Volume 6, Issue No. 2 - International Journal of Digital Information and Wireless Communications (IJDIWC)

Article · May 2016

READS
10

1 author:

Natalie Walker
SDIWC
55 PUBLICATIONS **0** CITATIONS

SEE PROFILE

ISSN 2225-658X (ONLINE)
ISSN 2412-6551 (PRINT)

International Journal of Digital Information and Wireless Communications

(IJDIWC)

Volume 6, Issue 2
2016

SDIWC
www.sdiwc.net

International Journal of
DIGITAL INFORMATION AND WIRELESS COMMUNICATIONS

Overview

The SDIWC International Journal of Digital Information and Wireless Communications is a refereed online journal designed to address the networking community from both academia and industry, to discuss recent advances in the broad and quickly-evolving fields of computer and communication networks, technology futures, national policies and standards and to highlight key issues, identify trends, and develop visions for the digital information domain.

In the field of Wireless communications; the topics include: Antenna systems and design, Channel Modeling and Propagation, Coding for Wireless Systems, Multiuser and Multiple Access Schemes, Optical Wireless Communications, Resource Allocation over Wireless Networks, Security, Authentication and Cryptography for Wireless Networks, Signal Processing Techniques and Tools, Software and Cognitive Radio, Wireless Traffic and Routing Ad-hoc networks, and Wireless system architectures and applications. As one of the most important aims of this journal is to increase the usage and impact of knowledge as well as increasing the visibility and ease of use of scientific materials, IJDIWC does NOT CHARGE authors for any publication fee for online publishing of their materials in the journal and does NOT CHARGE readers or their institutions for accessing to the published materials.

Publisher

The Society of Digital Information and Wireless Communications
Miramar Tower, 132 Nathan Road, Tsim Sha Tsui, Kowloon, Hong Kong

Further Information

Website: http://sdiwc.net/ijdiwc, Email: ijdiwc@sdiwc.net,
Tel.: (202)-657-4603 - Inside USA; 001(202)-657-4603 - Outside USA.

The issue date is April 2016.

IJDIWC

ISSN 2225-658X (Online)
ISSN 2412-6551 (Print)

Volume 6, Issue No. 2 2016

TABLE OF CONTENTS

Original Articles

International Journal of Digital Information and Wireless Communications (IJDIWC) 6(2): 63-77
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

An Open Source Architecture of a Wireless Body Area Network in a Medical Environment

George Kokkonis,
Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece
gkokkonis@uom.gr

Sotirios Kontogiannis,
Dept. of Informatics and Telecommunications Eng., University of West Macedonia, Kozani, Greece
skontogiannis@uowm.gr

Dimitrios Tomtsis
Department of Business Administration, Technological Educational Institute of West Macedonia, Grevena, Greece
tomtsis@teiwm.gr

ABSTRACT

This contribution describes the design and implementation of a novel Wireless Body Area Network system for healthcare monitoring, the architecture of which is based on commercially available and cost-effective hardware components, open-source software and a remote data repository for storage. In particular, the system utilizes a number of wearable medical sensors to measure different patient bio-signals which are all collected to a wearable microcomputer station. Medical sensor measurements are then categorized based on the criticality of the measurement and transmitted to a remote station. This long-range transmission is supported by a new network adaptive protocol designed to mainly transmit real-time medical data. In a prototype system, tests based on real-time and close to real-time transmission architectures have so far yielded favorable results in comparison to other existing protocols.

KEYWORDS

Medical sensors, monitoring protocols, Wearable Health Monitoring Systems, Body Area Networks.

1 INTRODUCTION

One of the most promising application domains of medical monitoring systems is the Wireless Body Area Network (WBAN), an emerging and promising health monitoring technology which promotes treatment of the specific and individual needs of each patient separately. A WBAN in a medical environment may consist of multiple wearable sensor nodes, strategically placed on the human body, which are capable of measuring and reporting the user's physiological state. The sensor nodes are designed as embedded or stand-alone devices which differ on the sensor type, size and weight. In general, each sensor node enables periodic or continuous sensing, sampling and bio-signal processing, while wirelessly transmitting one or more physiological signals (such as heart rate, blood pressure, body and skin temperature, oxygen saturation, respiration rate and electrocardiogram), but depending on the application, may also determine the user's location, state or physical activity.

In many cases, the obtained measurements are transmitted over a short distance via custom-made or industry standard wireless protocols to a small in size central node, such as a Personal Digital Assistant (PDA) or a microcontroller board, worn on the body or placed at an accessible location. Existing monitoring systems use the short-range wireless systems such as ZigBee (IEEE 802.15.4), WLANs, GSM and Bluetooth (IEEE 802.15.1). The central node may display the information on a user interface or transmit the collected vital signs to a remote medical center for diagnostic purposes by incorporating another wireless network for long-range transmission.

In this way, WBAN systems offer to the patient freedom of movement and assisted monitoring in selected living environments and also provide minimal feedback to help maintain an optimal health status. It is evident from the above, that a WBAN system may encompass a wide variety of components and advanced algorithms for data extraction, communication, storage and decision

support making, which when integrated all together, may suggest the usage of cloud computing as a potential technology.

Cloud computing can be described as an on-demand service model which is often based on virtualization and distributed computing technologies to provide information delivery and consumption where applications and information are accessed from a web browser while software and data are stored on web servers [1, 2]. Cloud components communicate and exchange data with one another using service-oriented architectures (SOA), as well as web services which allow self-contained business processes to operate over the internet [3]. The incorporation of cloud capabilities in medical applications means that medical data can be stored in a virtual archive, accessed and shared by healthcare providers and medical personnel when and as needed. The development of private cloud based services for health-care minimizes the limitations of traditional medical systems such as small physical storage, security and privacy, limited availability, processing limitations and lack of responsiveness.

This contribution describes the design and implementation of a novel WBAN system for healthcare monitoring, the architecture of which is based on low-cost off the shelf sensors and computing components, open-source software and a private repository for data storage in order to provide an economical solution to personalized health care monitoring problems, while retaining all the functionality and flexibility of more expensive systems. In particular, the system utilizes a number of wearable medical sensors to measure different patient bio-signals which are all collected to a wearable central microcomputer station. Medical sensor measurements are then categorized based on the criticality of the measurement and transmitted to a remote station. This long-range transmission is supported by a new network adaptive protocol designed to mainly transmit real-time medical data. Its novelty derives from its associated monitoring, notification, trend-calculation services and middleware protocols.

The rest of the paper is organized as follows: Section 2 presents existing health care monitoring systems and their monitoring capabilities. Section 3 describes the architecture of the proposed system and its framework, section 4 presents the classification of medical services while section 5 describes existing network protocols. Section 6 explains the design and implementation of the proposed network adaptive protocol. Section 7 presents experimental scenario results and section 8 summarizes and concludes this paper.

2 EXISTING MEDICAL MONITORING SYSTEMS

A substantial number of medical monitoring systems is described in literature, the majority of which may be grouped into the following categories:

2.1 Communication Systems

This category includes a collection of papers which reflect interesting concepts and recent advances in medical monitoring systems and applications concentrated on the design, simulation, and implementation of mobile and networking technologies.

In Yuce et al. [4] a multi-hoping 4-channel sensor network system is implemented to monitor physiological parameters from multiple patients using the Medical Implant Communication Service (402-405 MHz) to establish communication between the sensor nodes and a remote central control unit, and the Wireless Medical Telemetry Service (608-614 MHz) to exchange information between the remote central control unit and the remote base stations. These two services utilize frequency bands that are mainly assigned to medical applications. A single sensor channel is dedicated to support a continuous physiological signal such as ECC/EEG. Another example by Aydin et al. [5] present the development and implementation of a direct-sequence spread-spectrum system for real-time communication of physiological data such as temperature, PH and oxygen concentration using an implanted sensor within the system for diagnostic and monitoring purposes. Chu et al. [6] implements and evaluates the performance of a portable tele-trauma system to assist health-care centers. The system is based on commercially available 3G networks for simultaneous transmission of a patient's video, medical images and ECG signals.

2.2 Sensors

This category focuses on the implementation, issues, challenges and integration of general

purpose or application-specific sensors to the communication infrastructure of medical monitoring systems.
In Milenkovic et al. [7] a prototype sensor network is presented for health monitoring which discusses implementation issues and utilizes off the shelf 802.15.4 compliant network nodes and custom-built motion and heart activity sensors. In Jovanov et al. [8] the architecture of a wireless body area network is described composed of off the shelf sensor platforms with application-specific signal conditioning modules. A custom made sensor is developed, which can be used as a heart activity sensor and also to monitor position and activity of the upper and lower extremities. A three level wearable health system is presented using WBAN for patient monitoring. The first level consists of various physiological sensors, the second level contains the personal server and the third level includes the health care servers and other related services. In order to address some of WBAN challenges, C. Otto et al. [9] designed a system that includes accelerometer-based motion sensors, an ECG sensor, and a pocket PC based personal server. The WBAN hardware architecture is based on off the shelf commodity sensor platforms while the software architecture utilizes an open-source operating system for embedded sensor networks.

2.3 Systems

This category presents a collection of papers concentrated on the architecture and implementation of medical monitoring systems, addressing issues and challenges of system integration.
In Bourouis et al. [10] the architecture of a ubiquitous health monitoring system is presented which is based on wireless body area sensor networks to collect and send data through GPRS/UMTS to a cloud server. A smartphone application evaluates the state of a patient utilizing cloud technology, location data and a neural network running as a cloud service. In Choi et al. [11] a bedroom system for ubiquitous health monitoring is presented based on Bluetooth and Wireless LAN networks. Sensors attached to a patient's bed transmit all medical information to a remote monitoring station for further data processing and analysis. In Dai et al. [12], a wireless physiological multi-parameter monitoring system is based on mobile communication networks where the system monitors vital signs and transmits data via mobile communication networks to a mobile monitoring station and finally to a hospital central management system for data review and interpretation by medical personnel. A comprehensive review of the current research and development on medical monitoring systems is also available in Pantelopoulos et al. [13] where system implementations are compared with an emphasis given to multi-parameter physiological sensing system design in order to provide reliable vital sign measurements and to incorporate real-time decision support for early detection of symptoms.

2.4 Support Devices

The next group of papers covers different aspects of support devices, such as PDAs and tablets for medical monitoring applications.
Rodriguez et al. [14] present a location-aware medical information system that was developed to provide access to hospital information and services based on the user's location. A PDA was used to access and adapt user interfaces and resources such as patient's records or the location of a medical specialist, based on the user location detected by a trained backpropagation neural-network with radio-frequency signal strengths used as inputs. In order to reduce power consumption, connection time and at the same time increase autonomy and reliability of the monitoring system, local signal processing by PDAs was suggested by Lin et al. [15] and Rasid et al. [16]. Both systems are used for data acquisition, collection, processing and transmission.

3 SYSTEM ARCHITECTURE

The proposed system utilizes low-cost off the shelf sensors, computing and transmission components, as shown in Figure 1, to provide an economical solution to personalized health care monitoring problems, while retaining all the functionality and flexibility of more expensive systems. In particular, a WBAN is designed and developed based on wearable medical sensors to exchange patient information with a remote Central Station. Data transmission is performed using services over existing wireless technologies (3G or Wi-Fi) and protocols.

3.1 Wearable Sensors

It is the group of sensors that are attached to a patient's body and gather all medical information. In particular, the V2.0 e-Health sensors were used to monitor pulse and blood oxygenation (SpO2), airflow (breathing), body temperature, Electrocardiogram (ECG), blood sugar, galvanic skin response, blood pressure (sphygmomanometer), patient position accelerometer and electromyography (EMG). Commonly used sensor types are shown in Table 1.

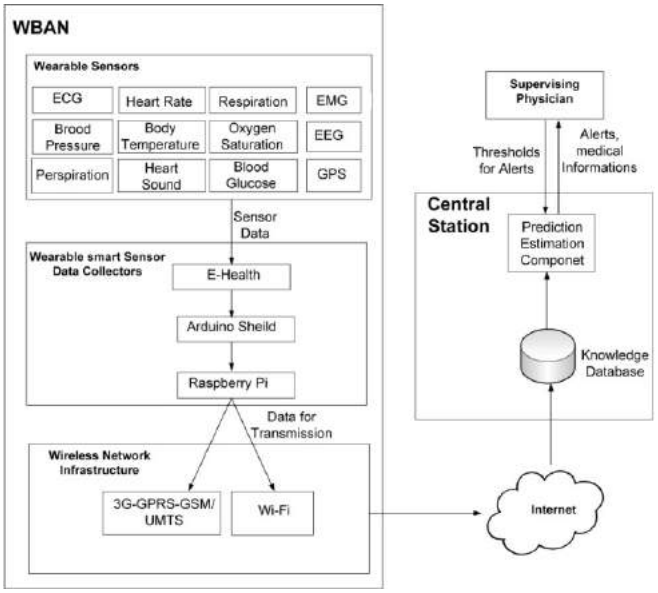

Figure 1. High level architectural diagram of the proposed WBAN system.

Wearable sensor bio-signals may be used for real-time monitoring of a patient status or to obtain close to real-time data in order to allow continuity in medical diagnosis.

3.2 Wearable Smart Sensor Data Controller

The wearable smart sensor data controller (WSSDC) is a credit card sized single-board microcomputer that collects all patient data gathered by the wearable sensors. The aggregated information is transmitted with the help of a 3G/Wi-Fi transceiver via the Internet to a remote central station. In general, biometric data can be sent wirelessly using any of the five available connectivity options (Wi-Fi, 3G, GPRS, Bluetooth and ZigBee) depending on the application. If required, a real time diagnostic image camera can be connected to the unit in order to send photos and videos of a patient to a medical diagnostic center. The WSSDC consists of the following devices:

Table 1. Wearable sensors which can be interfaced to WSSDC.

Bio Signal	Sensor Type	Description of measured data
Electrocardiogram (ECG)	Skin / chest electrodes	Electrical activity of the heart
Blood pressure (systolic and diastolic)	Arm cuff-based monitor	Refers to the force exerted by circulating blood on the walls of blood vessels, especially the arteries
Body and/or skin temperature	Temperature probe or skin patch	http://www.engadget.com/2013/06/17/geak-watch-android-china/
Respiration rate	Piezoelectric, Piezoresistive sensor	Number of movements indicative of inspiration and expiration per unit time (breathing rate)
Oxygen saturation	Pulse Oximeter	Indicates the oxygenation or the amount of oxygen that is being "carried" in a patient's blood
Heart rate	Pulse Oximeter/ Skin electrodes	Frequency of the cardiac cycle
Perspiration (sweating) or skin conductivity	Galvanic Skin Response	Electrical conductance of the skin is associated with the activity of the sweat glands
Heart sounds	Phonocardiograph	A record of heart sounds, produced by a properly placed on the chest microphone (stethoscope)
Blood glucose	Strip-base glucose meters	Measurement of the amount of glucose (main type/source of sugar/energy) in blood
Electromyogram (EMG)	Skin electrodes	Electrical activity of the skeletal muscles (characterizes the neuromuscular system)
Electroencephalogram (EEG)	Scalp placed electrodes	Measurement of electrical spontaneous brain activity and other brain potentials
Body Movements	Accelerometer Gyroscopic and magnetic sensors	Measurement of acceleration forces in 3D space
Weight measurement	Weighing scale	Measures the weight of the patient

a) E-health Arduino platform: The e-Health platform [17] is designed to measure biometric

data from medical sensors and transmit that data to the controller processor (RPI) via a Serial Peripheral Interface (SPI). Sensors connected to the E-health platform and through the RPI's communication channel (an external USB data transceiver) transmit medical data to the remote Central Station, as shown in Figure 1.

b) Arduino Shield: The connectivity between the Arduino E-health platform and the Raspberry Pi B+ was established with the use of a stackable Arduino shield.

c) Raspberry Pi B+: The Raspberry Pi B+ (RPI) is a low power (0.7-1.2W) single-board computer which includes a 32 bit, 700 MHz quad-core ARM processor with 512 MB of RAM, running a Rasbian Jessie Linux distribution. It consists of a 40 Pin GPIO that has 16 digital I/O pins used by digital sensors, an I2C interface for two wire sensors and an SPI interface for the connection for up to 8 analog sensors with the use of the MCP3008 8 channel 10-bit analog to digital chip. In this case though, a COTS Arduino based E-health platform was used. The RPI is responsible for the collection and transmission of all sensor data, from the WSSDC to the remote Central Station, using a USB Wi-Fi and a 3G dongle. The transmission protocol on the sender side is implemented on the Raspberry.

3.3 Wireless Network Infrastructure

It is the wireless network through which all data from the WSSDC is transmitted to the Central Station. All wireless communication standards that could be used for this transmission are shown in Table 2.
Existing standards fail to address the requirements of BANs, either due to security concerns [18] or interference problems [19] or due to the form factor of hardware modules or power consumption. As a response to these issues, the 802.15.6 IEEE Task Group [20, 21] is planning the development of a communication standard optimized for low power devices and operation on, in or around the human body. Furthermore, the Bluetooth Special Interest Group announced and implemented recently an alternative protocol, the Bluetooth Low Energy (BLE) [22], an ultra-low power technology for devices with limited battery capacity allowing consumption of only a small fraction of the power of the original Bluetooth products, which is targeting sports and wellness, and health care devices. Another promising wireless technology is the 5G (5th generation mobile network or 5th generation wireless system) [23] and its main target is to offer 1 GBps bandwidth with lower than 1 ms end-to-end delay. As far as its efficiency is concerned, it should be significantly enhanced compared to 4G.
Up to now, the most promising standards for long range transmission of patient information are the GPRS/3G/4G and the 802.11g because of the available range they offer. The GPRS/3G/4G uses the cellular network to offer unlimited range, while the range of a Wi-Fi network may be extended to several hundreds of meters.

Table 2. Wireless Communication Standards.

	Max. Theoretical Range	Max. Theoretical Data Rate	Power	Frequency
ZigBee	10 – 75 m	20 / 40 / 250 kbps	30 mW	868 MHz / 915 MHz / 2.4 GHz
Blue tooth	10-100 m	1-3 Mbps	2.5-100 mW	2.4 GHz
IrDA	1 m	16 Mbps	10 μW	Infrared
MICS	2 m	500 kbps	25 μW	402-405 MHz
802.11 g	200 m -1 Km	54 Mbps	100 mW-1W	2.4 GHz
GPRS/ 3G/4G	Cellular based	12Kbps/5 Mbps/12M Bps	32-200 mW	900MHz/ 1800MHz/2 .3GHz
5G	Cellular based	1 GBps	Lower than 4G	Undefined

3.4 The Central Station

This is the main station of the WBAN where all medical and positional patient data is stored. It is an Apache Tomcat JSP/PHP based application server with a MySQL database for medical data storage capabilities. A physician may access the history and real time medical information over the Internet, alter the thresholds and the destination of the medical alerts, as well as to start/stop monitoring a patient sensor.
The Central Station includes a knowledge database for the storage of current measurements and historical information which can also be used

offline either for locating an accurate and detailed patient medical history or for post processing data mining, statistical analysis and research. Such information may become useful by the Prediction and Estimation Component (PEC). Its main functionality is to perform prediction algorithms on historical database records in order to generate behavioral responses for each application incident. Data mining algorithms for statistical model development are also considered to be applicable.

4 MEDICAL SERVICES

All medical sensor bio-signals are classified into four distinct categories, depending on resolution, data update intervals and priority:

4.1 Real-time Services

In this category, real-time traffic is generated by sensors that are performing measurements at short time intervals. Examples of such sensors are the electrocardiogram, the accelerometer and the electromyography sensors. The WSSDC, either stores this data on a local drive and transmits it later on or sends it on the fly. The first approach demands storage space while the second approach demands good network conditions. Real-time services follow the second approach. Such devices may support a delay buffering mechanism of no more than 1-3 seconds per transmission.

4.2 Periodic Services

In this category, sensor data are periodically transmitted to the Central Station. Examples of such services are the notification service, the reminding service, the nutrition proposal service and the health cognitive service. Cognitive services provide a health model based on previous sensor measurements and use data mining to decide if a patient is in accordance with a specific model.
Periodic services require a shared timing mechanism set by the Central Station at the application layer. Periodic services are RESTful web services that take as input JSON data (JavaScript Object Notation) and reply with XML or JSON responses. Periodic services may include an authentication and encryption mechanism with the use of SSL/TLS protocols and x.509 certificates for authentication.
Furthermore, they are divided into periodic stream services (where the HTTP v.1.1 persistent session with the central session never expires) and periodic trend services (where the HTTP connection is re-initiated every time the WBAN system needs to transmit data).

4.3 Close to Real-time Services

In this category belong services that trigger on demand sensor measurements. A close to real-time service initiates sensor data measurement and transmission when there is indication of an abrupt medical status such as rapid temperature increase, fast posture and acceleration change. When a patient's results are not within the thresholds set remotely by medical personnel then alerts are produced and send by the system.
Close to real-time services need to be treated with respect to the content of the transmitted message. That is, a message acknowledgment mechanism is needed and a persistent retransmission mechanism for every packet loss is required.

4.4 Interactive Services

This category includes mostly multimedia medical services and action feedback services. That is, voice and video transmissions and Haptic services. Haptic data transmission is denoted by devices that provide real-time sense or touch or precise spatial position interaction between patient and doctor.
Interactive voice and video services use the RTP (Real-Time Transport Protocol) that includes required timing and synchronization mechanism for multimedia flows.

5 EXISTING REAL-TIME PROTOCOLS

Many interesting protocols have been introduced for the transfer of real-time medical data over the internet. The most important of them are:

5.1 Transmission Control Protocol (TCP) – HTTP/TCP

The transmission control protocol (TCP) is a transport layer network protocol that provides a reliable and connection-oriented service to medical internet applications. The TCP mechanism of

packet retransmissions assists on packets reception. TCP also performs congestion control in order to regulate the transmission rate and provides a typical flow authentication mechanism by establishing a 3-way handshake prior to data delivery. For real-time services, the TCP retransmission mechanism and the congestion control algorithms lead to excess jittering, which is not suitable for real-time traffic.

TCP is commonly used by periodic services, which use the HTTP protocol for the end-to-end delivery of medical data using XML and JSON data notation. This promotes the HTTP protocol as a reliable end-to-end protocol for periodic and close to real-time medical services.

5.2 User Datagram Protocol (UDP)

UDP is an unreliable and connectionless protocol for medical applications that transmits interactive and real-time data but it is considered to be a good candidate protocol for multimedia streaming and voice transmission over the internet protocol (VoIP). The UDP uses packet flows (datagrams) transmitted at constant rates without acknowledgments, loss or packet reorder provisions. Unlike TCP, it does not guarantee data packet delivery or rate adjustments in case of congestion. Hence, data transfers using UDP can be accomplished without significant time delay and variations, if indeed delivered.

