athanasios Vasilakos, and Sang-Soo Yeo. Context-aware middleware for multimedia services in heterogeneous networks. *IEEE Intelligent Systems*, (2):40–47, 2010.

[90] Kalevi Kilkki. Quality of experience in communications ecosystem. *J. UCS*, 14(5):615–624, 2008.

[91] Antonio J Jara, Miguel A Zamora, and Antonio Skarmeta. Glowbal ip: An adaptive and transparent ipv6 integration in the internet of things. *Mobile Information Systems*, 8(3):177–197, 2012.

[92] *EPCglobal.Inc.,"GS1 Object Name Service (ONS) Version 2.0.1",2013.1.*

[93] *"SWE", http://www.opengeospatial.org/ogc/markets-technologies/swe.*

[94] Jun Li, Yan Shvartzshnaider, John-Austen Francisco, Richard P Martin, and Dipankar Raychaudhuri. Enabling internet-of-things services in the mobilityfirst future internet architecture. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium on a*, pages 1–6. IEEE, 2012.

[95] Zhigeng Pan, Ruth Aylett, Holger Diener, Stefan Göbel, Xiaogang Jin, and Li Li. Technologies for e-learning and digital entertainment: First international conference, edutainment 2006, hangzhou, china, april 16-19, 2006, proceedings. *Lecture Notes in Computer Science*, 2006.

[96] Arun Kumar, Ashok Anand, Aditya Akella, Athula Balachandran, Vyas Sekar, and Srinivasan Seshan. Flexible multimedia content retrieval using infonames. *ACM SIGCOMM Computer Communication Review*, 41(4):455–456, 2011.

[97] Iso/iec tr 21000-1:2001, information technology multimedia framework (mpeg-21) part 1: Vision, technologies and strategy online; accessed 16-nov-2014.

[98] Anthony Vetro and Christian Timmerer. Digital item adaptation: overview of standardization and research activities. *Multimedia, IEEE Transactions on*, 7(3):418–426, 2005.

[99] Hassan Takabi, James BD Joshi, and Gail-Joon Ahn. Security and privacy challenges in cloud computing environments. *IEEE Security & Privacy*, 8(6):24–31, 2010.

[100] Alberto Coen Porisini and Sabrina Sicari. Sedap: Secure data aggregation protocol in privacy aware wireless sensor networks. In *Sensor Systems and Software*, pages 135–150. Springer, 2011.

[101] Muhamed Turkanović, Boštjan Brumen, and Marko Hölbl. A novel user authentication and key agreement scheme for heterogeneous ad hoc wireless sensor networks, based on the internet of things notion. *Ad Hoc Networks*, 20:96–112, 2014.

[102] Roberto Di Pietro, Claudio Soriente, Angelo Spognardi, and Gene Tsudik. Collaborative authentication in unattended wsns. In *Proceedings of the second ACM conference on Wireless network security*, pages 237–244. ACM, 2009.

[103] George Pallis and Athena Vakali. Insight and perspectives for content delivery networks. *Communications of the ACM*, 49(1):101–106, 2006.

[104] Michael Wooldridge. *An introduction to multiagent systems*. John Wiley & Sons, 2009.

[105] Kwang Mong Sim. Agent-based cloud computing. *Services Computing, IEEE Transactions on*, 5(4):564–577, 2012.

[106] Gerhard Weiss. *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT press, 1999.

[107] Katia P Sycara. Multiagent systems. *AI magazine*, 19(2):79, 1998.

[108] J Octavio Gutierrez-Garcia and Kwang-Mong Sim. Agent-based service composition in cloud computing. In *Grid and Distributed Computing, Control and Automation*, pages 1–10. Springer, 2010.

[109] Filip Pudlo and T Zabkowski. Multimedia data mining. *Information Systems in Management IX, Wydawnictwo SGGW, Warszawa*, pages 78–87, 2011.

[110] Jun Xin, Chia-Wen Lin, and Ming-Ting Sun. Digital video transcoding. *Proceedings of the IEEE*, 93(1):84–97, 2005.