In medical systems, UDP is useful for the transmission of real-time insensitive data, without a guarantee of a reliable transmission, but if channel capacity is exceeded then it leads to data loss and real-time flow corruption. UDP is a best effort protocol for the transmission of interactive services. Even if it does not guarantee data reliability, since it manages to maintain a constant flow rate by minimizing jitter.

5.3 Real-Time Transport Protocol (RTP)

RTP has been introduced as a successor of UDP, to carry interactive flow information between two end-systems. It includes flow control and packet retransmission provisions, in order to maintain real-time flow consistency [24]. It is augmented by another protocol called RTCP [25], which provides feedback to both communicating participants regarding data transmission such as the inter-arrival jitter estimate, information about the highest sequence number received, fraction of packets lost and cumulative number of packets lost. The RTP is mainly used by real-time medical applications that include video and audio streaming services.

5.4 Interactive Real-Time Protocol (IRTP)

The IRTP has been designed for interactive internet-based services [26] in order to reduce the end-to-end delay. It uses the TCP handshake for flow connection establishment, as well as the RTP flow control, and a re-transmission mechanism for crucial data. For real-time and interactive non-crucial data, the IRTP uses TCP in contrast to UDP for crucial data flow delivery.

5.4 Efficient Transport Protocol (ETP)

ETP is an alternative protocol to IRTP. It is also designed for internet-based interactive services. ETP focuses on reducing the end-to-end delay by introducing a time gap called the IPG (Inter Packet Gap), between two successive data packets [27]. This IPG-based control mechanism provides congestion control similar to the TCP window size based congestion control. ETP is UDP based. It does not differentiate real-time medical services into critical and non-critical, but in cases of network congestion attempts to provide a uniform end-to-end delay by modifying data packets in flight based on receiver IPG feedback.

6 PROPOSED PROTOCOL

In order to accomplish long-range transmission of medical data, from the WBAN controller to the remote Central Station, a protocol was designed and implemented, named MESETP (Medical Services Transport Protocol). This protocol includes three sub-protocols, where each one was designed to meet the requirements of a different medical service (periodic, close to real-time and real-time, described in sections 4.1 to 4.3). The following sections describe each sub-protocol but concentrate mainly on the development of the Adaptive Medical Sensor Transmission Protocol (AMESETP), a new network adaptive transmission protocol for real-time medical services.

6.1 MESETP for Periodic Services

For periodic services MESETP uses an application layer RESTful protocol with different options per service. If it is a non-streaming periodic service, then HTTP v.1.0 requests are used with keep-alive timeouts. If it is a streaming service, then HTTP v.1.1 persistent connections are performed to maintain connectivity at non-streaming time intervals.

6.2 MESETP for Close to Real-time Services

This protocol used for close to real-time services is based on a custom TCP half-open connection mechanism to transmit sensor data over an HTTP or an authenticated by SSL secure HTTP (ports 80 and 443 respectively). Data acquisition is accomplished with the exchange of only two data packets. The first one is used for a connection/data initiation transmission (TCP connection initiation SYN packet), while the second is used as a connection acknowledgement packet (TCP acknowledgement packet reception). Upon one packet exchange gesture a TCP RST packet is send by the sender indicating HTTP connection termination. Data collection of this arbitrary data exchange is performed and buffered at kernel level thus reducing unnecessary application processing delays during data exchange.

6.3 MESETP for Real-time Services

For real-time services MESETP protocol is renamed to Adaptive Medical Sensor Transmission Protocol (AMESETP). In general, the AMESETP reads the network status in frequent intervals and depending on the detected network conditions, it adapts the sending rate of the data packets and the quantization levels of the transmission in order to overcome congestion incidents. The header of the AMESETP protocol for a single data packet is depicted at Figure 2.

Bits	***0***	***1-4***	***5-7***	***8 - 15***	***16-23***	***24-31***
0	Sequence Number				CheckSum	
32	En	SRL	SenId	PatientId		-
	Data					

Figure 2. Header of AMESETP sub-protocol.

AMESETP protocol header includes the following fields:
The *Sequence Number*, used to reinforce the reliability of the protocol through the acknowledgment process.
The *Checksum* used for error-checking of both header and payload data.
The *Sending Rate Level* (SRL) field that informs the receiver for the sending rate level, described in following section. Based on this level the sending data rate and the packed data values per packet can be calculated (see Table 4).
The *Sensor ID* (SenID) field informs the receiver to which sensor the data correspond to.
The *PatientID* field informs the receiver to which patient the data correspond to.
The *Encryption* (En) field informs the receiver if encryption is used for security reasons. As the AMESETP carries private medical data, the symmetric 3DES with ECB mode encryption is enforced at the application layer.
Table 3 shows the efficiency of the AMESETP when it is compared to other common transport protocols such as the UDP, the RTP and the TCP.

Table 3. AMESETP Protocol Efficiency

	AMESETP	***UDP***	***RTP***	***TCP***
HEADER (bytes)	7	8	12+8(UDP)	20
PAYLOAD (bytes)	10-33	10-33	10-33	10-33
EFFICIENCY %	58.82- 82.50	55.55-80.48	33.33-60.00	33.33-60.00

The overhead of the AMESETP header is only 7 bytes. The payload of each packet varies from 60 bytes to 500 bytes depending on the number of values that are grouped and the quantization level that is used. The maximum transmission rate of a real-time sensor is about 2 bytes of sensor data per ms. Using this value as a nominal real time transmission value the result is a data recording rate of 2 Kbytes/s per sensor at the WBAN system. AMESETP protocol packs these values into packets using as initially transmission rate the Normal data at Table 5, Level 5.
The Normal data rate sends 10 packets/s (10 Hz) of 200 bytes of payload that correspond to 100 (2 byte) sensor values per packet and a quantization level of 16bit per value. The level 6 Quantization-Normal data rate has values of 8 bit that correspond to a smaller amount of payload data per packet. The combination of rates and bits per sensor value (quantization of value) that the

AMESETP protocol uses is presented at Table 5 and is consisted of 12 distinct levels. The higher the level the more frequent and more quantized the sensor values. The rate level that AMESETP uses is controlled by the acknowledgment packets send by the receiver and the rate control mechanism described in the following section.

6.3.1 AMESETP Rate Control Mechanism

Based on existing real-time protocol measurements, the maximum acceptable values of delay, jitter and packet loss of the video, audio and graphic data are shown in Table 4 [28]. These quantities were evaluated based on the Quality of Experience that a user feels when he uses the corresponding real-time or interactive services.

Table 4. Tolerable delay time reports for real-time audio, video and graphics [28].

QOS	*VIDEO*	*AUDIO*	*GRAPHICS*
JITTER (ms)	≤ 30	≤ 30	≤ 30
DELAY (ms)	≤ 400	≤ 150	≤ 100-300
PACKET LOSS (%)	≤ 1	≤ 1	≤ 10
REFRESH RATE (Hz)	≥ 30	≥ 50	≥ 30

When network conditions deteriorate, then the proposed protocol lowers its sending rate in order to avoid congestion [28]. To lower the sending rate, sensor generated measurements (at a maximum steady rate of 1 KHz) are grouped together into bigger data packets or quantized to a number of less accuracy bits. The size of bytes per packet and quantization of sensor values changes according to the network conditions shown in Table 5.

Table 5. AMESETP Protocol Rates and Compressed Rates for Real-time Services.

Sending Rate	Level	Packets/s (Pkt-Hz)	Values/ Packet	Bytes/ packet	Quantization Bits/value
High data rate	1-0000	33	30	60	16
Q-High data rate	2-0001	33	30	30	8
Medium data rate	3-0010	20	50	100	16
Q-Medium Data rate	4-0011	20	50	50	8
Normal data rate	5-0100	10	100	200	16
Q-Normal data rate	6-0101	10	100	100	8
Q-Slow data rate	7-0110	5	200	300	12
Q-very slow data rate	8-0111	5	200	250	10
Q-very very slow data rate	9-1000	5	200	200	8
Extra slow data rate	10-1010	4	250	250	8
Compressed 1 data rate	11-1011	2	500	375	6
Compressed 2 data rate	12-1100	2	500	250	4

The mechanism for AMESETP rate level adaptation is the following: At the receiver, the AMESETP records periodically packet loss, as well as the inter-package delay variation time based on Eq. 1. The inter-packet delay variation is measured using the average packet reception gap among two consecutive packets, as shown by Eq. 2:

$$Td_i=(R_i-R_{i-1})-(S_i-S_{i-1}+Ki+IPGi) \quad (1)$$

$$Td_i=dt(R_i-R_{i-1})-0.0013 \quad (2)$$

Where Tdi is the inter-package delay time for packet i, R_i is the reception time of packet i, S_i is the transmission time of packet i and IPG is the inter-packet gap set by the technology used. K_i is the user to kernel space delay insertion of a packet at the transmitter, an average experimental measurement of which is 120.5 usec for each packet insertion at kernel level for the RPI.
Because data transmission by the sender is performed at a maximum of 1ms intervals and since the IPG during packet insertion is close to 9.2 sec, the negative term of eq. 1 is a constant value (Eq. 2). It should be noted that the average 10BASE-T IPG value of 9.6 us is used for both 3G and Wi-Fi technologies for packet insertion. The difference between two consecutive time inter-packet delays corresponds to the value of inter-arrival jitter estimate or inter-packet delay variation J_i. J_i value is then smoothed (SJ_i) for AMESETP based on Eq. 3.

$$J_i=Td_i- Td_{i-1},$$

$$SJ_i=(15/16)J_{i-1}+(1/16)Td_i \quad (3)$$

Eq. 3, is similar to the equation used by the RTP protocol for the calculation of the inter-arrival jitter at the receiver in order to adjust its flow rate.
SRL metric (Eq. 4) defines AMESETP transmitter rate level (Table 5). When an increased packet loss is detected then the transmitter is notified by an acknowledgment packet to lower its rate (the number of the grouped values is increased based on Table 5). There are 12 levels of grouping and/or compression of sensor values per packet transmission and is changing according to Table 5 based on Sending Rate Level (SRL) and calculated from Eq. 4.

$$SRL=\begin{cases} l_0=5 & \\ l_{i-1}+1 & ,pl<\frac{pl_{max}}{2} \ \ or \ \frac{SJ_{i-1}}{4}>SJ_i \\ l_{i-1} & ,\frac{pl_{max}}{2}<pl<pl_{max} \\ l_{i-1}-1 & ,pl_{max}<pl \end{cases} \quad (4)$$

AMESETP SRL metric divides packet loss into three intervals: The first packet loss interval $pl<pl_{max}/2$ or $SJ_i<SJ_{i-1}/4$ corresponds to perfect network conditions and the protocol attempts to increase the sending rate. The second interval $pl_{max}/2<pl<pl_{max}$ corresponds to good network conditions and the protocol attempts to keep the sending rate steady. The last interval $pl>pl_{max}$ corresponds to unacceptable network conditions and the protocol lowers the sending rate in order to avoid congestion.
For example when AMESETP detects a decrement on packet loss based on Eq. 4 or an average jitter J_i decrement of Ji/4 for a measurement receiver period then it signals via ACK the transmitter to increase its sending rate by one level based on Table 5. If packet loss is more than maximum packet loss (pl_{max}) then the sending rate is decreased according to Eq. 4 by one level based on Table 5 and so forth.
The packet loss *pl* can be evaluated by the negative acknowledgements of the lost packets that the receiver sends to the sender. The factor pl_{max} is the maximum acceptable packet loss taken from Table 4. For real-time graphics the maximum acceptable packet loss is set to $pl_{max}=10\ \%$.
Another method the AMESETP uses to avoid congestion is the network adaptive quantization where quantization levels are changing according to detected network conditions. When network conditions are good, the quantization levels increase and vice versa. Three different levels are used, denoted as the minimum, medium and maximum quantization levels ($ql_{min}=16=2^4$, $ql_{med}=4096=2^{12}$, $ql_{max}=65536=2^{16}$) at Table 5 respectively.

6.3.2 AMESETP Flow Control Mechanism

AMESETP flow control mechanism uses a unique sequence number for every packet and the method of the cumulative negative acknowledgement (CNAK) for packet drops signaling. Moreover, the sender transmits the packet with the grouped values as soon as they are created. A copy of this packet is stored in a buffer in case a negative acknowledgement is received. The maximum size of the sender buffer is set to 65536 packets. The maximum packet size and headers is 241 bytes. Consequently, the maximum buffer size is 16 Mbytes.
The receiver sends a CNAK with the packets that have not reached their destination every k sec (where k=2-12sec based on protocol configuration). The receiver also sends the sequence number of the last received packet to help the sender to empty its buffer.

Bits	***0-3***	***4-7***	***8 – 15***	***16-31***
0	Sensor Id	ACK Sequence Number		CheckSum
32	PatientId			-
	Data (Last Received Sequence Number, Dropped Sequence Numbers)			

Figure 3. The AMESETP Header of Cumulative Negative Acknowledgment Packets (CNAK).

The maximum period of time for the CNAK cannot exceed the $k_{max}=212$ sec as the sequence number will be reset at this time for a packet rate of 33 Hz. In our case study we set the period for cumulative acknowledgment *k=1 sec.* The packet header for the CNAK is shown in Figure 3.
To enforce reliable transmission of the CNAK packet an acknowledgment is sent by the sender. If the receiver doesn't get the acknowledgment in a specific time period, a copy of the CNAK is sent again.

A flow diagram of the proposed protocol with the CNAK is shown in Figure 4.

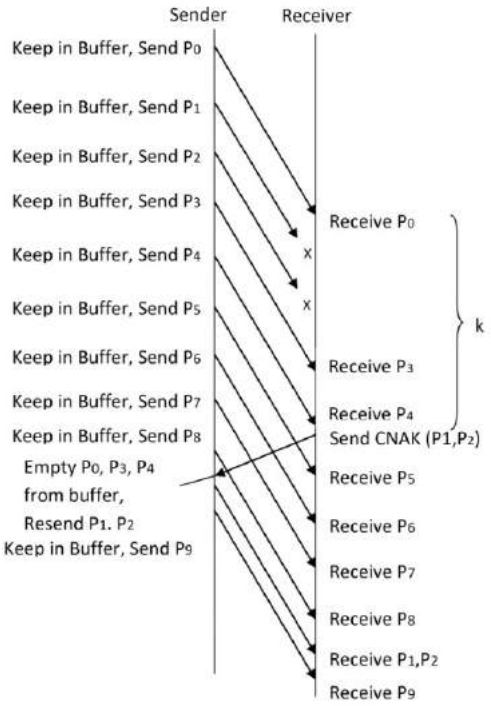

Figure 4. Flow Diagram of the AMESETP.

7 AMESETP PERFORMANCE

The AMESETP performance was experimentally tested against the UDP and the RTP for real-time services (Scenario I, II), as well as the TCP protocols for close to real-time services. All scenarios include transmission flows from the WSSDC to the Central Station (an Intel core I7 32bit system at 3.2GHz and 8GB DDR3 RAM – running Ubuntu server 14.10 Linux distribution). Two different types of communication channels were used during experimentation:

Type A - 3G HSDPA (High Speed Downlink Packet Access): HSDPA provides packet data transmission over UMTS – W/CDMA cellular systems, by utilizing reduced delays, and a peak raw data rate of 14 Mbps. It also provides around three times the capacity of the 3G UMTS technology. That is a maximum of 14Mbit downlink speed and 512Kbit/s uplink (Average speeds tests showed for the test link 4.5Mbit/s Downlink and 420Kbit/s Uplink speeds). A 3G USB dongle was plugged into the RPI controller to transmit data.

Type B – Wi-Fi IEEE802.11n wireless network transmission over ADSLv2: Wi-Fi transmission provides a symmetrical radio channel at an average of 54Mbit/s from the medical data collector to a central access point where data are redirected to the central medical station using ADSLv2 wired transmission. ADSL utilized a maximum downlink capacity of 24Mbps and an uplink of 1Mbit/s (Average speed tests for the test link showed 12Mbit/s Downlink and 784Kbit/s Uplink).

7.1 Scenario I Experimental Results

In this scenario, 20 EMG real-time flows were transmitted from the WSSDC to the Central Station using UDP protocol. The same transmission was then repeated but this time using the AMESETP for real-time services. In addition, a single real-time EMG flow was transmitted using an HTTP/TCP communication channel. All the transmissions shared the same measurements from a 2byte/ms sensor source which was transmitting a single data packet every 100ms (19.6Kbit/s per flow). Table 6 summarizes the experimental results. As it is seen, the AMESETP implementation outperforms the simple UDP transmission in the case of less packet drops due to the NACK periodic retransmission mechanism.

Table 6. AMESETP performance comparison results with other common real-time protocols.

Wi-Fi	3G	Min Inter-packet delay (msec)		Max Inter-packet delay (msec)		% Packet Drops	
UDP		48	124	1048	1214	2	8
AMESETP		73	114	912	998	-	2
TCP RTT		260	448	389	1042	-	-

As far as the 3G transmission is concerned, latency is introduced by the communication channel. Continuous system testing over long periods of time showed that the 3G transmission can tolerate spurious or continuous latencies up to 890msec before packet drops occur. 3G communication channel is also jitter tolerant maintaining similar inter-packet delays for AMESETP and UDP flows. In general, 3G is preferable in a mobile environment due to its extended range of transmission. However, in a good 3G signal reception, the packet delivery delay time is less than 130 msec. In environments with increased contention, low signal and/or mobility the inter packet delay time may reach up to 1200 msec. Furthermore, significant jitter was spotted in a 3G

network with increased mobility. In addition, Wi-Fi networks with the same environmental conditions, present spurious and burst drops.

As far as the Wi-Fi transmission is concerned, continuous system testing over long periods of time showed that it is preferable in less mobile and more stationary environments. When the WSSDC was in optical contact with an Access Point with a signal power more than -68dBm, then the Wi-Fi network provided fair real-time services with inter-packet delays less than 120 msec. The Wi-Fi network is less tolerant to excessive jitter or delays and responded with packet drops for simple UDP flows when the packet inter-arrival time was more than 168 msec. Moreover, the Wi-Fi transmission medium introduces important jitter when there is no optical contact (multipath propagation) or interference from other sources close to the Wi-Fi access point. If such jitter persists more than 2-3 sec it may lead a UDP flow to burst drops and the AMESETP protocol to a full stop.

7.2 Scenario II Experimental Results

In this scenario, Real-Time Medical sensor data flows (RTMs) are transmitted from the WSSDC to the Central Station in order to experimentally compare the performance of the UDP, the RTP over HTTP tunneling and the AMESETP protocols. Such flows simulate real ECG (Electrocardiogram) transmissions when a three-electrode wearable sensor is attached to a patient. Each flow transmits a total 1 Mbyte of medical data then it stops. Sensor data are transmitted every 1msec and the sensor resolution is 16bits (2 bytes of data every 1 us, equivalent to sensor real-time rate of 2 Kbyte/s of data per second and per flow). The experiment was repeated using 5, 10, 25, 50, 75 and 100 concurrent flows of AMESETP protocol and then the same experiment was repeated using the same number of flows for UDP and RTP protocols respectively.

Figures 5 and 6 show typical results per flow of the average Goodput of the AMESETP, the UDP and the RTP using type-A 3G and type-B Wi-Fi networks respectively. (Goodput is the payload reception rate at the receiver end in terms of Kbit/sec–without the protocol headers). It is obvious that in both cases the AMESETP outperforms the UDP and performs at least as equally as the RTP for a low number of flows or even better for a higher number of flows.

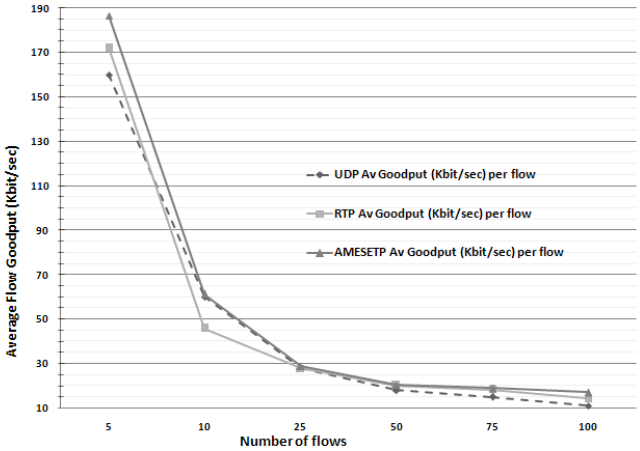

Figure 5. Per flow average Goodput of AMESETP, UDP, and RTP protocols over number of flows, using a Wi-Fi network

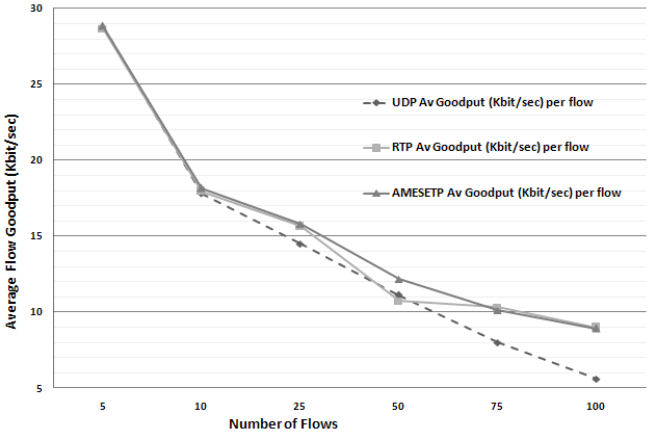

Figure 6. Per flow average Goodput of AMESETP, UDP, and RTP protocols over number of flows, using 3G network

In Table 7, Goodput gain results for both type-A and B networks are shown. Goodput gain is the average % better performance of an AMESETP flow over UDP and RTP flow average Goodput accordingly. It is expressed as the average Goodput difference of AMESETP flows minus average Goodput of UDP-RTP flows over AMESETP flows average Goodput.

Table 7. AMESETP per flow % Goodput gain over UDP and RTP for Wi-Fi and 3G transmission medium over number of concurrent flows

Wi-Fi	3G	AMESETP Per flow % Goodput Gain Over RTP Protocol		AMESETP Per flow % Goodput Gain over UDP Protocol	
AMESETP 5 flows		0.7	-	1.4	0.5
AMESETP 10 flows		2.5	0.5	2.0	1.9
AMESETP 20flows		3	1.0	2.8	8.2
AMESETP		4.1	5.0	8.2	8.8

50 flows				
AMESETP 75 flows	5	6.1	12	12.8
AMESETP 100flows	7.2	9.7	19.4	17.14

It is obvious that the AMESETP significantly outperforms UDP real-time flows for both 3G and Wi-Fi networks. The Goodput gains for up to 10 flows is 1-1.5% more Goodput on average per flow over a UDP flow carrying the same data using the same initial transmission rate with the AMESETP.

For less than 3 flows and non-congested links the AMESETP and the UDP perform similarly (no Goodput gain of AMESETP over UDP was detected). For more than 10 flows the UDP transmission deteriorates significantly to burst drops (10-15% packet drops per flow on average – Table 8).

Table 8. AMESETP, UDP, RTP cumulative %packet drops for all flows for Wi-Fi and 3G transmission medium over number of concurrent flows

Wi-Fi	3G	%Packet Drops UDP		%Packet Drops RTP		% Packet Drops AMESETP	
5 flows		0.1	0.2	0	0	0	0
10 flows		0.5	1.0	0.1	0.1	0.5	0.5
20 flows		1.1	5.0	0.2	0.1	1.0	2.0
50 flows		2.0	9.0	1.2	4	1.5	3.5
75 flows		4.7	12	2.0	8	3.1	8
100 flows		9.1	20	4.1	11	7.0	9

The RTP performance is significantly better than that of UDP. For less than 10 flows the RTP presents similar performance characteristics with the AMESETP in terms of Goodput and packet drops (Tables 7, 8). As the number of flows increases the RTP abrupt rate scaling mechanism decreases significantly RTP transmission rates. This can be seen both at Goodput gains of Table 7 for rates of 50 flows and over. This rate decrement leads also to significant burst packet drops at least 2 times more than that of the AMESETP.

Figures 8 and 9 show the average inter-packet delay time for the Wi-Fi and the 3G networks respectively. It is evident from these figures that the UDP presents less jitter in terms of inter-packet delay time than the AMESETP, which is more than 20% as the number of flows increases, but there is also a significant increase of packet drops (10-20%) which lowers its Goodput.

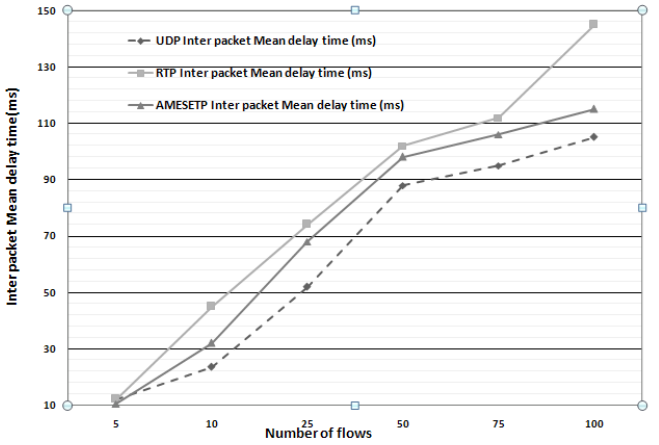

Figure 7. Average inter packet delay time (ms) per flow of AMESETP, UDP, and RTP protocols over number of flows, using Wi-Fi network

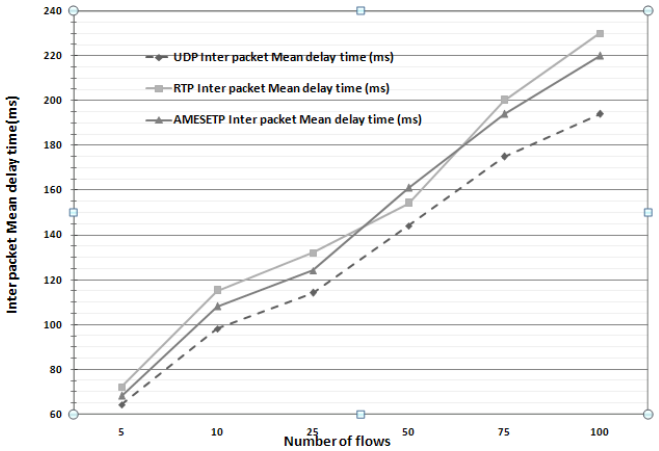

Figure 8. Average inter packet delay time (ms) per flow of AMESETP, UDP, and RTP protocols over number of flows, using 3G network

The same figures show that the RTP performance is comparable to that of the AMESETP as far as inter-packet delay time is concerned. For the Wi-Fi transmission, the AMESETP performs better than the RTP in terms of jitter for both low and high number of flows, presenting at an average 10-15% less inter-packet delay. For the 3G transmission, the inter-packet delay of the RTP is less than that of the AMESETP for low flow numbers, but the situation reverses for higher than 50 flows numbers. This is an indication that the AMESETP attempts to increase the sending rate level (SRL), but in this case it fails to do so. Possibly a more aggressive adaptation rate is required based only on jitter or inter-packet delays.

8 CONCLUSIONS

This contribution describes the design and implementation of a novel WBAN system, the architecture of which is based on low-cost off the

shelf hardware components, open-source software and programming environments such as the Linux (RPI) and the Arduino IDE, to provide an economical solution to personalized health care monitoring problems.

In particular, the system utilizes the V2.0 e-Health wearable sensors to measure different patient bio-signals (airflow, body temperature, electrocardiogram, blood sugar, etc.), which are all collected to a wearable microcomputer station. Medical sensor measurements are then categorized into real-time, close to real-time and periodic based on resolution, data update intervals and priority of the measurement and transmitted to a remote station. The remote station serves as a repository for data storage, as well as a data processing unit where a physician may access the history and real time medical information, alter thresholds and send alerts.

The long-range transmission between the microcomputer and the remote station is supported by a new real-time network protocol, which reads the network status in frequent intervals and depending on the detected network conditions, such as packet loss, jitter and inter-package delay time, it adapts the sending rate of the data packets and the quantization levels of the transmission in order to avoid congestion.

In a prototype system, experimental tests have shown that the proposed protocol is a promising candidate for real-time Wi-Fi or 3G transmission of medical data since it outperforms in most cases other existing protocols, such as the TCP, UDP and RTP, although a more aggressive adaptation rate may improve its performance. However a better calibration of AMESETP rate level change for non congestive intervals may further increase AMESETP performance characteristics.

REFERENCES

[1] R. Buyya, C.S. Yeoa, S. Venugopala, J. Broberga, I. Brandicc, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems, Vol. 25, pp. 599-616, 2009.

[2] P. Bruening, B. Treacy, "Cloud Computing: Privacy, Security Challenges". In The Bureau of National Affairs, 2009.

[3] M. Poulymenopoulou, F. Malamateniou, G. Vassilacopoulos, "Emergency Healthcare Process Automation using Workflow Technology and Web Services". Int. J. Med. Inform., Vol. 28, no. 3, pp. 195-207, 2005.

[4] M.R. Yuce and C.K. Ho, "Implementation of body area networks based on MICS/WMTS medical bands for healthcare systems", 30th Annual International IEEE EMBS Conference, August 20-24, 2008.

[5] N. Aydin, T. Arslan, and D.R.S. Cumming, "A direct sequence spread-spectrum communication system for integrated sensor microsystems", IEEE Transactions on Information Technology in Biomedicine, Vol. 9, No 1, pp. 4-12, March 2005.

[6] Y. Chu and A. Ganz, "A mobile teletrauma system using 3G networks", IEEE Trans. Inform. Technol. Biomed., vol. 8, no. 4, pp. 456–462, Dec. 2004.

[7] A. Milenkovic, C. Otto, E. Jovanov, "Wireless sensor networks for personal health monitoring: Issues and an implementation", Elsevier, Computer Communications 29 (2006), pp. 2521–2533.

[8] E. Jovanov, A. Milenkovic, C. Otto and P. C. de Groen, "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation", Journal of NeuroEngineering and Rehabilitation, Vol. 2, March 2005.

[9] C. Otto, A. Milenkovic, C. Sanders, E. Jovanov, "System Architecture of a Wireless Body Area Sensor Network for Ubiquitous Health Monitoring", Journal of Mobile Multimedia, Vol. 1, No.4 (2006) pp. 307-326.

[10] A. Bourouis, M. Feham and A. Bouchachia, "A new architecture of a ubiquitous health monitoring system: a prototype of cloud mobile health monitoring system." International Journal of Computer Science Issues, Vol. 9, Issue 2, No 2, March 2012.

[11] J.M. Choi, B.H. Choi, J.W. Seo, R.H. Sohn, M.S. Ryu, W. Yi, K.S. Park, "A System for Ubiquitous Health Monitoring in the Bedroom via a Bluetooth Network and Wireless LAN", Proc. The 26th Annual International Conference of the IEEE Eng. Med. Biol. Soc. (EMBS), Vol. 5, 2004, pp. 3362-3365.

[12] S. Dai, Y. Zhang, "Wireless Physiological Multi-Parameter Monitoring System Based on Mobile Communication Networks", In 19th IEEE Symposium on Computer-Based Medical Systems Based on Mobile Communication Networks, Washington, DC, USA: IEEE Computer Society, 2006, pp. 473-478.

[13] A. Pantelopoulos and N. G. Bourbakis, "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 40, No. 1, January 2010.

[14] M. Rodriguez, J. Favela, E.A. Martinez, and M.A. Muñoz, "Location-aware access to hospital information and services," IEEE Transactions on Information Technology in Biomedicine, Vol. 8, no. 4, pp. 448–455, Dec. 2004.

[15] Y.H. Lin, Y.Y. Chen, P.C.I. Ko, I.C. Jan, J.M. Wong, and G.J. Jan, "A wireless PDA-based physiological monitoring system for patient transport," IEEE Trans. Inform. Technol. Biomed., vol. 8, no. 4, pp. 439–447, 2004.

[16] M.F. Rasid, B. Woodward, "Bluetooth telemedicine processor for multi-channel biomedical signal transmission via mobile cellular networks", IEEE Trans. Inform. Technol. Biomed., 2005 Mar;9(1):35-43.

[17] E-Health Sensor Platform V2.0 for Arduino and Raspberry Pi. https://www.cooking-hacks.com/-documentation/tutorials/ehealth-biometric-sensor-platform-Arduino-raspberry-pi-medical, 2013.(Accessed 31 Dec 2015)

[18] C. Hager and S. Midkiff, "An analysis of bluetooth security vulnerabilities", IEEE Wireless Communications and Networking, vol. 3, pp. 1825–1831, March 2003.

[19] A. Sikora and V. F. Groza, "Coexistence of IEEE 802.15.4 with other systems in 2.4 GHz ISM-Band", In Proc. of Instrumentation and Measurement (IMTC), vol. 1, pp.1786-1791, May 2006.

[20] S. Ullah, P. Khan, N. Ullah, S. Saleem, H. Higgins, and K. S. Kwak, "A review of wireless body area networks for medical applications", CoRR, abs/1001.0831, 2010.

[21] IEEE 802.15 WPAN Task Group 6. Available at: http://www.ieee802.org/15/pub/-TG6.html, 2009. (Accessed 31 Dec 2015).

[22] C. Gomez, J. Oller, J. Paradells, "Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-power Wireless Technology", Sensors2012, vol 12, pp-11734–11753, August 2012.

[23] J.G. Andrews, S.Buzzi, Wan Choi; S.V. Hanly, A.Lozano, A.C.K. Soong, J.C. Zhang, "What Will 5G Be?", IEEE Journal in Selected Areas in Communications, vol. 32, no. 6, pp.1065-1082, June 2014

[24] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson "RTP: A Transport Protocol for Real-Time Applications", RFC 3550, 2003.

[25] J. Ott, C. Perkins, "Guidelines for Extending the RTP Control Protocol (RTCP)", RFC 5968, September 2010.

[26] L. Ping, L. Wenjuan, and S. Zengqi, "Transport layer protocol reconfiguration for network-based robot control system," in Networking, Sensing and Control, 2005. Proceedings. 2005 IEEE, Mar. 2005, pp. 1049 – 1053.

[27] R. Wirz, M. Ferre, R. Marin, J. Barrio, J. Claver, and J. Ortego, "Efficient transport protocol for networked haptics applications," in Haptics: Perception, Devices and Scenarios, ser. Lecture Notes in Computer Science, M. Ferre, Ed. Springer Berlin / Heidelberg, 2008, vol. 5024, pp. 3–12.

[28] G. Kokkonis, K. E. Psannis, M. Roumeliotis,"Network Adaptive Flow Control Algorithm for Haptic Data Over the Internet–NAFCAH", Genetic and Evolutionary Computing, vol. 388, pp 93-102, Sept. 2015.

International Journal of Digital Information and Wireless Communications (IJDIWC) 6(2): 78-86
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

Enhancement of the Fusion of Incompatible Lists of Results

Guezouli Larbi and Azzouz Imane
LaSTIC, University of Batna 2,
Algeria
{larbi.guezouli,imane.azzouz}@univ-batna.dz

ABSTRACT

This work is located in the domain of distributed information retrieval (DIR). A simplified view of the DIR requires a multi-search in a set of collections, which forces the system to analyze results found in these collections, and merge results back before sending them to the user in a single list.
Our work is to find a fusion method based on the relevance score of each result received from collections and the relevance of the local search engine of each collection, which is the main issue of our work.

KEYWORDS

Text mining; distributed information retrieval; content-based retrieval; merging results.

1 INTRODUCTION

The growth of the number of servers on the worldwide network makes the management of a huge quantity of information an obligation. For this, the use of information retrieval systems became indispensable especially in the distributed world. To organize this huge quantity of information, web search engines find their origins in the information retrieval systems developed to find, from a database of documents, relevant documents to a user request. [1], [2], [3], [4], [5], [6]
Information retrieval systems can be categorized into classes based on their architectures: centralized and distributed. Centralized systems require that all documents were located on the same site. But distributed systems were considered as a set of independent information retrieval systems allowing the simultaneous access to distributed collections of documents on local or wide network. [7], [8], [9]

1.1 Problematic

One of essential problems of distributed information retrieval systems is the fusion of results obtained from a set of local information retrieval systems. The user must receive one list as response to his request. This list is the fruit of the fusion of different result's lists of different local information retrieval systems.
For example, user presents a query to the distributed information retrieval system. The system forwards the query to selected servers. Each server calls its local information retrieval system to search for the query. Each server provides a list of results to the distributed system. Finally, the distributed system merges all the lists of results in a single list and returns it to the user.
Therefore, the main problem is to define a set of rules for merging these lists.

1.2 Background

The field of information retrieval touches the distributed world in order to try to find relevant information contained in distributed servers. A user can't search the information in all these distributed servers. He needs a system which can do an automatic search for him.
A distributed information retrieval system [10] can do this work for the user, where the system receives the query of the user, then, and according to the query, it selects the pertinent servers and tries to find the needed information in these servers. Each server sends back a list of results.
The user can't process all the lists of results by himself, the system has to do this task for him.

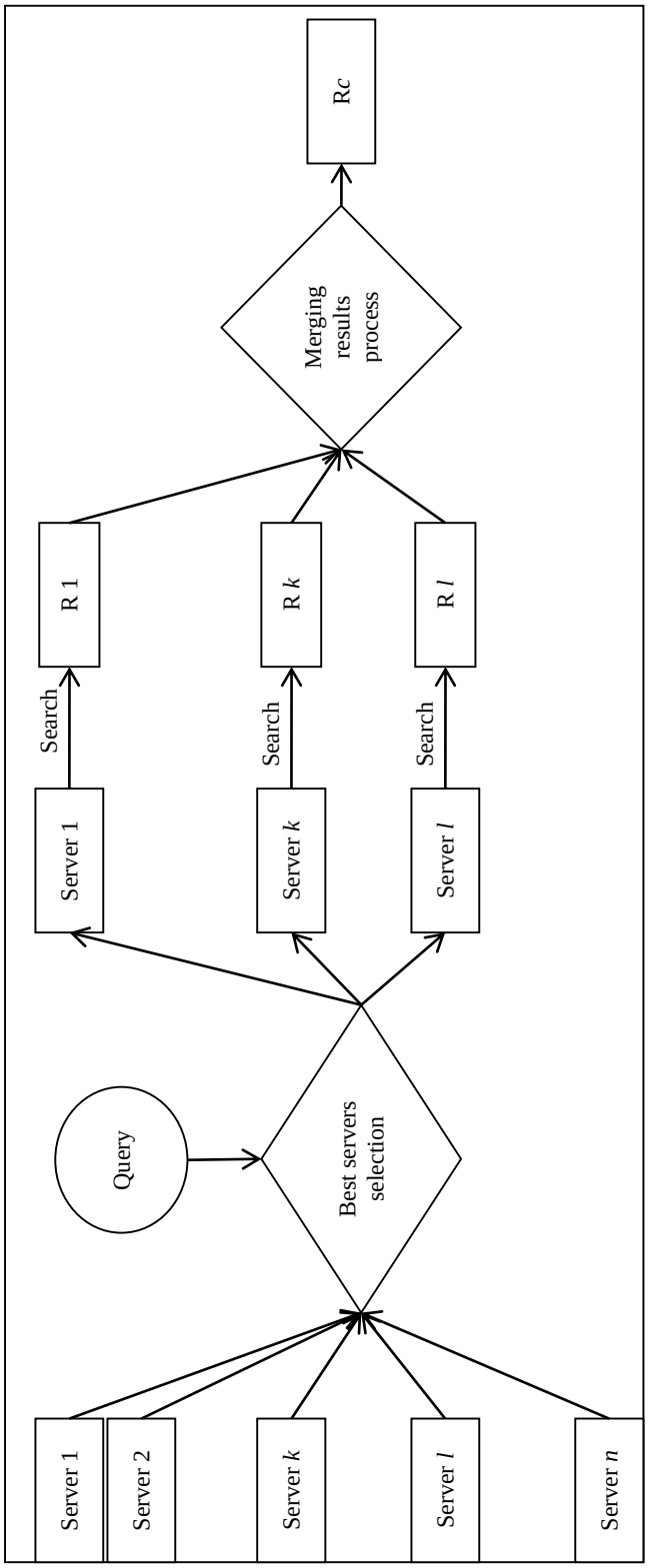

Figure 1. Information search process in distributed information retrieval system

Figure 1 shows the details of information search process in distributed information retrieval. In the first part, there is an important process called "Best servers selection". In this part, according to the query's content, the relevant servers will be selected.
The query will be sent to the selected servers in order to find relevant documents.
An important component of the system was the merging (or combining) results process. This module receives all the lists of results from selected servers and generates a combined list.
The system returns this combined list to the user who can use it easily.
In this work, we address issues pertaining to the merging results process in distributed information retrieval system.

1.3 Challenges

Merging results is challenging in particular in the context of a heterogeneous lists of results for several reasons.
The challenges are (1) make lists of results compatible, (2) sort merged results and (3) have a good relevance.
As we have said in the last section, the received lists of results from selected servers are not compatible because each server uses his own method of information retrieval. Therefore, merging and sorting results in one list is a problem.
We have to resolve this problem without losing the relevance.
The rest of the article is organized as follow. Section 2 discuses related work. We describe our proposals in section 3 and we report our experimental results in section 4. Finally, section 5 concludes the article.

2 RELATED WORK

Recently, there are a lot of researches realized in the domain of distributed information retrieval [3], [9], [11], [12], [13], [14], [15], [16], [17].
We focus our search on the last step of the process of distributed information retrieval systems which is the fusion of results.
The architecture of fusion process is shown in Figure 2.

Firstly, Roa et al.[11] have presented the results of the research carried out on the field of ranking strategies.
The analysis reported in their survey aims at providing a starting point towards future developments on benchmarking and empirical evaluation of ranking solutions for the Web of Linked Data.

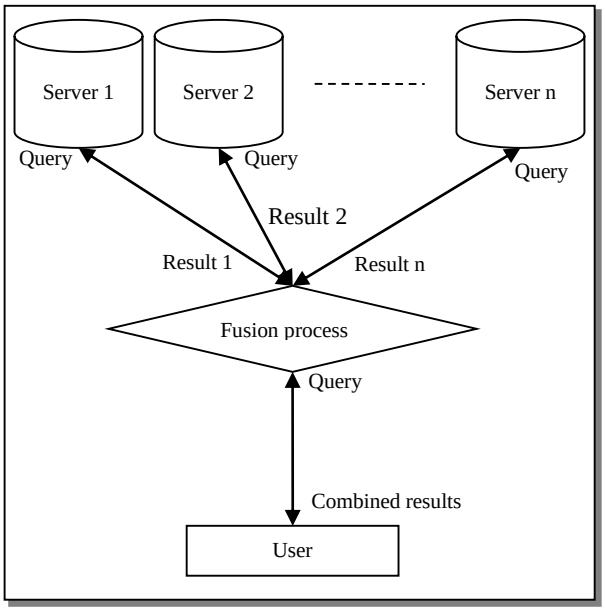

Figure 2. Architecture of fusion process

Dragut et al. [13] describe a system that integrates spatial objects from multiple local search engines.
They find the set of neighborhoods in a search engine that best approximates the area of a neighborhood in a different search engine.
Huo et al. [14] propose a fusion model to improve the tail queries by introducing new results from their reformulations. In our case, we have one query and several collections.
Savoy et al. [3] propose a method called RSM (Raw Score Merging) where they use local scores to sort results of different servers. They suppose that all servers use the same method to sort results and scores are compatible.
In our case, servers don't use the same method to sort results.
Other methods exist exploiting the heterogeneity of the sources of the documents. We can cite the CombSUM method and CombMNZ method.

Hubert et al. [18] propose CombSUM method which is based on the local scores obtained by different servers.
The score of the document D_j is calculated as follow:

$$ScoreCombSUM_j = \sum_{i=1}^{nbr_{servers}} score_{ij} \quad (1)$$

Hubert et al. [18] propose also CombMNZ method which is based on the local scores and the number of servers which have returned the query document in their results.
The score of the document D_j is calculated as follow:

$$ScoreCombMNZ_j = \left(\sum_{i=1}^{nbr\_servers} score_{ij}\right) \times Count_j \quad (2)$$

where $Count_j$ is the number of servers which have returned the document D_j.
This normalization isn't very important for our goal.
There is another kind of methods [3] based on the length of results of different servers. In this kind of methods, the local score of returned documents is weighted by the coefficient W_i calculated as follow:

$$W_i = 1 + \left(\frac{(S_i - S_m)}{S_m}\right) \quad (3)$$

where S_i is the score of the i^{th} server calculated as follow:

$$S_i = ln\left(1 + \left(\frac{L_i \cdot k}{\sum_{j=1}^{db} L_j}\right)\right) \quad (4)$$

where L_i is the number of documents returned pas the i^{th} server
and *db* is the number of interrogated servers.
These methods aren't pertinent for our case.
Rasolofo proposes in its thesis [19] the Round Robin approach where the merged list of results is constructed by taking alternately one document of each list returned by servers.
The disadvantage is that this method doesn't take into account the difference between the scores of the documents.
The CORI [20] algorithm is one of best used algorithms. The score of a document is calculated as follows:

$$D = \frac{D' + 0.4 \times D' \times C'_i}{1.4} \quad (5)$$

where:

$$D' = \frac{S - S_{min}}{S_{max} - S_{min}} \quad (6)$$

and S is the local score of a document;

S_{max} and S_{min} are the maximal and the minimal values of local scores of documents.

$$C'_i = \frac{C_i - C_{min}}{C_{max} - C_{min}} \quad (7)$$

and C_i is the relevance of the algorithm used by the i[th] server.

The final CORI score *D* is normalized between 0 and 1. According to the studies, the computational time of this method increases rapidly.

Yuwono and Lee propose a method called "Rank_based" [21] based on the conversion of ranks of documents in the list of results into a value of similarity.

This conversion isn't important for our case.

Rasolofo speaks about a method called Normalized Raw Score [19]. In this method the local scores are normalized by using maximal score of returned documents of all servers.

This normalization also isn't important for our case.

3 PROPOSED APPROACHES

In the distributed information retrieval we need to merge incompatible lists of results returned by several servers.

The problem of merging results is difficult because the lists of results are incompatible. They come from different information retrieval systems. Therefore, we can't gather all received lists of results in a single list without pretreatment.

To resolve this problem, we have to mathematically model the merge process which is the last step in the process of a distributed information retrieval system.

3.1 Problem Formulation

In the rest of this article we represent the merge process as a system defined as follows:

$$MP = <SR, CR, wf>$$

where:

MP is the Merge Process system;

SR is the Set of lists of Results returned by consulted servers (called: local results);

CR is the list of Combined Results (called: global results);

wf is the weight function used to make local results compatible.

In order to improve the relevance of the merging process and minimizing the calculation time, we had two ideas:

3.2 Approach Based on the Relevance of Local Information Retrieval Systems

The first idea is an approach based on the relevance of local information retrieval systems used by servers and local scores. We call it "M1" for (Method 1).

We define the weight function as the difference between the relevance of local information retrieval system and the maximal value of relevance of all used information retrieval systems. It is defined as follows:

$$wf(i) = (S_{max} - S_i) \times 100 \quad (8)$$

where:

wf(i) represents the weight of the i[th] server;

S_i is the relevance of the information retrieval system used by the i[th] server. It is restricted between 0 and 1;

S_{max} is the maximal relevance of information retrieval systems used in all known servers.

The constant 100 is used to make wf_i as a percentage since S_i is restricted between 0 and 1.

The algorithm 1 gives the details of calculation of the weight function.

Algorithm 1
Calculate $wf(i) = (S_{max} - S_i) \times 100$

Require: i
Ensure: $wf(i)$
Find S_{max}
return $(S_{max} - S_i) \times 100$

And then, the local results (*SR*) of documents returned by the i[th] server are weighted by *wf(i)*.

$$SR_{ij} = SR_{ij} \times wf(i) \quad (9)$$

Finally, documents selected by servers are sorted by their weights to build the *CR* list (global results).

The algorithm 2 shows how to combine the lists of local results in a single list.

Algorithm 2 Combine results

```
Require: SR, S
Ensure: CR
k ← 0
for I = 0 to nbServers do
for j = 0 to sizeof(SR[i]) do
CR[k] ← SR[i][j]*wf(i)
k++
end for
end for
Sort the table CR
return CR
```

3.3 Approach Based on the Relevance of Local Information Retrieval Systems and the Ranks of Selected Documents

The second idea is based on the relevance of local information retrieval systems used by servers and the ranks of selected documents in local results. We call it "M2" for (Method 2).

We define the weight function as the ratio between the relevance of local information retrieval system and the maximal value of relevance of all used information retrieval systems.

To make this weight dependent on the rank of documents, the ratio is multiplied by the complement of the rank of a document in its list of results.

The weight function is defined as follows:

$$wf(rank_i, i) = \left(rank_{max} - (rank - 1)\right) \times \frac{S_i}{S_{max}} \quad (10)$$

where:

$rank_i$ is the rank of a document in the local list of results *SR* of the ith server (ie. the rank in $SR[i] \geq 1$);

$wf(rank_i,i)$ is the new score of the $rank^{th}$ document of the ith server;

$rank_{max}$ is the maximal value of rank;

S_i is the relevance of the information retrieval system used by the ith server;

S_{max} is the maximal value of the relevance of all consulted servers $S_{max}=MAX(S_1,S_2,...,S_n)$.

We note that the new score $wf(rank_i,i)$ depends, at the same time, on the local ranks of documents in the local lists of results and the relevance of the information retrieval systems used by consulted servers.

The algorithm 3 gives the details of calculation of the weight function.

Algorithm 3 Calculate $wf(rank_i, rank_{max}, i, S)$

```
Require: rank_i, i, S, rank_max
Ensure: wf(rank_i, i)
Find S_max
return (rank_max − (rank − 1)) × S_i / S_max
```

Finally, documents with good scores are the closest to the query.

Using the new score $wf(rank_i,i)$, we try to make results provided by different servers compatible.

The algorithm 4 shows how to combine the lists of local results in a single list.

Algorithm 4 Combine results

```
Require: SR, S
Ensure: CR
k ← 0
for i = 0 to nbServers do
rank_max ← sizeof(SR[i])
for rank_i = 1 to sizeof(SR[i]) do
CR[k] ← wf(rank_i, rank_max, i, S)
k++
end for
end for
Sort the table CR
return CR
```

4. EXPERIMENTAL RESULTS

In this section we present results of performed experiments in order to validate the performances of our approaches.

We need to evaluate our approaches in real conditions by taking into account the factor of computational time and the sizes of the corpora.

4.1 Evaluation of Execution Time

In order to compare our propositions with existing approaches of fusion of results, we begin by calculating the execution time using the same corpora with these methods.

The following graphs show the evolution of execution time of developed methods.

The graphs of Figure 3 show that the fusion time of each method rises almost linearly depending on the number of servers. (Except the CORI method which rises exponentially).

We observe that our first proposition called "M1" presents a good execution time comparing to the other methods.
Our second proposition "M2" is situated between "CORI" and "Ranked based" methods which is a good result.

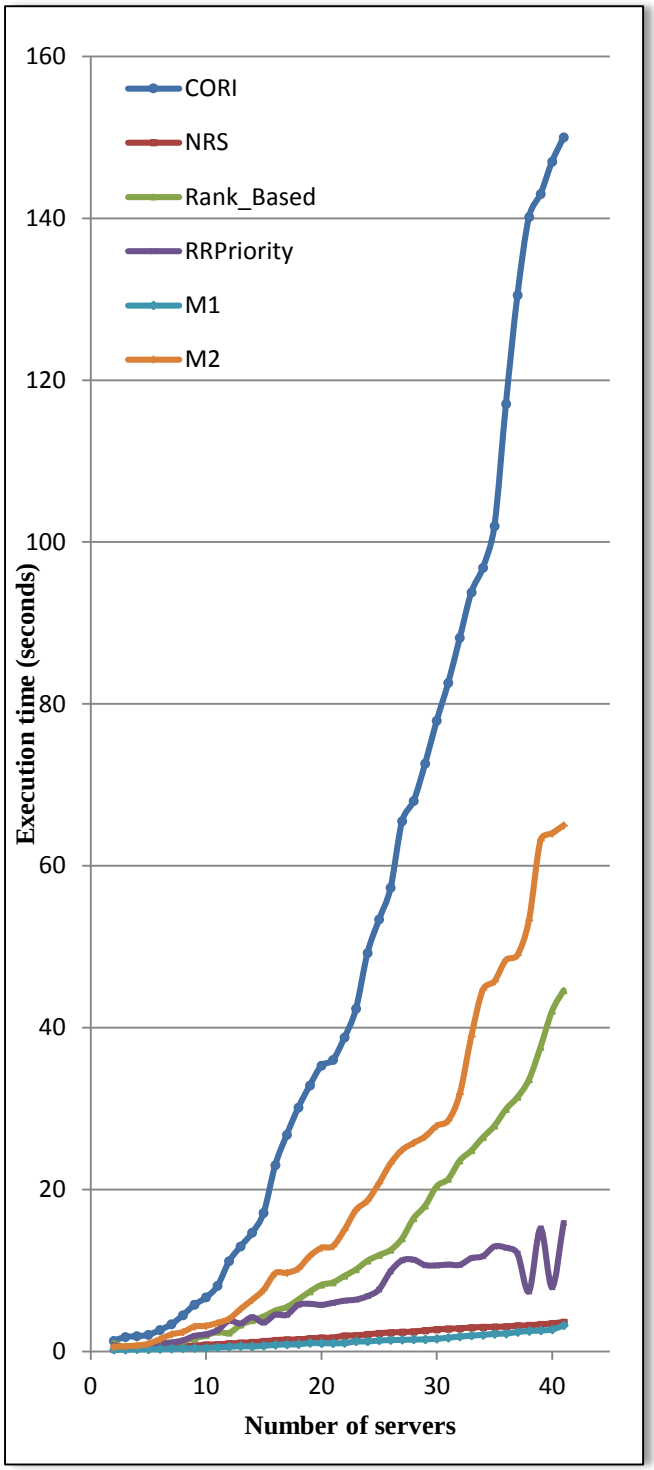

Figure 3 Evolution of computing time of fusion process according to the number of servers

4.2 Evaluation of the Relevance of Merge Process

The relevance of a method isn't based only on the computing time. For this, we propose to use another metric called the "Mean-Squared Error" (MSE).
Considerate that S is the number of documents sorted by a method of fusion. We propose to calculate the error rate of fusion process by the following formula:

$$MSE = \frac{1}{|S|} \times \sum_{i \in S} (Id_i - O_i)^2 \qquad (11)$$

In Eq. (11), Id_i is the rank of the ith document in the ideal sorted list of results; and O_i is the rank assigned to the ith document by the fusion method to evaluate.
In order to evaluate the relevance of implemented methods of fusion, we use the set of servers showed in Table 1, where sp_i is the relevance of the ith server.

Table 1. Results obtained by different servers

Server 1		**Server 2**		**Server 3**	
sp_1 = 90		**sp_2 = 70**		**sp_3 = 40**	
doc	**score**	**doc**	**score**	**doc**	**score**
LA123	65.5	FR453	87.54	FT567	87.54
LA673	57.8	FR012	75.5	FT195	51.64
LA946	35.7	FR673	11.84	FT548	40.9
LA765	19.81				
LA546	10.74				

Server 4		**Server 5**	
sp_1 = 35		**sp_2 = 60**	
doc	**score**	**doc**	**score**
DTR318	42.9	MHT217	90.43
DTR707	24.95	MHT232	15.54
DTR850	29.15	MHT305	22.56
DTR964	44.54	MHT471	13.07
DTR123	83.64		

After several tests with changing the number of servers, we obtained the results shown on the Table 2. Each column represents the sorted list of results obtained by one method.

Table 2. Comparison table of fusion results obtained by several methods

CORI Modified	NRS	RRPriority	Rank_Based
MHT217	LA123	LA123	LA123
DTR123	FR453	FR453	LA673
FT567	LA673	MHT217	LA946
FR453	MHT217	FT567	MHT217
LA123	FR012	DTR123	FR453
LA673	FT567	LA673	MHT305
FR012	DTR123	FR012	LA765
FT195	LA946	MHT305	DTR123
LA946	DTR964	FT195	DTR964
DTR964	DTR318	DTR964	FR012
DTR318	FT195	LA946	MHT232
FT548	LA765	FR673	FT567
DTR850	MHT305	MHT232	DTR318
LA765	DTR850	FT548	LA546
DTR707	MHT232	DTR318	FT195
MHT305	MHT471	LA765	DTR850
MHT232	DTR707	MHT471	FR673
LA546	FT548	DTR850	MHT471
MHT471	FR673	LA546	FT548
FR673	LA546	DTR707	DTR707

CORI Modified	M1	M2
MHT217	FR453	FR453
DTR123	LA123	LA123
FT567	MHT217	MHT217
FR453	FR012	FR012
LA123	LA673	LA673
LA673	FT567	FT567
FR012	LA946	DTR123
FT195	DTR123	LA946
LA946	FT195	FT195
DTR964	LA765	FT548
DTR318	FT548	DTR964
FT548	DTR964	LA765
DTR850	DTR318	DTR318
LA765	MHT305	MHT305
DTR707	DTR850	DTR850
MHT305	LA546	DTR707
MHT232	MHT232	MHT232
LA546	DTR707	LA546
MHT471	FR673	FR673
FR673	MHT471	MHT471

According to the prepared query, we find that results of "CORI modified" method are the best comparing to other methods. For this, we consider that results of "CORI modified" method represent the ideal list of results, and we compare the results of all other methods with those of "CORI modified".

On Table 2 we can see that the 8 first answers of our methods are almost equivalent to answers of "CORI modified" method. And we can see clearly that our two approaches are better than the other methods (NRS, RRPriority, Rank_Based).

After these tests, we can apply the Eq. (11) to calculate the error rate of each method. We found results shown in Table 3.

Table 3. Error Rates of Results Returned by Several Methods

Methods	MSE
CORI Modified	00.00
M1	05.80
M2	05.80
NRS	10.40
RRPriority	10.60
Rank_Based	13.50

We observe on Table 3 that the error rates of proposed methods "M1" and "M2" are better than those of NRS, RRPriority and Rank_Based methods.

Finally, according to these tests, we conclude that our approaches are good in computing time and in terms of relevance comparing to existing methods.

4.3 Discussions

Results obtained by performed experiments show that the calculation time of merge process rises almost linearly depending on the number of servers, except the "CORI modified" method which rises exponentially.

Graphs of Table 3 show that our first proposition called "M1" presents a good execution time comparing to the other methods. Our second proposition "M2" is situated between "CORI modified" and "Ranked based" methods which is a good result.

Results of the relevance tests show that the error rates of proposed methods "M1" and "M2" are

better than those of NRS, RRPriority and Rank_Based methods. The MSE of our proposed approaches is 5.8, but the MSE of NRS method is 10.4, the MSE of RRPriority method is 10.6 and that of Rank_Based method is 13.5.

Based on these tests, we conclude that our proposed approaches are good in computing time and in terms of relevance comparing to existing methods.

5 CONCLUSION

In our work, we were interested to the distributed information retrieval systems and especially to the last step in the process of these systems which is result's merging.

After studying different existing approaches of result's merging, and in order to ameliorate the performances of these methods, we proposed to add the relevance of local information retrieval systems of different servers.

By using the relevance of local information retrieval systems, the merging process provides good performance when we merge different lists of results obtained by different servers.

The highlight of this proposition is to taking into account the heterogeneity of local information retrieval systems.

To evaluate our propositions and to compare them with existing approaches, we used the computing time and the error rate of each method.

Our tests showed that proposed approaches "M1" and "M2" are better than existing methods: NRS, RRPriority and rank_based in terms of computing time - relevance ratio.

REFERENCES

1. Craswell, Nicholas Eric: Methods for Distributed Information Retrieval. 2000, The Australian National University.
2. Hassina, Aliane, Alimazighi Zaia, Boughacha Rime, and Djeliout Toufik. Un système de reformulation de requêtes pour la recherche d'information. Revue d'Information Scientifique et Technique (RIST), 2003. **13**(1): p. 25-33.
3. Savoy, Jacques, Yves Rasolofo, and Faiza Abbaci: Recherche d'informations dans des sources distribuées. In: *XIXème Congrés Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID)*. 2001: Martigny, Suisse. p. 237–252.
4. Si, Luo, Jamie Callan, Suleyman Cetintas, and Hao Yuan. An effective and efficient results merging strategy for multilingual information retrieval in federated search environments. Information Retrieval journal, 2008. **11**(1): p. 1-24.
5. Ming-Feng, Tsai, Wang Yu-Ting, and Chen Hsin-Hsi: A Study of Learning a Merge Model for Multilingual Information Retrieval. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2008. Singapore, Singapore: ACM.
6. Paltoglou, Georgios, Michail Salampasis, and Maria Satratzemi. A results merging algorithm for distributed information retrieval environments that combines regression methodologies with a selective download phase. Information Processing and Management, 2008. **44**(4): p. 1580-1599.
7. Abbaci, Faiza: Méthodes de sélection de collections dans un environnement de recherche d'information distribuée. 2003, Neuchâtel University.
8. Paltoglou, Georgios: Algorithms and strategies for source selection and results merging (collection fusion algorithms) in distributed information retrieval systems. In: *Department of applied informatics*. 2009, University of macedonia Thessaloniki.
9. Meng, Weiyi, Clement Yu, and M. Tamer Ozsu: Advanced Metasearch Engine Technology. 2010: Morgan & Claypool.
10. Zygmunt, Mazur. On a Model of Distributed Information Retrieval Systems Based on Thesauri. Information Processing Management journal, 1984. **20**(4): p. 499-505.
11. Antonio, J. Roa-Valverde and Sicilia Miguel-Angel. A Survey of Approaches for Ranking on the Web of Data. Information Retrieval journal, 2014. **17**(4): p. 295-325.
12. Alonso, Inostrosa-Psijas, Wainer Gabriel, Gil-Costa Veronica, and Marin Mauricio: DEVs Modeling of Large Scale Web Search Engines. In: Proceedings of the 2014 Winter Simulation Conference (WSC '14). 2014. Savannah, Georgia: IEEE Press.
13. Eduard, Constantin Dragut, Dasgupta Bhaskar, Beirne Brian P., Neyestani Ali, Atassi Badr, Yu Clement, and Meng Weiyi. Merging Query Results From Local Search Engines for Georeferenced Objects. ACM Trans. Web, 2014. **8**(4): p. 20:1-20:29.
14. Shuai, Huo, Zhang Min, Liu Yiqun, and Ma Shaoping: Improving Tail Query Performance by Fusion Model. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14). 2014. Shanghai, China: ACM.
15. Horatiu, Bota, Zhou Ke, Jose Joemon M., and Lalmas Mounia: Composite Retrieval of Heterogeneous Web Search. In: Proceedings of the 23rd International Conference on World Wide Web (WWW '14). 2014. Seoul, Korea: ACM.
16. Savoy, Jacques and Le Calvé: Recherche d'informations dans un environnement distribué. Cahier de recherche en informatique. 1996: Université de Neuchâtel, Faculté de

droit et des sciences économiques, Division économique et sociale.

17. Wu, Shengli and Jieyu Li: Merging Results from Overlapping Databases in Distributed Information Retrieval. In: 21st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP). 2013.
18. Hubert, Gilles, Yannick Loiseau, and Josiane Mothe: Etude de diffrentes fonctions de fusion de systmes de recherche d'information. In: *Le document numérique dans le monde de la science et de la recherche (CIDE'10)*. 2007: Nancy, France. p. 199–207.
19. Yves, Rasolofo Omega: Recherche d'information distribuée: approches pour la sélection de collections et la fusion de listes de résultats. 2002.
20. SI, LUO and JAMIE CALLAN. A semisupervised learning method to merge search engine results. ACM Trans. Inf. Syst., 2003. **21**(4): p. 457–491.
21. Yuwono, Budi and Dik Lun Lee: Server ranking for distributed text retrieval systems on the internet. In: Proceedings of the Fifth International Conference on Database Systems for Advanced Applications. 1997.

International Journal of Digital Information and Wireless Communications (IJDIWC) 6(2): 87-96
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

Ontology for Performance Measurement Indicators' Comparison

Vanderlei Freitas Junior[1], Victoria Uren[2], Christopher Brewster[2], Alexandre Leopoldo Gonçalves[3]
[1]Catarinense Federal Institute of Education, Science and Technology, Brazil; [2]Aston University, United Kingdom; [3]Federal University of Santa Catarina, Brazil
junior@ifc-sombrio.edu.br, v.uren@aston.ac.uk, C.A.BREWSTER@aston.ac.uk, a.l.goncalves@ufsc.br

ABSTRACT

Indicators are widely used by organizations as a way of evaluating, measuring and classifying organizational performance. As part of performance evaluation systems, indicators are often shared or compared across internal sectors or with other organizations. However, indicators can be vague and imprecise, and also can lack semantics, making comparisons with other indicators difficult. Thus, this paper presents a knowledge model based on an ontology that may be used to represent indicators semantically and generically, dealing with the imprecision and vagueness, and thus facilitating better comparison. Semantic technologies are shown to be suitable for this solution, so that it could be able to represent complex data involved in indicators comparison.

KEYWORDS

Performance measurement; semantic indicators; vagueness; imprecision; ontology; comparison.

1 INTRODUCTION

Performance measures have been recognized as important tools for organizational development [1]. Velimirovic, Velimirovic and Stankovic [2] state that continuous monitoring of performance measures has been the basis for improving organizational performance, and qualitative and quantitative expression of selected results. In other words, the authors state that performance measures enable organizations to effectively express their success through numbers.

The performance, in this context, can be understood as "sum of all processes that lead to a potential or future sequence of outcomes and results" [3]. Kaplan and Norton [4] indicate that performance can be expressed "only as a set of parameters or indicators that are complementary, and sometimes contradictory, and that describe the process through which the various types of outcome and results are achieved" [5], [4].

Thus, performance measures enable organizations to carry out a diagnosis of strengths and weaknesses, and to determine alignment with their goals by the analysis of indicators, which plays a central role in assessment processes.

Some authors, however, have pointed out difficulties in the use of indicators, since they lack semantics, and may be vague or imprecise in the statement of values that they are intended to represent [6], [7], [8], [9], [10], [11].

Pintzos, Matsas and Chryssolouris [6] have pointed out the difficulties and challenges posed by the lack of semantics in indicators, in particular the difficulty of comparing indicators and results internally, between sectors or departments of the organization itself, or between external agents. We will explore specific examples of this issue in the usage scenario presented in Section 4.

So, despite their importance in performance measurements, indicators have vagueness and imprecision, and lack semantics and these insufficiencies hinder their comparison. We will demonstrate that other information, for example, attributes and relationships between indicators, about these indicators is needed in order to allow their full comprehension and more accurate comparison.

Knowledge Engineering is important in order to face these challenges, which, according to Schreiber [12], "has evolved into the development of theory, methods and tools for developing knowledge-intensive applications. In other words, it provides guidance about when and how to apply particular knowledge representation techniques for solving particular problems". Therefore, the purpose of this paper is to present a model to address these challenges, that can represent indicators in a semantic way. This model has to deal with vagueness, imprecision and temporality, to

allow improved comparison between performance measurement indicators.

An important consideration about the use of semantic technologies to address the challenges of this study comes from the fact that a main aspect from our research problem is the need to compare two indicators. To make any comparison, it is important to know if the indicators are equivalent, if they have the same background.

Therefore, in this paper we present a new approach to performance measurement representation, based on an ontology, which increases semantic representation of indicators, in particular specifying background information about measurements, and thereby reducing vagueness and imprecision. This in turn supports more accurate comparison. Our proposal is general and reusable, so a wide range of domains could be represented by the proposed model.

The remainder of this paper is structured as follows. Section 2 provides some preliminaries on the fundamental theoretical aspects underlying this paper. In Section 3 we present our model. Section 4 demonstrates the application of the model in a usage scenario. Some conclusions are drawn and ideas for future research are proposed in Section 5.

2. PERFORMANCE MEASUREMENT

The daily work of organizations requires their efforts to face the challenges ahead. The success of organizations, then, is connected directly to the capability to assess its reality and the variables that affect it internally and externally. In this respect, it follows that the practice of evaluation, in the broadest sense, is part of human nature and underpins decision-making tasks [1] [13].

According to Neely, Platts and Gregory [14], performance measurement may be defined as a process of quantifying the efficiency and effectiveness of an action, which leads to performance.

So, performance measures have wide application in modern organizations, being understood as "qualitative, and quantitative expression of some results by chosen indicators. Performance measurement enable to effective organizations to express their success by numbers" (*sic*) [2]. In other words, it has the goal to translate organizational results into numbers and concepts that best express the satisfaction of the goals and objectives set by the company. Indicators play a central role in organizational performance assessment processes [7]. These authors also emphasize the role of indicators, stating that organizational performance measurement constitutes a development of measurable indicators that can be used to assess the progress toward a particular goal, as well as in achieving these goals.

2 SEMANTIC REPRESENTATION AND COMPARISON OF INDICATORS

As noted in the introduction, Pintzos, Matsas and Chryssolouris [6] have pointed out the difficulties caused by the lack of semantics in indicators. For these authors, the main characteristics that describe the context of an indicator are its name, calculation formula and measurement units. These attributes could situate an indicator, giving some meaning, allowing, as consequence, its comparison.

On the other hand, Opoku-Anokye and Tang [7] recognize that the task of setting indicators for organizational performance evaluation is multidisciplinary, since it is developed from various sources. This diversity often puts performance measurement primarily in the control of the functional units of organizations. Thus, the technical measures, models and approaches are developed with the same approach, i.e., based on functional and divisional structures of organizations, rather than semantics, business processes and product life cycle.

Pintzos, Matsas and Chryssolouris [6] present another limitation involving the lack of semantics in indicator development and performance measurement. According to the authors, the decision-making process in industry is based on performance requirements, which specify the values of the relevant production attributes. These values are expressed by internal performance indicators. The indicators may express the same metrics and use the same information, but have different definitions and calculation methods. Just to give an example, the "cost" indicator is often a monetary value and means an amount of resource needed to create some product or service. However,

the method of cost calculation could vary even in different sectors of the same company. This situation hinders performance comparisons, not only between different companies, but within the organization itself. In this context, the authors claim that semantic modeling is able to provide a broad overview of the data needed for a company to operate [6], facilitating the understanding and standardization of indicators.

Various tools and frameworks for performance measurement are available to guide planning, indicators definition, data collection and results monitoring. Among them, one of the best known and most referenced in the literature is the Balanced Scorecard (BSC). Developed by Kaplan and Norton [4], the BSC is a decision support tool at the level of strategic management, which supports improved satisfaction of an organization's strategic objectives, starting from appropriate planning and selection of indicators.

Despite the robustness of these tools, the literature indicates some difficulties to be addressed by academic research. Bobillo et al. [11] point out that, some variables or indicators are associated with some vagueness and imprecision, for example where it is more natural to refer to their values through linguistic expressions (e.g. low, medium, high) instead of numerical values. Furthermore, the authors also point out the difficulty caused by the lack of an explicit representation of their semantics.

This situation hinders performance comparison not only between different companies, but within the organization itself. In this context, Pintzos, Matsas and Chryssolouris [6] claim that semantic modeling is able to provide a broad overview of the data needed for a company to operate, facilitating the understanding and standardization of indicators.

It becomes clear that there exists a need for a generic semantic indicators model that would be sufficient to attenuate the vagueness and imprecision of indicator values by supplying background information on measurement conditions, allowing a more accurate comparison between indicators.

Bobillo et al. [11] also point out that the variables involved in planning and performance measurement may have inaccuracies and vagueness, which can interfere in results analysis and interpretation. On the other hand, they also claim that the data collected by the performance indicators and variables in the main methodologies are lacking in semantics, damaging the interpretation and analysis of these data. This difficulty makes the process of comparison between these indicators fragile.

The indicators, in summary, are widely used in organizations to assess institutional processes. However, the phenomena of imprecision and vagueness, and the lack of semantics, render their interpretation and comparison difficult.

3 SEMANTIC TECHNOLOGIES AND THE PROPOSED MODEL

Our proposed model is based on semantic technologies. These technologies were selected for three main reasons outlined below.

First, according to Belhadef, Eutamene and Kholadi1[15], "the concept of ontology is a concept that is not always easy to characterize. Indeed, it is used in different contexts: philosophy, linguistics, intelligence (AI), and each one's have its particular definition" (*sic*). But, in the context of this study, ontologies are considered as a well-established technology to represent knowledge in a specific domain. According to Gomez-Perez [16], an ontology can be used as a knowledge base skeleton, where inference processes or reasoning are executed. The use of ontologies for domain knowledge representation allows improvement in the information extraction process and the exchange of knowledge, which is one of the major motivations of this study. Gobin [17] reports that an ontology definition is based on the idea of conceptualization, i.e., a simplified version of the real world needing to be represented by providing "a shared and common understanding of a domain that can be communicated across people and application systems".

The second aspect of the use of semantic technology is the easy comprehension of the model. An ontology is used to represent consensual knowledge in a specific domain, and this aspect makes them easy to understand. Even users without prior knowledge of the technology or the domain could understand the meaning of all concepts expressed by the ontology. Gomez-Perez [16]

reports that clarity and objectivity are principles that have proved useful in development of ontologies. It "means that the ontology should provide the meaning of defined terms by providing objective definitions and also natural language documentation".

The third reason for the use of semantic technologies is the opportunities for reuse of existing components. The process of ontology design takes into account the possibility of the reutilization of some parts of other ontologies. Guarino [18] confirm that "an important benefit of using an ontology at development time is that it enables the developer to practice a "higher" level of reuse than is usually the case in software engineering (i.e. knowledge reuse instead of software reuse). Moreover, it enables the developer to reuse and share application domain knowledge using a common vocabulary across heterogeneous software platforms."

An important consideration about the use of semantic technologies to address the challenges of this study comes from the fact that a major aspect from our research problem is the need for comparison between two indicators. To make any comparison, it is important to know if the indicators are equivalent, if they have the same background, if they are using the same attributes, the same formulae etc.

Furthermore, Tang [19] indicate two advantages of the ontologies over databases, as a well-known techonology to represent and organize data. The author emphasizes that "as a comparison, database technology does not readily illustrate relationships among data entities. Another advantage of this approach over that of the database is that complex relationships between and among classes may not be so easily defined" in a database.

Thus, taking into account the scientific literature presented, we can identify the necessity to create a knowledge model for generic representation of indicators, considering the treatment of background information in order to allow the comparison between indicators with improved accuracy.

Knowledge Engineering is important in order to face the challenges described in this paper. It aims to provide systems capable of explicitly representing and storing the knowledge of the organization, considering all the systemic organizational context of knowledge intensive tasks [12].

Knowledge Engineering, therefore, provides instruments for knowledge-based systems modeling and developing that are able to explicitly formalize and represent knowledge for knowledge intensive tasks.

Taking into account all these tools, we highlight, within the limits of this work, the ontologies.

Ontologies are explicit specifications of the resource types and the possible relationships between them. In addition, they may include specific instances of concepts in the ontology [20].

An ontology can also be defined as a set of terms hierarchically ordered to represent a specific domain. It can be used as a knowledge base skeleton where inference processes are executed (reasoning). So, the use of an ontology allows the knowledge engineer to define a domain, allowing improvement in the information extraction process and the exchange of knowledge [21].

Another defnition widely accepted by ontology engineers [17] is authored by Gruber [22] who argues that ontology is an explicit specification of a conceptualization. Conceptualization is an abstract model of the world to be represented and that representation must explicitly specify the concepts, properties and relations. Borst [23] modified this definition, stating that an ontology is a formal specification of a shared conceptualization. He emphasized that there must exist a model in the specification of the ontology and the conceptualization should be done in such way to allow sharing. The expressed knowledge should be common sense and not particular to the person writing. Studer, Benjamins and Fensel [24] complement the definition asserting that an ontology is a formal, explicit specification of a shared conceptualization, reinforcing the requirement that the specification be explicit.

To Gobin [17], this definition is based on the idea of conceptualization, i.e., a simplified version of the real world represented by providing a common and shared vision of a domain that can be communicated between people and systems.

To Dillon and Simmons [25], ontologies support sharing a common understanding of the

information structure among people or software agents, but not only that, they also make it possible to reuse the given knowledge domain, and explain the assumptions, separating the operational knowledge from the domain knowledge, in addition to permit the analysis of domain knowledge. The ontologies engineering, in turn, is defined by the same authors as a highly collaborative process, since an ontology developed fairly accurately will be useless if it will not be accepted by domain experts, who should be directly involved in its development.

Bobillo *et al.* [11] state that ontologies allow data to be enriched with semantics, permitting automatic checking of data consistency, and giving an easier way to maintain the knowledge base and reuse of components.

Thus, assuming ontologies as tools for computational representation of specific knowledge of a domain, giving it meaning, ontologies can be used to provide semantics to indicators modeling.

In order to address the representation of performance measurement indicators, considering imprecision, vagueness, temporality and relationship between other indicators, with the aim of improving comparison of indicators, we propose a model based on ontology, as shown in Figure 1.

According to the literature, we can say that we have to know more about a given indicator, to understand its real meaning and the implications of its value. More complete knowledge about the origins of the data involved in the calculation of the value of an indicator could create the basis for a more accurate comparison between indicators, improving comprehension of its meaning and allowing the decision on what indicators could be compared and used for a decision-making process.

We also believe that we should know, at least, the following information about an indicator in order to facilitate improved comprehension and comparison:

- Entity interested in an indicator.
- Variables the indicator is related to.
- The description, the value and the importance (weight) of the indicator.
- What kind of criteria are used in a given domain to assess the indicator.
- What kind of formula or mathematical calculus is executed to calculate the indicator value.
- What other attributes compose the indicator value.
- Which point in time the indicator is related to.

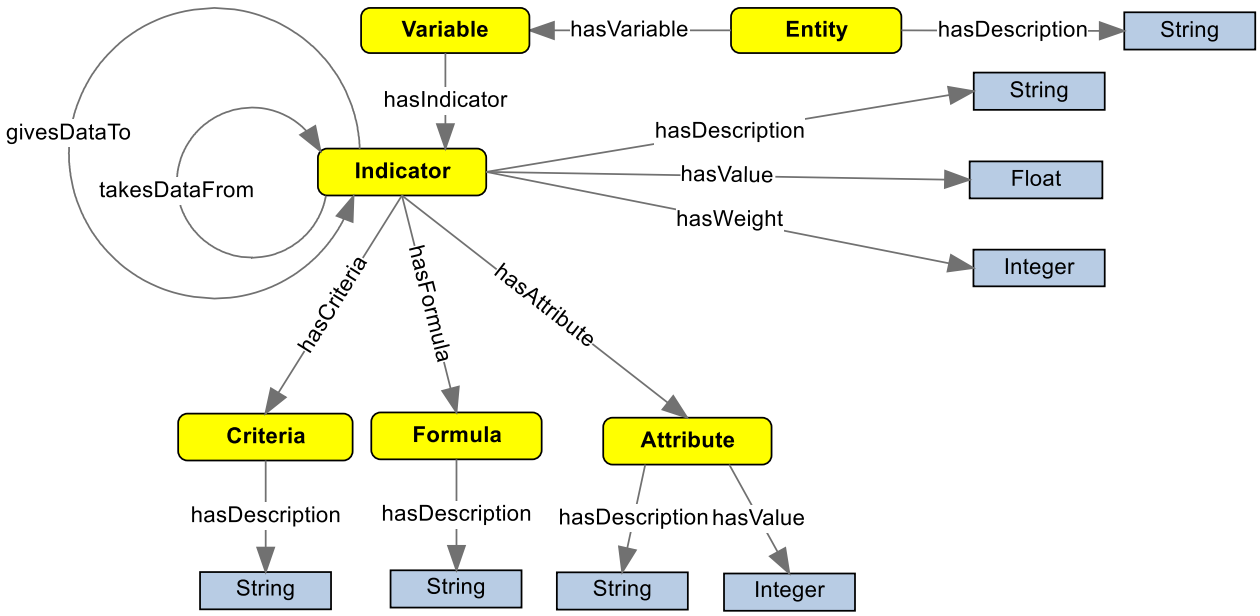

Figure 1. Proposed model

- What relations between other indicators need to be known.

All of these needs are represented by classes and properties in the proposed model, shown in Figure 1.

The ontology consists of six classes.

The "Entity" class is related to the "Variable" class by the "hasVariable" property in order to represent the relationship between an entity and the variable to be analyzed, which in turn, comprises one or more indicators.

The "Variable" class is related to the "Indicator" class through the "hasIndicator" property in order to represent the relationship between an indicator and the assessed variable referred to by this indicator.

The "Indicator" class, in its turn, has three properties. The first, "hasDescription" has the objective of providing a text description of the indicator. The second, called "hasValue" indicates the final value of the indicator calculation. Finally, the property "hasWeight", seeks to establish the importance of this indicator in an instantiated set of indicators by the ontology.

Also, in order to represent the relationships between indicators, the model implements the properties "givesDataTo" and "takesDataFrom" which indicate the links between indicators, their relationships and mainly represent the need to share data between them.

The "Indicator" class relates to the other three classes, which seek to ensure their meaning: "Criteria", "Formula" and "Attribute".

The "Criteria" class stores the necessary information for classification of represented indicators, from which you can set the sufficiency of the calculated values.

The "Formula" class tries to show the calculation methods of an indicator, ranging from simple arithmetic expressions such as sum or average, to complex mathematical formulas needed to calculate the numerical value represented by the indicator. This class is linked to the class indicator for the property "hasFormula".

The "Indicator" class is also related to the "Attribute" class through the property "has Attribute". This class has the objective of representing involved features and variables in the calculation of the modeled representation indicator value. These attributes are therefore part of the process of semantic enrichment of indicators. They allow the description of elements that complement and ensure meaning, and give the necessary data to determine indicator values. The "Attribute" class has two properties: "hasDescription" and "hasValue". The first is responsible for describing the feature to be modeled, and the second for assigning a value.

In the next section we will see an usage scenario of our proposed ontology.

4 USAGE SCENARIO

An example could be given to demonstrate the use of the proposed model in order to represent indicators, dealing with imprecision, vagueness and the relations between indicators, which shows how comparison is supported. A given company with operations across the country, needs to analyse its fleet fuel consumption, and needs this to be done car by car.

A report will show the fuel consumption of each car, telling that a car "A", located in the North of the country, had an average fuel consumption of 12 km/l. And a car "B", located in the South of the country, had a consumption of 13 km/l.

Now, in order to analyse this data, we have to make comparisons, and the question we have is: are both car equally economical? Is it possible to compare the consumption of car "A" and car "B"?

We have to know some more details about these indicators in order to make a comparison between these two values. All this information is represented in the ontology shown in Figure 2.

According to the specialists [26], a car's fuel consumption could be influenced by a lot of factors, like seasonal features, (e.g. temperature, wind), and the use of some components of the car, like air conditioning.

To establish an accurate comparison, variables must be defined by domain specialists or according to company interests. The ontology allows these adjustments, adapting to each situation. Thus, all of these variables must be analysed to determine if both indicators are equivalent and could be fairly compared.

Relations and information about the indicators are shown in the Figure 2.

In our case of comparison, we have two indicators that could be assessed as the same, but each one has its own particularities, as we can see in Figure 3.

The car "A" average consumption and the car "B" average consumption indicators has seven attributes that must be considered in order to be assessed: consumption, distance, location, average time of air conditioning use, average temperature and average wind speed and period.

Apparently, car "B" is more economical then car A, because it has travelled 130 km instead of the 120 km of car A, using 10 l of fuel. But we have to consider other variables involved in the representation of these indicators.

Car "A" has travelled 120 km using 10 l of fuel. It is located in North of the country, where the average temperature is around 35ºC, which demands 8 hours of use of air conditioning per day. The wind, in this part of the country, is around 10 km per hour.

On the other hand, car "B" has travelled 130 km using the same 10 l of fuel in a very different environment. This car is located in the South of the country, where the average temperature is 15ºC. In such temperatures, cars do not need to use air conditioning. The wind speed is around 20 km per hour in this part of the country at this period of the year.

So, as we can see, some factors that could increase a car's fuel consumption are affecting the indicators related to car "A", and apparently making it less economical then car "B". However, analysing all variables we could say that, despite of all the factors, car "A" is just 1 km/l, on average, less economical than car "B".

Dealing with all these variables, a lot of other calculus could be performed to achieve an accurate comparison, considering, for example, what is the impact of the use of air conditioning in the car consumption, or even what is the influence of the wind or temperature on the performance of

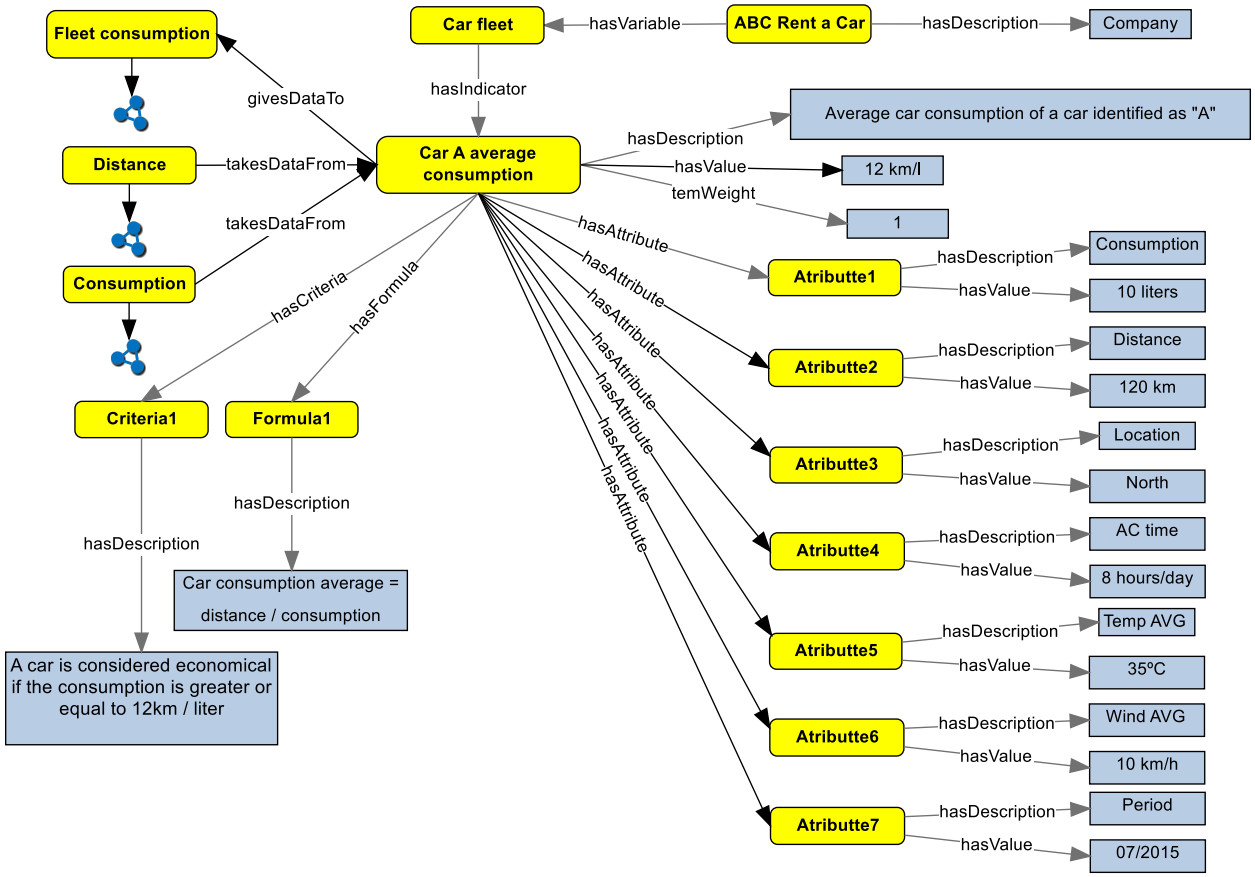

Figure 2. Car consumption ontology

the engine, and how the company cost is increasing because of these variables.

Figure 4 shows an another analyses that could be performed over our model. As we may see, some attributes were considered higher than in Car "B", others were considered equal or lower. Two attributes can be considered as just information in order to increase the semantic of the indicator. Analysing this attributes, we can say that in North of this given country, the temperature is higher in July then it´s in the South, what demands more time of air conditioning. In this case, as the cars have used the same amount of fuel, this difference of temperature and air conditioning can be reflected in lower covered distance by car A.

This usage scenario has demonstrated the importance of considering a set of information when making a comparison between indicators. We have demonstrated that ontologies can represent the background, increasing semantics and exactly situating the indicator in its domain, reducing vagueness and giving improved comprehension of the background of the measurements.

5 CONCLUSIONS AND FUTURE RESEARCH

In this paper, we proposed a semantic model to represent performance measurement indicators, based on an ontology, which addresses vagueness, and the lack of semantics concerning the background of measurements, in order to allow improved comparison between indicators.

The semantic technologies have been shown to be able to represent all the complex data involved in the comparison of indicators.

In the usage scenario proposed, we demonstrated that in a huge country like Brazil, with different seasonal conditions in different regions (or even in a comparison between the consumption of cars of the same company in different countries in Europe) the task of comparing indicators is non-trivial. It is important to take into account factors involved in the use and fuel consumption of the car. In the same country, or region, the same season may have different weather conditions, temperature or wind, for example. Such variables could affect fuel consumption, e.g. where a higher temperature could increase the regular consumption due to use of air conditioning. Background information is therefore needed about an indicator to be able to make comparisons, and semantic technologies are ready to address this sort of challenge.

The main contribution of this study is a new approach to indicators representation in order to support comparison. The model has the potential to allow computer-based comparison between two or more indicators. An interesting direction of further

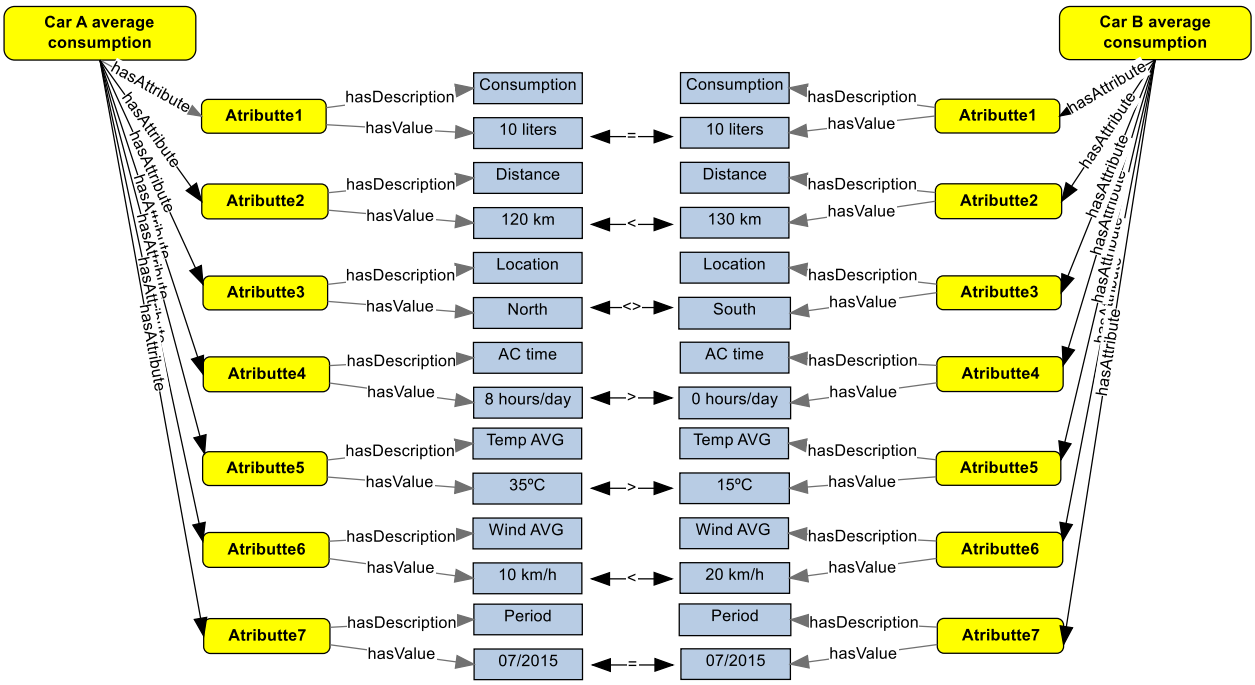

Figure 3. Comparison between attributes from two indicators

research would be to implement a knowledge-based system to simulate the use of the model in a real and computational situation.

As future research we propose two main aspects of the model: time and imprecision. We know that time plays a central role in indicators comparison, because a given indicator represents a specific point in time, or even a specific period of time. So, a model that could expand the representation of time, allowing different granularities, could be more effective in order to give a complete semantic representation of indicators.

On the other hand, dealing with imprecision could be a good way to improve representation of indicators. In our usage scenario, some questions could be answered, like: what is the economical level of car "A" related to car "B"? What is the impact of the variables in the consumptions of the car? Some linguistic variables could be associated to this analysis, like "high", "low", "better". The specific literature about fuzzy logic offer some answers to this questions, so a model that could represent this imprecision could improve the indicators' representation and increase the performance of the reasoning.

6 ACKNOWLEDGEMENTS

This work was partially supported by the Ministry of Education of Brazil. Vanderlei Freitas Junior holds a PDSE scholarship from CAPES Foundation, Ministry of Education of Brazil, Brasília - DF 70.040-020, Brazil.

7 REFERENCES

[1] Z. Jin, F. Deng, H. Li and M. Skitmore. "Practical framework for measuring performance of international construction firms". Journal of Construction Engineering and Management**.** n. 139, 2013, pp.1154-1167.

[2] G. Velimirovic, G. M. Velimirovic, and R. Stankovic. "Role and importance of key performance indicators

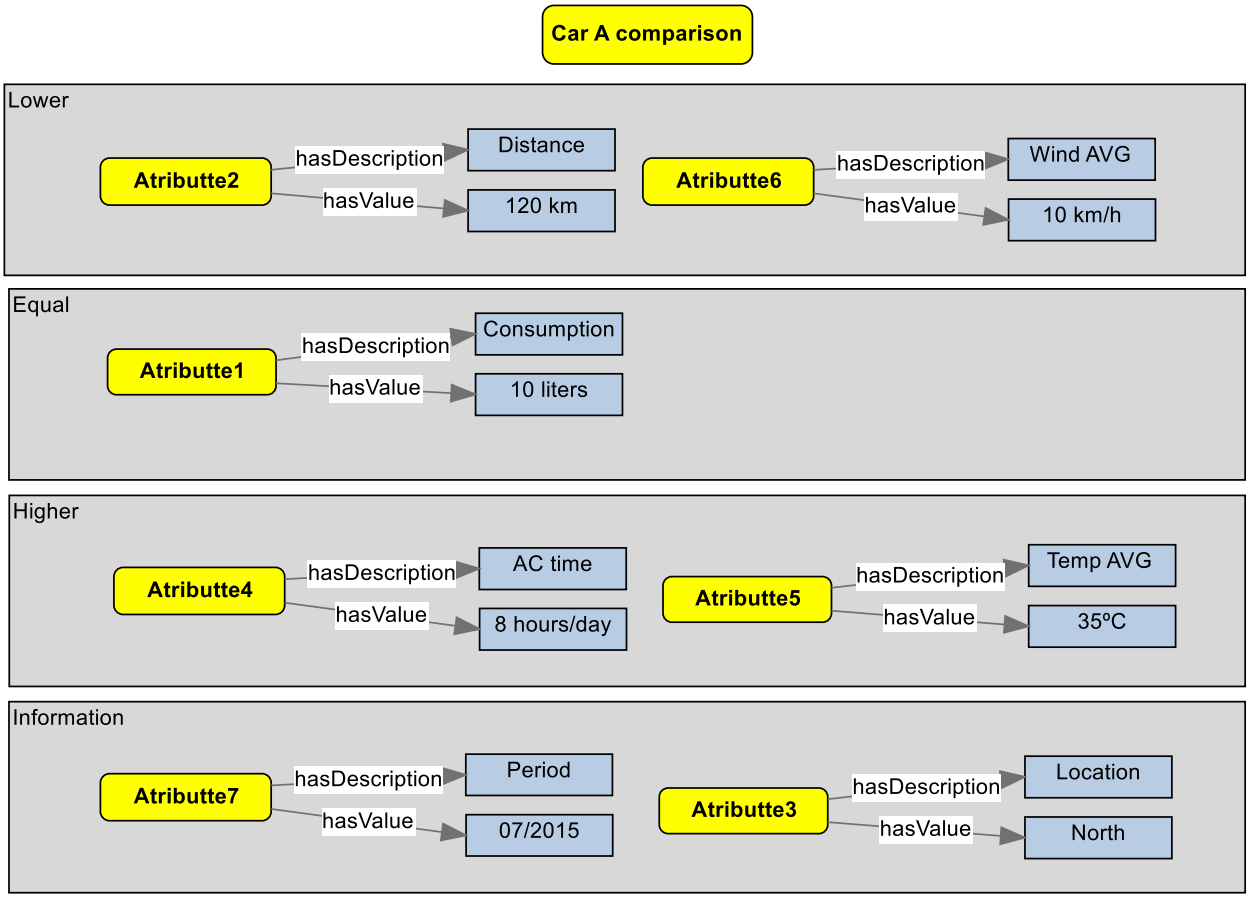

Figure 4. Car A comparison

measurement". Serbian Journal of Management**,** v. 6, n 1, 2011, pp. 63-72.

[3] A. Neely. Business Performance Measurement: Theory and Practice. Cambridge, United Kingdom: Cambridge University Press, 2002.

[4] R. S. Kaplan and D. P. Norton. "The balanced scorecard: Measures that drive performance". Harvard Business Review, Boston, MA, USA, v. 70, n. 1, 1992, pp. 71-79.

[5] M. J. Lebas. "Oui, il faut définir la performance". Revue Française de Comptabilité, 269, 1995, pp. 66-71.

[6] G. Pitzos; M. Matsas and G. Chryssolouris. "Defining manufacturing performance indicators using semantic ontology representation", in CIRP 3, 2012, Athens, Greece. Proceedings… Athens, Greece, v. 3, 2012, pp. 8-13.

[7] S. Opoku-Anokye and Y. Tang. "The design of a semantic-oriented organisational performance measurement system", in 14th Internacional Conference on Informatics and Semiotics in Organisation (ICISO)**,** Stockholm, Sweden, p. 45-49. 2013. Available from http://centaur.reading.ac .uk/31975 [Accessed 10th March 2014].

[8] F. Bobillo and M. Delgado. "Tractability of the Crisp Representations of Tractable Fuzzy Description Logics", in CEUR WORKSHOP. Proceedings**...** URSW, v. 654, 2010, pp. 109-112.

[9] M. Tavana; N. Mousavi and S. GOLARA. "A fuzzy-QFD approach to balanced scorecard using an analytic network process". International Journal of Information and Decision Sciences, v. 5, n. 4, 2013.

[10] Y. Shen; D. Ruan and E. Hermans. "Modeling qualitative data in data envelopment analysis for composite indicators". International Journal of System Assurance Engineering, v**.** 2, n. 1, 2011, pp. 21-30.

[11] F. Bobillo, M. Delgado, J. Gómez-Romero and E. López. "A Semantic Fuzzy Expert System for a Fuzzy Balanced Scorecard". Expert Systems Application, Lousiana, USA, v. 36, n. 1, 2009, pp. 423–433.

[12] G. Schereiber. "Knowledge Engineering", in Frank van Harmelen, Vladimir Lifschitz and Bruce Porter, editors. Handbook of Knowledge Representation (Foundations of Artificial Intelligence). Elsevier Science, 2008, pp. 929-946.

[13] O. M. Toledo; C. A. N. Cosenza. Metodologia de avaliação de desempenho baseada em lógica fuzzy. In: Congresso Brasileiro de Ensino de Engenharia Cobenge, 2004, Brasília. Proceedings... 2004.

[14] A. Neely, M. Gregory, K. Platts. Performance measurement system design: A literature review and research agenda. International Journal of Operations & Production Management, n. 25, p. 1228–1263. 1995.

[15] H. Belhadef, A. Eutamene, M. K. Kholadi. "Character Recognition Approach Based on Ontology". International Journal of Digital Information and Wireless Communications (IJDIWC), n. 1, v. 2, 2015, pp. 572-581.

[16] A. Gómez-Pérez. "Ontological engineering: a state of the art". British Computer Society, London, UK, v. 2, 1999, pp. 33-43.

[17] B. A. Gobin. "Reusing OWL-S to model knowledge intensive tasks performed by Knowledge Based Systems", in ICT and Knowledge Engineering, 2012, Bangkok, Thailand. Procedings**…** Piscataway, NJ, USA: IEEE, 2011, pp. 34-42.

[18] N. Guarino. "Formal ontology and information systems", in International Conference on Formal Ontology in Information Systems - FOIS´98, 1998, Trento, Italy. Proceedings**…** Amsterdam: IOS Press, 1998, pp. 3-15.

[19] A. Tang. "Towards the Use of a Computing Ontology: A Curricula Management System Proposal". International Journal of Digital Information and Wireless Communications (IJDIWC), v. 1, n. 1, 2015, pp. 94-108.

[20] A. Abramovich. Human experience representation via upper level ontologies. In: International Conference on Integration of Knowledge Intensive Multi-Agent Systems. Procedings of IEEE. Piscataway, NJ, USA: IEEE, 2005. p. 187-190.

[21] A. Gómez-Pérez, A. Ontological engineering: a state of the art. British Computer Society, Londres, Inglaterra, v. 2, p. 33 – 43, 1999.

[22] T. R. Gruber. A Translation Approach to Portable Ontology Specification. Knowledge Acquisition, v. 52, n. 6, p. 1111-1133. 1993.

[23] W. N. Borst. Construction of engineering ontologies for Knowledge sharing and reuse. 1997. 243 f. Tese (Doutorado em Telemática e Tecnologia da Informação) - Universidade de Twenty, Enschede, Holanda, 1997.

[24] R. Studer; R. Benjamins; D. Fensel. Knowledge Engineering: Principles and Methods. Data & Knowledge Engineering, v. 25, n. 1-2, p. 161-197. 1998.

[25] T. S. Dillon; G. Simmons. Semantic Web Support for Open-Source Software Development. In: Signal Image Technology and Internet Based Systems, 2008, Bali. Proceedings… Piscataway, NF, USA: IEEE, 2008. nov./dez. p. 606-613.

[26] INMETRO. "Metodologia para divulgação de dados de consumo veicular. Methodology for vehicle consumer data dissemination". Available from http://www.inmetro.gov.br/consumidor/pbe/Metodologia_Consumo_Veicular.pdf [Accessed 24th August 2015.

International Journal of Digital Information and Wireless Communications (IJDIWC) 6(2): 97-111
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

A Case Based Reasoning Architecture and Component Based Model for Green IS Implementation and Diffusion in Organisation

Bokolo Anthony Jnr. and Noraini Che Pa
Department of Software Engineering and Information Systems
Faculty of Computer Science and Information Technology
University Putra Malaysia (UPM)
43400 Serdang, Selangor, Malaysia
result4real@yahoo.com and norainip@upm.edu.my

ABSTRACT

Green Information Systems (IS) offers promise for IS scholars to make significant contribution in reducing greenhouse gas emissions, achieve sustainability and reducing the effects of global climate change. Green IS has therefore become one of the latest considerations to improve organization's environmental sustainability whilst reducing the cost of IT processes. Due to recent research centered on sustainability, a few number of Green IS models and frameworks have been developed for addressing environmental sustainability, however there is still problem relating to provision of adequate data that can train, educate and support practitioner in diffusing and implementing Green IS in organisations. Thus after reviewing existing research of Green IS, this paper proposed a Case Based Reasoning (CBR) architecture for Green IS Implementation in organisations to train, educate and support practitioner in implementing Green IS in their organisation and a Component Based Model to support Practitioners in diffusing and achieving Eco-Sustainability goal. Contribution from this paper can also be used as a guide for future research in this domain.

KEYWORDS

Case Based Reasoning, Green IS, Green IT, Sustainability, Green IS Components.

1 INTRODUCTION

The society and organisations are entirely dependent on the environment for existence. In spite of this crucial dependency, human activities are degrading the environment. The deprivation now threatens the wellbeing of all people and possibly the long-term survival of the human race.
It is argued that establishing environmental sustainability is one of the most important global challenges of the 21st century and that the information systems discipline has both a obligation and an prospect to contribute to solving this challenge. On the one hand, IT contributes approximately 2% of global greenhouse gas emission. Reducing these effects through more energy-efficient systems with decreased environmental impact is the domain of Green IT.
More broadly, Green IS is suggested as a means to also resolve the remaining 98%. Green IS has evolved as a subfield of the information systems (IS) discipline investigating how IS can support sustainability. It contrasts from Green Information technology (IT) in the sense that it's not limited to the energy efficiency and equipment utilization of IT, but focuses on the design and implementation of information systems that contribute to the implementation of sustainable business processes and also enhance business productivity. Green IS is considered to have high transformative power, as it addresses the revolution of socio-technical systems in a number of application areas [1].
Information Systems (IS) were the ultimate force to improve productivity in the last decade. Currently, organizations have more opportunities to benefit from sustainable development while improving productivity, reduce costs and increase profitability. However, practitioners poor environmental practices results in many types of waste; Balances energy inefficiency, noise,

friction, and the emission of all products, waste subtracted from economic efficiency. Such poor environmental practices could be improved with the implementation and diffusion of Green and sustainable practices in their business processes. Presently in IS Community sustainability has become an important issue in the strict sense of reducing the direct environmental impacts of IT use. Energy efficiency in hardware and data centers continues to receive a great deal of research, given the potential to decrease emissions and lessen energy costs [2]. Recycling of computer hardware, network infrastructure and electronic waste is also a necessity for enterprises to implement a sustainable practice. However IT based firms such as Dell and Toshiba promote environmentally-friendly designs emphasizing reusability and elimination of toxic chemicals, while IT service enterprise offer green IT consultation to end users.

Green IS can therefore minimize the total cost of technology usage and increase the total environmental value. Therefore Green IS approaches support business's sustainability initiatives, examples of such initiatives include the deployment of analytical tools and IS that support dynamic routing of automobiles to reduce energy consumption, the execution of emission management systems and the supplanting of carbon emitting business practices through videoconferencing and other online collaborative facilities. As a result, the implementation and diffusion of Green IS can be considered as a critical factor not only for the sustainability of businesses but also the success of the low carbon economy [3].

In diffusing and implementing Green IS practice organisations are committed to reducing the concentration of harmful substances, saving the use of natural resources, creating a safe working environment to make their products and service compliant with applicable laws and regulations. Thus in diffusing and implementing sustainable practices each year, organisations not only reduce the carbon-dioxide emissions but also save the material, energy, and disposal costs [4]. IS includes information technology (IT) hardware and software, people, processes, and have specific purposes. Green IS focuses mainly on environmental sustainability all over the IT lifecycle, and aim to enable and transform entire organisations toward environmental sustainability hub. Green IS also presents a fundamental solution to organizational environmental degradation [5]. But currently practitioners are faced with various issues when implementing and diffusing sustainable practices in their organisation, such as insufficient information to provide support on how to implement Green practices. According to [6] there is lack of information for making decision relating to sustainability diffusion in enterprise, thus enterprise need to integrate data from multiple information systems to support their multiple business processes and attain sustainability. But the availability of information can be useful to practitioners in various ways, such as practitioners gaining access to cost and energy systems working in real-time to provide support to practitioners to measure the cost and energy utilized in their business.

[7] mentioned that one of the problems of attaining an environmental sustainability is lack of information to enable and motivate economic and behaviorally driven solutions. [8] stated that in an IS based Green supply chain management (SCM) there is need for an enterprise system which collects, restore, coordinate, classify and forecast the information resources, that is crucial to provide support for the practitioners in attaining sustainability. This is based on the development and the change of the facing environment, in which practitioners will need a support system.

Thus this paper contributes to environmental sustainability by synthesizing the existing Green IS frameworks and models into a single component based model that is purposefully aligned to environmental sustainability. Thus this paper is theoretical and exploratory in nature and makes an original contribution by providing a synthesized Green IS component based model and a case based reasoning architecture aligned to the concept of environmental sustainability to provide adequate data to train, educate and support practitioner in diffusing and implementing Green practices in organisations.

The structure of this paper is organized as follows: Section 2 is literature review, section 3 is the methodology, section 4 is case based reasoning

architecture, section 5 is Green IS component based model, section 6 is discussion and finally, section 7 is the conclusion and future works section.

2 LITERATURE REVIEW

This section lays a theoretical outline of Sustainability, Green IS, Green IT and CBR.

2.1 Information System for achieving Sustainability in Organisation

Sustainability means meeting the needs of the present without compromising the ability of future generations to meet their own needs. [9] stated that for an enterprise to achieve sustainability triple bottom-line, also known as the *3Ps; planet, people, and profit* must be considered by the management in the enterprise. IS also considers the 3Ps as the major force driving growth and productivity in the last decade has a critical role in creating a sustainable organizational business practice. A global UN survey to determine the issues controlling the future identified sustainable economic development as the dominant issue. The report notes, never before has world opinion been united on a single goal as it is on attaining sustainable development.

Currently humans' consumption of the earth's resources is unsustainable and is creating major environment problems. Climate change, resource depletion, loss of biodiversity, and air pollution has a major impact on many citizens and the earth, thus organisation and the society need to change the current behavior, if we want to achieve sustainability. The present consumption of the earth's finite resources cannot be maintained. Therefore there is need to move to sustainable development, which meets the needs of the present without compromising the ability of future generations to meet their own needs [10].

At the moment, the disposal of equipment by organisations is a major environmental problem because of the toxic products in computers hardware. However, IS has been the major contributor to productivity growth in many countries over the last half century. IT is also needed to deploy the IS that will support sustainable business practices. Organisations are linking sustainability to their corporate strategy. They recognize that they have key responsibility to participate in solving this critical global problem and end users expect them to provide Green products and services. Sustainability requires sustainable business practices because of the dominant role of organisations in the global economy. Thus it is fair to conclude that IS is a major element in the transition to a sustainable economy [10].

2.2 Green IS and Green IT in Organisation

IT specifically focuses on computer devices that store, transmit, or process information, IS is more encompassing, as it includes the integrated and cooperating set of people, processes, software, and information technologies to support individual, organizational, or societal goals [11]. IT and IS contribute to the environmental issues differently, For the most part, it appears that IT is part of the problem; such as data centre's energy consumption, electronic waste. While IS can be part of the solution; such environmental management systems, telecommuting programs [9]. Table 1 shows the role of Green IT and Green IS in organisations;

Table 1. Role of Green IT and IS in Organisation

Green IT	Green IS
Green IT is mainly focused on energy efficiency and equipment utilization [10].	Green IS refers to the design and implementation of information systems that contribute to sustainable business processes [10].
Green IT is perceived as focusing on minimizing the harm of technology to the environment [12].	Green IS may be viewed as an enabler of environmental sustainability [12].
Is the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems such as monitors, printers, storage devices, and networking and communications systems efficiently and effectively with minimal or no impact on the environment [13].	It's the conversion of physical (individual, business, community) activities to more efficient digital alternatives to reduce their carbon footprint, and on how information can be leveraged and managed to enable sustainability [9].

	Refers to the design and implementation of systems that contribute to sustainable business processes or enables environmental sustainability [9].
	Green IS tends to be less direct in its impact. It focuses on designing and initiating systems to enable environmental management and includes the concept of Green values enacted during design and implementation [12].

Reducing the energy consumption of data centers.	Track environmental information (such as toxicity, energy used, water used, etc.) about the creation of products, their components, and the fulfillment of services.
Promotes usage of renewable energy sources to power data centers.	Monitor enterprise's operational emissions and waste products to manage them more effectively.
Reduces electronic waste from obsolete computing equipment.	Provides information to stakeholders, so they can make green choices more conveniently and effectively.

From Table 1 it can be seen that Green IT involves lowering of energy consumption and cost reduction within data centres but Green IS takes the scope a step broader, by incorporating the possibility of entire systems, viewed holistically and thus potentially able to offer more comprehensive solutions for environmental sustainability. Presently IT and IS are often at the forefront of managerial practice, in supporting sustainable economic development in organisation. Table 2 shows the contributions of Green IT and IS presently in organisations.

Table 2. Contributions of Green IT and IS in Organisation [10]

Green IT	Green IS
Promotes the design of energy efficient chips and disk drives.	Reduce transportation costs with a fleet management system and dynamic routing of automobiles to avoid traffic congestion and reduce energy consumption.
Replacing personal computers with energy efficient thin clients.	Support team work and meetings when employees are distributed throughout the world, and thus reduce the impact of air travel.
Promotes telecommuting and remote computer administration to reduce transportation emissions.	IS can move remote working beyond telecommuting to include systems that support collaboration, group document management, cooperative knowledge management etc.

Table 2 shows that information technology (IT) supports organisations to transmits, processes, or stores information and information system (IS) is responsible for the integration and collaboration of software, using information technologies to support individual, group, organizational, or societal goals.
Similarly, Table 2 shows that Green IT mainly focused on energy efficiency and equipment utilization and Green IS, in contrast, refers to the design and implementation of information systems that contribute to sustainable business processes [10].

2.3 Overview of Case Based Reasoning (CBR)

Case-based reasoning (CBR) is the process of solving new problems based on the solutions of similar past problems, by retrieving similar cases from a case base. Hence, practitioners can reuse and revise solutions from similar cases and obtain the final confirmed solution for the new problem [14].
Research studies have utilized CBR in various applications in architectural design and Green application management, such as estimation of energy consumption, cost and duration estimation of software projects.
CBR is a problem-solving and continuous learning methodology. In CBR, new solutions dealing with a current situational circumstance are generated by

retrieving the most similar cases from the case base and familiarizing and reusing them in the present circumstances. The CBR technique can be explained as a four-step process as shown in Figure 1;

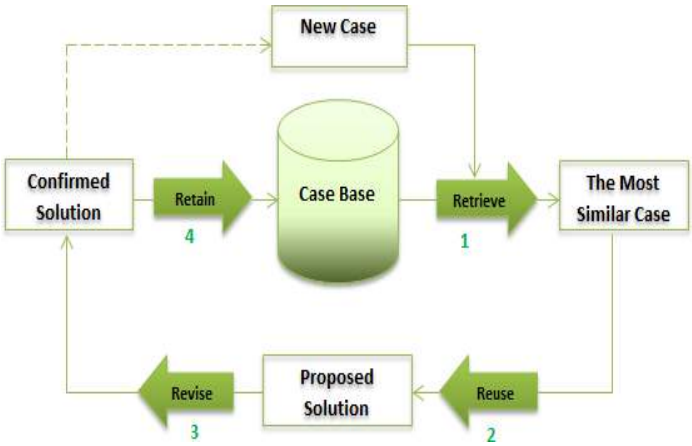

Figure 2. Case Based Reasoning (CBR) System Adopted from [14]

Figure 1 shows the CBR system's composed of four steps: *retrieve, reuse, revise and retain*, which are termed by many researchers as the CBR cycle. CBR is outstanding at dealing with real world, open-ended, difficult problems (such as Green IS implementation). Thus CBR should be used whenever possible, because it is often the fastest approach for gaining a adequate solution among the available techniques [15]. Below are the CBR phases;

i. **Retrieve:** Given a target problem or a new case, retrieve from the case base to get the most relevant and similar cases. Usually, a case consists of a problem description and its solution. Thus this phase mainly involves the retrieval of one or several cases that contain solutions that can be reused in the current problem solving context. Since it is usually unlikely that the case base already has a problem that matches the new one exactly, the concept of similarity is normally used to execute the retrieval task, CBR technique employ similarity measures that allow computation of the similarity between two problem descriptions. In general, similarity can be divided into three types: (1) surface similarity, derived similarity, and structural similarity.

   a. ***Surface similarity*** is mostly the similarity of the characteristic that are provided as part of the description of the case, such as the height of a material, the cost of a implementing green practice in a project, etc.

   b. ***Derived similarity*** is the similarity of derived capabilities, which are obtained from the description of the case by inference, based on domain knowledge. For example, in the mobile field, to find a similar mobile device, the area of a screen may be more significant than simply its length and width.

   c. ***Structural similarity*** is the similarity of the characteristic of the object. This similarity is used at where cases need to be presented in complex configurations, such as charts or histogram. In these cases, the retrieval step requires a measurement or a learning procedure of the structure of the cases first and then computes the characteristic similarity.

ii. **Reuse:** Plots the solution from the retrieved cases to the target problem and offer a proposed solution. This may involve adapting the solution, as needed, to fit the new situation. This second step of the CBR leads to proposing a solution for a new problem from the solutions in the retrieved cases. In most situations, individuals will simply use replacement or conversion. In some special cases, the reuse step is modified to assist in a better retrieval.

iii. **Revise:** Test the suggested solution in the real world or in a model and, if necessary, revise in order to have the confirmed solution. The third step, revise, is highly user based. In some cases, the revise step is integrated into the reuse step. The retain step is the last step in the CBR cycle. In many Internet-based systems, such as electronic-commence endorsement systems, in order to maintain a high quality case base which not only maintains an acceptable diversity but also reduces the

number of cases and thus reduces the computation time, different methods have been recommended to selectively retain and even replace the case in the original case base.

iv. **Retain:** After the solution has been successfully applied to the target problem, the solution is store and the resulting outcome is a new case in the case base.

2.4 Existing Green IS Models and Frameworks

Green IS has gain reliance for the past decade, researchers, academicians and organisations are concerned on how to utilize information system to reduce the impact of CO2 generated from the use age of IT.

In view of these issues [4] designed a conceptual model of reasoning that is based on the variables for Green Supply chain and the impact of Green supply chain. The researchers used the model to test its influence on firm's Business Process Implementation (BPI), business innovation and value added performance. The model aims at reducing cost, expanding business and enhances globalization based on Green IS.

[16] developed a practice oriented Green IS framework aimed at lowering organisational Greenhouse Gas. The framework consists of six components. The component includes business and IS strategy that determines the people, dematerialisation; energy efficiency, Green operations categories and waste and recycling. [6] proposed an entity centric approach to Green IS to help organisations in forming a cohesive representation of the environmental impact of their business operations at both micro and macro levels. Their approach focuses on the concepts that define particular areas of interest in an organisation, for example, business entities like employees, products, customers, intellectual property, assets, etc.

[5] synthesised existing Green IS framework and proposed an IS framework that purposely aligned to strong environmental sustainability. The proposed framework comprises of motivating forces, Organisation (People and IS) and Strong environmental sustainability. The framework serves as a guide for further empirical research; and as a prescriptive for informing management practice. In practice, the framework provides main focal points for management in the pursuit of environmental sustainability.

[2] contributed to Green IS research by presenting a model of IT investment and carbon productivity. The model comprises of Government policy and individual behaviour that are dependent on the inputs of the labour, Capital, Energy and GHG emission of the people involved in the organisation, investments in production process implemented by the people and value added output that is based on the demand pattern of the government and individual behaviour.

[8] develops a three layers enterprise application information system integration model based on the green supply chain management (SCM) combining with the relationship of enterprise application information systems, and describes the function and relation of each layer to help enterprise realize the sustainable development, through the green grade barriers and heighten the competition ability. The model assists SCM in help enterprises facing the serious marker competition and economic crisis and enhances the ability to face the pressure of environment and energy.

[10] demonstrate how the transformative power of IS can be leveraged to create an ecologically sustainable society by suggesting a new subfield of IS, energy informatics, reducing energy consumption, and thus CO2 emissions. This can be expressed as *Energy + Information < Energy*. The concept for Energy informatics is concerned with analysing, designing, and implementing systems to increase the efficiency of energy demand and supply systems.

This requires collection and analysis of energy data sets to support optimization of energy distribution and consumption networks. The researchers proceeded to develop an energy informatics framework which comprises of the stakeholders (consumers, Suppliers and Governments), Eco-goals (Eco-efficiency, Eco effectiveness and Eco-equity) which are influenced by the following attributes; policies & regulations, economics and corporate norms. These attributes affects the supply and demand in the organisations information system.

3 METHODOLOGY

The objective of this section is to explain how the chosen research methodology addresses the research problem. The paper assumes constructivism, which refers to inquiry about a domain that is dependent on human perception and to epistemology where knowledge about the world is gained subjectively [5]. Such a basis fits the review of existing literatures for interpretation and explaining social phenomena using existing data. Literature review is also appropriate for IS research focusing on managerial and organisational issues. The paper is theoretical and exploratory in nature, and conducts a research synthesis on relevant and quality literature retained from a systematic literature search strategy. Research syntheses make a valuable contribution to the academic body of knowledge, and further academic discourse by uncovering meaningful, abstract features and patterns, and by providing new theories, model or frameworks. The research synthesis begins by analysing the existing Green IS model and frameworks' purposes and characteristics. The frameworks or models purposes and characteristics are then iteratively distilled to extract the essential components and attributes related to achieving sustainability in organisation.

3.1 Impact of Information Technology (IT) on the Environment

Researchers in sustainability have framed the impacts of IT on the environment as first, second, and third-order effects. First-order effects are direct impacts from IT hardware during the product lifecycle, including production, use and dumping of computer equipment. These effects are similar to the scope of Green IT. Second-order effects are the effects of ICTs on other processes such as transportation or manufacturing production, influencing their environmental impacts. This is similar to Green IS. Third-order effects are longer term. They occur when widespread use of ICTs leads to changes in routines and economic structures. Third-order effects may be seen in specific cases, such as telecommuting or the growth of domestic businesses built on e-commerce platforms such as Alibaba. So far, these effects are limited, but widespread adoption of the internet is only a decade or less old in many countries [2].

3.2 Green Information Systems (IS) Process

In Green IS process, management have to work closely with practitioners and end users to effectively manage the environment. Figure 2 shows the summarized Green IS process.

Figure 2. Green IS Process [4]

i. Green Design

Refers to the use of natural, organic, and recyclable material, which aims for zero waste designs; that is making old outputs new inputs.

ii. Green Procurement

Green procurement process mainly integrates environmental principles to restrict the use of harmful substance on the procured material, components or products to decrease the generation of unsafe chemicals.

iii. Green Development

Green development is a business approach to reserve resources, lessens consumption of new raw material, and limits the use of lethal chemical substance in development and design decisions to minimize negative impacts over the environment.

iv. Reuse and Recycling

At the end of a design and development phase, practitioners shall recycle old products and reduce components to reduce resource waste and pollution.

4 CASE BASED REASONING (CBR) ARCHITECTURE FOR GREEN INFORMATION SYSTEMS

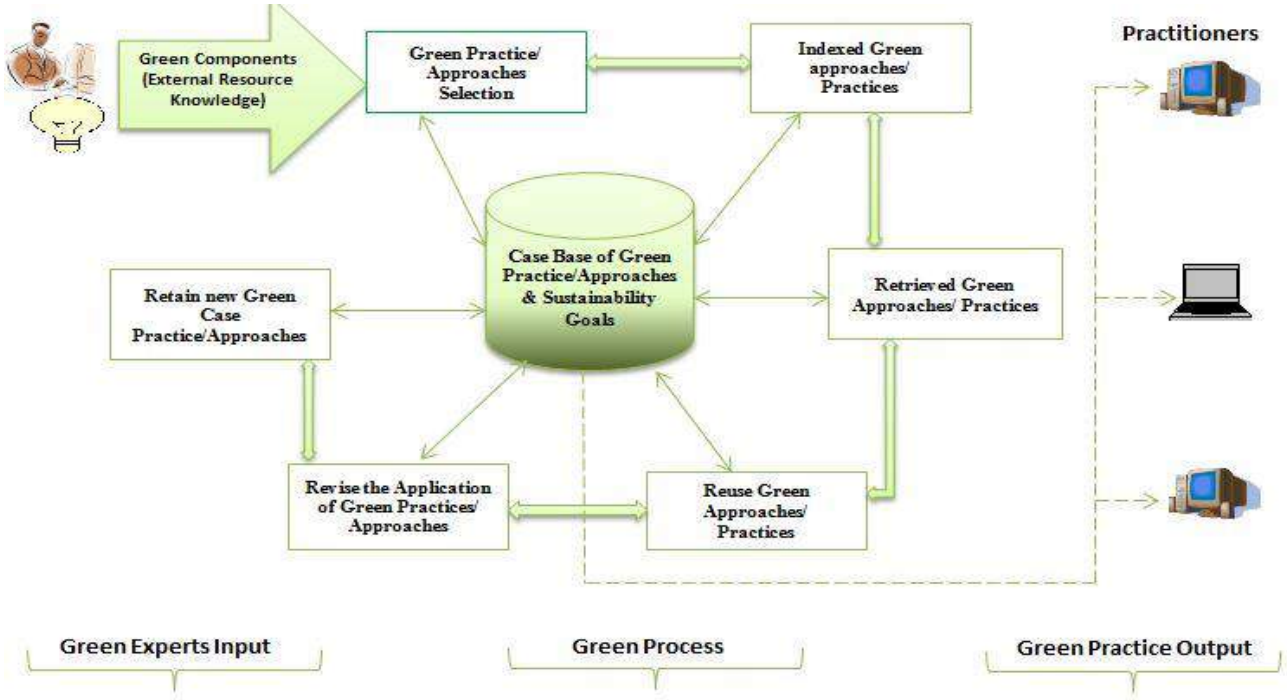

Figure 3. Case Based Reasoning (CBR) Architecture for Green IS Implementation in Organisation

Figure 2 shows the case based reasoning (CBR) architecture for Green IS implementation in organisation. The architecture uses the concept of *New Green Case Problem→ Retrieved Green Case Problem→ Retrieved Green Case Solution.* Below are the modules in the Green IS CBR architecture;

1. Green Experts Input

This is the first module in the CBR architecture, it involves;

a. Green Practice/Approaches Selection – in this phase the new Green cases are selected by the Green IS experts. These cases are added into the case base.

b. ***Indexed Green approaches/Practices*** – in this phase the added Green cases are indexed into the case base. In this phase the case base automatically assigns a unique case number to each case that is been added in to the case base.

c. Green Components (External Knowledge Source) – The Green components involves motivating forces, business strategies, technology & systems, management and practitioners as seen in Section 5 of this paper. Previous Information on these components are added into the case base. The added information aids and supports the organisation in attaining sustainable practice in their business process and are reused by current practitioners in carrying out sustainable practices in their organisation. This phase also involves inputting previous Green experience to solve new sustainability problems. This is the first phase of CBR for Green practice and is particularly appropriate for learning, since learning is already part of the life cycle of the CBR architecture. The Green data from the external source is used to improve the decision making process relating to Green IS implementation in the organisation.

2. Green Process

a. Retrieval Green Approach/Practice – involves the retrieval of similar Green cases from the case base. This phase is triggered by a query that consists of a new Green case describing the problem to be solved. Then, the case representation mechanism constructs an internal representation of the new case, which is compared with cases in the case base of the CBR system to identify similar Green cases.

b. Reuse Green Approach/Practice – involves the application of the green cases retrieved from the case base. In this phase a solution for the new Green case-problem is obtained from the similar retrieved Green cases. This Green solution will be eventually customized through the adaptation mechanism in order to be reused by the practitioners for sustainable business process.

c. Revise Green Approach/Practice– in this phase the Green solution from the case base is used and summarized. However, if the solution is suitable for the problem (achieving sustainability), the learning mechanism of CBR will evaluate the possibility of including the new Green solved case in the case base, which can be adopted in future.

d. Retain Green Approach/Practice–this phase formulates new Green solution from the practitioners. These solutions are retained and stored in the case base. However, in some cases the Green CBR architecture may not necessarily retain all new solved cases because it will not automatically imply the solution as an improvement of the quality of existing solutions that was retrieved.

e. Case Base- contains case base of Green Approach/Practice and how to achieve sustainability goals. It also provides different practical suggestions for business people to implement eco-entrepreneurial strategy including: green brainstorming, cost reductions, stimulation of innovation through green design and networking (interaction with stakeholders), attraction of interest of overwhelmed consumers in an emerging attention economy' through green marketing and through green start-ups (green-green businesses), and providing energy efficient computing. The case base of the CBR architecture is represented allows effective searches and inferences of Green practice thus facilitating Green understanding and reuse in the organisation.

3. Green Practice Output

a. Practitioners- these are the staffs and management involves in ensuring that Green practices are implemented in the organisations business process. The practitioners' objective is to ensure that sustainability is achieved in the organisation. These practitioners utilize the case base to accomplish the eco-sustainability goal (Eco-efficiency, Eco-effectiveness and Eco-equality). Thus it's the practitioners' duty to use IS to achieve a sustainable environment.

4.1 Benefits of CBR Architecture towards Sustainable Environment

Currently practitioners uses their intuitively and experience in implementing diffusing Green standards, since practitioners are faced with new types of projects with different project characteristics and requirements. To promote sustainability, practitioners could be provided with information on previously cases, which are similar to the new project. In this case, they can systematically capture and organize knowledge from the past cases to meet the new requirements. However, existing Green approaches don't utilize previously certified cases to provide suggestions for the practitioners in implementing green procedures.

Since practitioners frequently meet new types of projects with different project characteristics and requirements. It would be helpful if practitioners could refer to other similar green diffusion cases when planning, designing and implementing projects in their organisation. Thus the proposed case-based reasoning (CBR) architecture provides case studies of similar Green projects. Figure 3 shows the case based reasoning architecture, it consists of CBR elements which are cases and case base. Where, cases are practical and valuable

sources of knowledge; they contain the structured knowledge and the collection of information to support Green decision-making. The Case base in the CBR architecture can be used as a storage of cases that provides training and education on how practitioners can implement Green practices in their organisation to support sustainable practice. The case-based reasoning CBR architecture can provide case studies of similar certified green practices, procedures and suggestions on how to resolve Green challenges such as providing information and support to practitioners in managing their business process.

Thus CBR architecture provides a practical means to realize environmentally friendly operations and logistics. Via a knowledge pool that can be used to inform practitioners by facilitating practitioners in understanding and implementing greening strategies/approaches and to play a critical role in realizing a Green business process. The CBR has the potential and opportunity to support practitioners in implementing Green practices.

The CBR architecture will play an important role in supporting organisations and its staffs to learn more of the approaches to improve their environmentally friendly operations and logistics activities. Thus to search and share best Green practices, comprising relevant strategies, approaches, etc. in the (CBR) architecture the case base is used to store all Green practices and procedures which holds the good practices that can be shared by all practitioners in the organisation.

Thus experts in the organisation can provide their good practices into the common case base for sharing, through the organisations network. But in order to ensure the appropriateness of information, the organisation should conduct a screening process before adding their successful practices as useful knowledge input into the common case base. This also requires that besides a common case base, each practitioner may have his/her own individual case base, to store its good practices at first instance. The CBR architecture is based one of the most popular approaches for the development of knowledge-based systems and also the CBR architecture is considered similar to the way human beings solve problems.

4.2 Benefits of CBR Architecture in Green Practices Implementation

a. CBR assists practitioners in implementing green initiatives and approaches.

b. CBR aids to identify the status of their operations and logistics processes with regard to the green aspects.

c. CBR provides methods in facilitating and improving organisations green activities.

d. CBR facilitate practitioners in understanding, selecting, developing and implementing Green strategies/ approaches.

e. CBR also provides reference/ guidance/ learning to practitioners for successful Green practice.

5 COMPONENT BASED MODEL FOR GREEN INFORMATION SYSTEMS

This paper also proposes component based model for achieving eco-sustainability goal using green IS Components as seen in Figure 4. The model assists practitioners, and management to disseminate and transfer information with the support of green IS components in their organisation without compromising the benefit of the society and environments now and in the future. In eco-sustainable, Green IS plays a key role in the dissemination of information among practitioners that is the practice of using computing resources in an eco-friendly and efficiently manner. The proposed case based model provides a guide for researchers; academia and practitioner interested in Green IS research. The component based model can be utilized as a road map to build the beliefs and norms of practitioners, stack holders and end users about the environment which is essential to diffuse and adopt a sustainable practice in their organisations [17]. The component based model is adopted from previous works from [17]

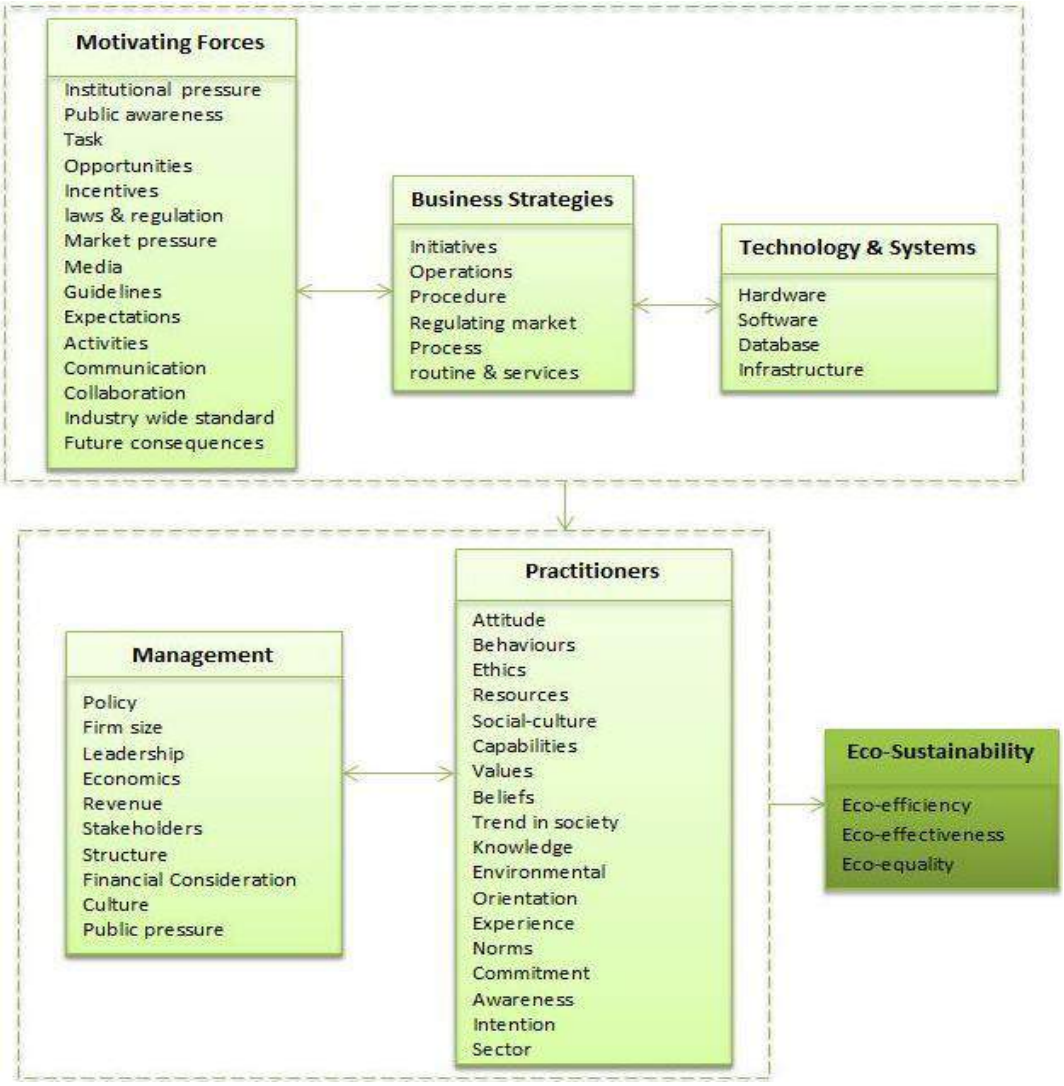

Figure 3. A Component Based Model for Green IS Diffusion in Organisation (Adopted From [17])

Figure 3 shows the component based model for Green IS diffusion in organisation. The five components in the model influence the adoption of Green IS in organisations. The component based model for green IS diffusion considers the attributes that influence the components of Green IS, which are shown in figure 3 as Motivating Forces, Business Strategies, Technologies & Systems, Organization and Employee in context to the practice of Green IT/IS in organizations. Thus the component based model can assist practitioners in having a positive impact, with the potential to reduce global production of CO2 in their organisational business continuity.

5.1 Proposed Model Components

Below are the components for Green IS diffusion in organisation;

1. Motivating Forces

Motivational Forces controls how a firm conducts its business relation to adoption Green IT/IS based on standard set mostly by government or industries. These standards and Regulation are dedicated to sustain external groups and would include responding to pressure from government, regulatory bodies, and staffs of organizations. Furthermore, these regulations also refer to additional measures targeting the long term

sustainability, avoiding fines and penalties and also conserving the environment when mitigating risk. However, government regulation and policies are put in to place to ensure managers and practitioners respond to environmental issues and incorporate it into their business strategy, which is usually required, but also in the form of guidelines. Government regulations may apply to the state, the entire territory, or district-wide. Another type of standard is the Industry based standard which includes compulsory formal rules [17].

2. Business Strategies

Mainly involves description of the organization in terms of its scope, size and management structure. This phase involves how the organization implements their business strategies. The strategy used by the organization to accomplish its objectives is very important in the adoption Green IT/IS practices. Business strategy mainly aims to support organizations in reducing their operating costs in product development. Thus green IT/IS can be used as a strategic decisions tool. Business strategy is the most prominent drivers of green IT/IS and it's aimed to reduce costs and carbon emissions, and achieve sustainable development [17].

3. Technologies & Systems

Technologies and system defines the features of technology which would affect Green IS adoption and implementation. Therefore Technology and Systems refers to Infrastructure that enable sustainable-related businesses, these technology and systems are utilized by IT professionals and practitioners processing the knowledge and skills to implement green IS practices in their organisation [17].

4. Organization

Green IT/IS practices is able to redesign how practitioners in the organization advancing their future eco-sustainability procedures and practices, especially those facilitated by information systems. The management works together to realize the social, economic as well as environmental benefits of accomplishing the aims and objectives of the organisation. The management must provide training successions and campaigns to inform the employee on how techniques such as telematics can improve Green IT/IS and why commitment of the organization is essential in implementing Green IT/IS [17].

5. Employee

These are IT/IS managers and professionals (system designers and developers), environmental stewards and top management. The employees' commitment is mandatory in planning, implementing, validating, deploying and maintaining the organisations system with eco sustainability considerations in mind. Thus the employees' collaboration is based on their shared interest and stakes in fulfilling the objectives or requirements of the organization; this can be seen as part of the environmental ethics/stewards in the organization [17].

5.2 Green IS Components and Attributes

Table 3 shows the Green IT/IS components and their attributes.

Table 3. Green IT/IS Components and Attributes

Component	Attributes
Motivating Forces	Institutional pressure, public awareness, task, opportunities, incentives, laws & regulation, market pressure, media, guidelines, expectations, activities, communication, collaboration, industry wide standard and future consequences.
Business Strategies	Initiatives, operations, procedure, regulating market, process, routine & services.
Technologies & Systems	Hardware, software, database & infrastructure.
Organization	Policy, firm size, leadership, economics/, revenue, stakeholders, structure, financial consideration/cost saving, culture & public pressure
Employee	Attitude, behaviours, ethics, resources, social-culture, capabilities, values, beliefs, trend in society, knowledge, environmental orientation, experience, norms, commitment, awareness, intention & sector.

These attributes influences and determines how the components are diffused to achieve a

sustainable environment in the usage of IT infrastructures in organisations. The attributes are to be considered in implementing a Green and sustainable practice.

5.3 Eco Sustainability Goals

Eco-sustainability is the capability of one or more object or beings, either independently or mutually, to exist either unaffected or in changed forms for lengthy time frame. Eco-sustainability aims to assist organisations diffuse and implement green practice implementation [17]. Table 4 shows the eco sustainability goal. The first goal, eco-efficiency, has received more attention in the latest years. Indeed, Green IT has been working on the alignment of its business function with this goal. According to [18] Green IS can potentially address the three goals.

Table 4. Eco Sustainability Goals

Eco Sustainability Goals	Description
Eco-efficiency	Refers to organization's ability to provide competitively services that satisfy end users' needs and bring value of life while gradually decreasing ecological impacts [17]; [18]; [10], [7]; [11]. [10] added that in eco-efficiency, financial goals still remain foremost in organization's mind, but it should be mindful of the need to pursue sustainable practices where they do not interfere with financial considerations. Thus Eco-efficiency combines traditional efficiency goals with ecological considerations, and is defined as the delivery of competitively products and services that satisfy human needs and bring quality of life, while progressively reducing ecological impacts and resource intensity throughout the life-cycle to a level at least in line with the earth's carrying capacity [10].
Eco-effectiveness	Aims to stop pollution and reduction of natural resources by guiding practitioners and firms' responsiveness to the fundamental and important factors of environmental issues via a fundamental redesign of green system [17]. Eco-effectiveness is the most promising goal. It means doing the right thing, e.g. making the choices that make possible long-term prosperity and environmental sustainability. It requires a shift in societal values and business models to transform our current economy [17]; [18]; [10], [7]; [11]. Eco-effectiveness means that we end current practices that result in ecological degradation. There is need to mimic nature and create ongoing healthy systems where the waste products of one process become inputs to other processes [10]. Eco-effectiveness involves the life-long concern with educating managers on the difference between efficiency (doing things right) and effectiveness (doing the right things). While eco-efficiency might focus on reducing the energy consumption of artificial lighting, the eco-effective approach would design work places that use natural lighting. Eco-effectiveness, considered the ultimate solution for ecological problems, requires a shift of mind set and transformation of business models [7].
Eco-equity	Is concerned with the equal rights of people to environmental resources and a business's social responsibility for the future generation [17]; [18]; [10], [7]; [11]. It aims for fair distribution of natural resources between current and future generations. One group in society should not consume so much that it denies other members of its generation their fair share of that resource. To implement eco-equity as a societal goal, there is need to know what resources we have and how rapidly they are being consumed. Thus data base is needed for the full range of global resources with details of available stocks and depletion rates. Then, Green IS can develop methods for determining equitable distributions between and across generations [10].

5 DISCUSSIONS

Green IS sees IT primarily as a problem to be treated; for example, data centers are a rapidly growing source of carbon emissions and need to be made more energy efficient to reduce their impact on the environment. Green IS utilizes information systems as a possible solution to many environmental problems. Presently, electricity generation is a major source of CO2 emissions; smart grid technology can employ information systems to increase efficiency in the production, transmission, and use of electricity [2]. Green IS can contribute as solution to

environmental degradation and highlights conservation of the environment in order to support human life into the future [5]. Green IS, and sustainable development, should not been seen as a cost of doing business. Rather, they are opportunities for organizations to improve productivity, reduce costs, and enhance profitability.

Poor environmental practices result in many forms of waste. Unused resources, energy inefficiency, noise, heat, and emissions are all waste products that subtract from economic efficiency. Less waste means a more efficient enterprise. Firms that actively pursue green IS to create sustainable business practices are doing the right thing for their community, customers, investors, and future generations. Managers seeking to create sustainable organizations and Green IS should find the CBR architecture very useful for thinking about problems, brainstorming solutions, and planning implementation of innovations. Hence, it is important to provide an approach for assisting management and practitioners in the development of Green IS. We start by recognizing the four fundamental drives of information systems [10].

The developed CBR architecture can provide practitioners and management with references of previous cases for new green projects, so that they can learn and utilize good practices from the past. Since each project is unique and system design is the result of a creative effort, the practices in previous cases may not be directly applicable to the new cases. Therefore, practitioners are allowed to modify the suggested solution according to the nature and needs of the new cases, as specified in the *revise* phase of the CBR cycle.

Practitioners and management can use the *retrieved* similar cases as the base, and *revise* their design by considering the project-specific characteristics and requirements. However CBR technique does not encourage practitioners, academicians or researchers to use past technology to deal with similar issues but encourages them to learn from the past, and eventually *revise* in order to have an authentic Green design or Green procedure. It is believed that this type of knowledge technique can promotes better green implementation and diffusion [14].

6 CONCLUSION AND FUTURE WORKS

To better understand the impacts of IT on the environment, practitioners should be aware that the use of IT results to 2 % of the total CO2 emissions in the U.S., similar to the emissions created by the airline industry. Thus, Green IT seems to address a very small part of the overall climate change problem, although the environmental impacts of IT could grow very rapidly as hundreds of millions of people in the developing world go online for the first time, and data intensive applications such as video streaming demand more computing power. By contrast, the use of Green IS throughout the economy potentially addresses much of the other 98 % of emissions [2].

Green IS mainly includes the use of information systems to enhance sustainability across the economy. This view includes enhancing efficiency in industries that are major sources of CO2 emissions, such as the transportation, manufacturing, and energy sectors. Green IS can therefore lead not only to reduction in carbon emissions but also to significant cost savings. This implies that Green IS can decrease organization's environmental footprint.

Green IS has a greater potential than green IT because it tackles a much larger problem. It can make entire systems more sustainable compared to reducing the energy required to operate information technologies. Thus Green IS can be implemented in organisation using CBR, which is a problem-solving paradigm in which, CBR use knowledge from past experiences to solve new problems. CBR consists of two basic parts: one corresponding to the description of the problem that is used to map similar cases, and other, which specifies the solution to the problem. It comprises of a case-based mechanism for case representation; a mechanism for retrieving cases and a similarity measure to identify relevant cases to a current problem by retrieving existing cases based on the current situation; a learning mechanism enabling the system to remember successful solved cases in order to reuse them in solving new problems [19].

CBR architecture was developed to provide adequate data that can train, educate and support

practitioner in implementing Green practices in organisations by suggesting ways of promoting sustainability in their organisation and increase the environmental orientation of practitioners in their organisation. A component based model was also designed to assist practitioners and management in the adoption of Green practices in in their organisational. The component based model creates a profounder understanding of how Green IS can support the sustainability of the organization and identify the components and attributes that affect Green IS use based on a comprehensive literature review. The component based model can also help close the gap in practices, issues related to diffusion Green IS in organization, hence serving as a guide for practitioners in achieving sustainability in their business process.

Future work will involve the development of a CBR algorithm and implementation of a CBR system to support practitioners in utilizing information systems to achieve sustainability. The designed component based model will be refined and validated based a case study that is to carried out in the future aimed to test the Green IS components and attributes.

REFERENCES

[1] B. J. Vom, R. Watson, C. Dwyer, S. Elliot, and N. Melville, "Green Information Systems: Directives for the IS Discipline," Communications of the Association for Information Systems (CAIS), vol. 33, pp.509-520, 2013.

[2] D. Jason, "Green IS: Concepts and Issues for Information Systems Research," Communications of the Association for Information Systems: vol. 27, pp.173-184, 2010.

[3] A. Molla, "GITAM: A Model for the Adoption of Green IT," 19th Australasian Conference on Information Systems Green IT Adoption Model 3-5 Dec 2008, Christchurch Molla, pp. 658-668.

[4] S. Q. Robert, L. L. Hsu, J. C. H. Chen and M. D. Chang, "A Case Study on Green Information System," Proceedings of ASBBS Annual Conference: Feb 2012, Las Vegas, pp.512-519.

[5] H. R. Grant and L. Sam, "Synthesis of Green IS Frameworks for Achieving Strong Environmental Sustainability in Organisations," SAICSIT'12, Oct 01 - 03 2012, Pretoria, South Africa.pp. 306-315.

[6] C. Edward, S. Hasan, U. U. Hassan, M. Herstand, S. O'Riain, "An Entity-Centric Approach to Green Information Systems," pp.1-7, 2011.

[7] R. T. Watson, M. C. Boudreau and A. J. Chen, "Information Systems and Environmentally Sustainable Development: Energy Informatics and New Directions For The Is Community," MIS Quarterly, vol.34, pp. 23-38, 2010.

[8] J. Xiuna and L. Bai, "The Enterprise Application Information System Integration based on the Green Supply Chain Management," 2009 International Conference on Information Technology and Computer Science (ITCS 2009) IEEE, 2009, pp. 433-435.

[9] C. Hong-Mei and R. Kazman, "Architecting Ultra-Large-Scale Green Information Systems," GREENS 2012 IEEE, 2012, Zurich Switzerland, pp.69-75.

[10] R. T. Watson, M. Boudreau, A. Chen and M. Huber,"Green IS: Building Sustainable Business Practices," Information Systems: A Global Text, pp.1-17, USA, 2008.

[11] J. C. Adela, R. T. Watson, M. C.Boudreau, E. Karahanna "An Institutional Perspective on Tthe Adoption of Green IS & IT," Australasian Journal of Information Systems, vol. 17, pp. 5-27, pp. 2011.

[12] P. Chris, C. McGibbon and I. Brown, "Adoption of Green IS in South Africa – An exploratory study," SAICSIT '11, Oct 3–5, 2011, Cape Town, South Africa, pp. 330-333.

[13] S. Murugesan, "Harnessing Green IT: Principles and Practices," IEEE IT Professional, pp. 24-33, 2008.

[14] C.P. C. Jack, J. M.Lucky, "A non-linear case-based reasoning approach for retrieval of similar cases and selection of target credits in LEED projects," Building and Environment vol.93, pp.349-361, 2015.

[15] W. Chengbo, J. Vaughan, J. Mercer and Y. Zhao, "A Case-Based Model Facilitating Retailing Operations Going "Green": A Proposed Research Agenda with a Consideration of Recession," IEEE, pp. 1-4.

[16] B. Tom, "Towards a Practice-Oriented Green IS Framework," ECIS 2011, Proceedings, pp.1-12.

[17] A. J. Bokolo and C.P Noraini, "A Framework for Adoption and Implementation of Green IT/IS Practice in IT Governance," Proceedings of Third International Conference on Green Computing, Technology and Innovation (ICGCTI2015), SDIWC, Serdang Malaysia, 2015, pp. 38-48.

[18] C. C. Ana and J. Á. Carvalho, "Green Information Systems: The Use of Information Systems to Enhance Sustainable Development,"Engineering Week 2010 Guimarães, 11- 15 Oct 2010, pp. 1-2.

[19] L. Adriana and R. Girardi, "Case-Based Reasoning Architecture of an Hybrid Software Agent Software Agent," 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) 2010, pp.79-86.

International Journal of Digital Information and Wireless Communications (IJDIWC) 6(2): 112-121
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

Simple Method For Determining Harmonic Sequences in a Machine, Transformer or Network

Claude Ziad Bayeh
EEE GROUP-R&D department
LEBANON
Email: c.bayeh@eee-group.net
c.bayeh@hotmail.com

ABSTRACT

Harmonics in power systems are very dangerous and may cause different problems for equipment, i.e. the third harmonic has the same amplitude and angle in each phase of a three phase system. It enters into the neutral and may damage or overheat it if the system is unbalanced. The fifth harmonic causes an inverse polarity in machines where the magnetic field rotates in the opposite direction than the normal magnetic field, this causes the rotor to reduce its speed. Harmonics can produce different distortions for all equipment, and may breakdown some of them and damage other sensitive equipment. They reduce the lifetime of super-capacitors and other equipment, etc. These harmonics are caused by non-linear loads or natural disasters. Thus, it is very important to analyze these harmonics and filter them. For many engineers, it is necessary to arrange harmonics into three sequences, direct sequence, inverse sequence and homopolar sequence. Therefore, to arrange these harmonics, they have introduced many methods, but unfortunately, till now the traditional methods are complicated and take a longer time to find the harmonic sequences. Moreover, students find difficulties in the traditional methods and need some faster and simpler method to help them determining the sequence of harmonics, whatever is their range. For this reason, this paper proposes a new method to determine harmonic sequences in the easiest and fastest way.

KEYWORDS

Determining Harmonics, Direct sequence, Inverse sequence, Homopolar sequence, Simple method, Transformer, Machine.

1 INTRODUCTION

In Electrical Engineering, it is very important to analyze unbalanced Three-Phase power systems and to know the sequence of harmonics and phases. The first paper about the symmetrical components was introduced by Charle L. Fortescue [1]. He demonstrated that any asymmetrical set of unbalanced M phasors could be decomposed into a linear combination of the same number M of a balanced symmetrical set of phasors. So the symmetrical components are a Direct sequence (also called Positive sequence), an inverse sequence (also called negative sequence), and a zero sequence (also called Homopolar sequence). These 3 sequences are the basis for analyzing power systems and their effect on equipment such as machines [2-7], transformers [8-11] and power systems [12-13].

The proposed method by Fortescue allows determining the harmonic sequences in order to filter them, because they cause problems on networks, transformers, and machines, and because they introduce heating, electrical pollution, inverse polarity in machines (i.e. Fifth Harmonic) and may affect the functionality of some sensitive loads. Thus, by knowing the sequence of harmonics, it permits to filter the unneeded harmonics. Hence, active and passive filters may be designed in a way to eliminate or reduce these harmonics.

For students, engineers and specialists, the time needed to find the harmonic sequences is very

precious; the less time consumed to find harmonics, the better the achievement is for them.

This paper proposes a new and simple method to find Harmonic Sequences, just with one operation for each sequence, which makes the calculation much easier and faster. And it doesn't need even a calculator or any medium to calculate it. This can be also applied on programming part which finds the harmonic sequences, the program will be much shorter, more efficient and faster.

In the second Section, the Symmetrical Components are presented; the normal current is function of Harmonic Sequences. The Third Section discusses the most popular methods of determining the harmonic sequences. In the Fourth Section, this paper introduces a new formula called "Nth Fractional Part of Real Number", this new formula is used in Section Five to determine the harmonic sequences rapidly. The Fifth Section is the most important Section in this paper, where a new equation is defined to determine in one operation, whether the Harmonic has a direct sequence, an inverse sequence or a zero sequence. In the Sixth section, a discussion of the four methods is presented. And finally, a conclusion is developed in Section Seven.

2 SYMMETRICAL COMPONENTS

Fortescue [1] has decomposed any three phase vectors for an electrical system into three balanced systems which are (direct, inverse, and homopolar) in the following way:

$$V_a = V_{a0} + V_{a1} + V_{a2} \quad (1)$$
$$V_b = V_{b0} + V_{b1} + V_{b2} \quad (2)$$
$$V_c = V_{c0} + V_{c1} + V_{c2} \quad (3)$$

Where,
V_a, V_b, V_c are the three phases of the system,
V_{a0}, V_{b0}, V_{c0} are the homopolar components of the system,
V_{a1}, V_{b1}, V_{c1} are the direct components of the system,
V_{a2}, V_{b2}, V_{c2} are the inverse components of the system,

For the direct sequence
$V_{a1} = V_1, V_{b1} = a^2 V_{a1} = a^2 V_1, V_{c1} = aV_{c1} = aV_1$

For the inverse sequence
$V_{a2} = V_2, V_{b2} = aV_{a2} = aV_2, V_{c2} = a^2 V_{c2} = a^2 V_2$

For the homopolar sequence
$V_{a0} = V_0, V_{b0} = V_{a0} = V_0, V_{c0} = V_{c0} = V_0$

Where,
$a = 1\angle 120^0$,
$a^2 = 1\angle 240^0$,
$a^3 = 1\angle 360^0$,
$a^4 = a = 1\angle 120^0$.

The same concept can be applied to the current. Therefore, the Symmetrical components matrices of the voltages and currents are as following:

$$\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & a^2 & a \\ 1 & a & a^2 \end{bmatrix} \cdot \begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & a^2 & a \\ 1 & a & a^2 \end{bmatrix} \cdot \begin{bmatrix} I_0 \\ I_1 \\ I_2 \end{bmatrix} \quad (5)$$

And

$$\begin{bmatrix} V_1 \\ V_2 \\ V_0 \end{bmatrix} = \begin{bmatrix} 1 & a & a^2 \\ 1 & a^2 & a \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} I_1 \\ I_2 \\ I_0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & a & a^2 \\ 1 & a^2 & a \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \quad (7)$$

The current in each phase is:
$$I_a = \sum_{n=1}^{\infty} (I_{0n} + I_{1n} + I_{2n}) \quad (8)$$
$$I_b = \sum_{n=1}^{\infty} (I_{0n} + a^2 I_{1n} + aI_{2n}) \quad (9)$$
$$I_c = \sum_{n=1}^{\infty} (I_{0n} + aI_{1n} + a^2 I_{2n}) \quad (10)$$

The current of each sequence is:
$$I_0 = \sum_{n=1}^{\infty} I_{0n} \quad (11)$$
$$I_1 = \sum_{n=1}^{\infty} I_{1n} \quad (12)$$
$$I_2 = \sum_{n=1}^{\infty} I_{2n} \quad (13)$$

Where,
n is the harmonic number,
I_0 is the current of homopolar sequence,
I_1 is the current of direct sequence,
I_2 is the current of the inverse sequence,
I_a is the current of the phase a,
I_b is the current of the phase b,
I_c is the current of the phase c.

This method is the most used method to determine the harmonic sequences. An example of this method is presented in the following section.

3 TRADITIONAL METHODS FOR DETERMINING THE HARMONIC SEQUENCES

There are many developed methods to determine the sequence of harmonics, three of the most famous methods which are used by all engineers, researchers and students are discussed in this section:

3.1 Mathematical Method

The mathematical method based on Fortescue equations consists of determining the harmonic sequences using equations that calculate the vectors of currents, their magnitudes and their phase angles, this method is very long and it takes approximately 7 to 10 minutes to calculate the sequence of one harmonic, it is described as following:

3.1.1 First step-Finding the phase currents

This step consists of finding the phase currents for the three phase system I_a, I_b, and I_c, where,

$$I_a = I \cdot \cos(wt - 0^0) = I\angle 0^0 \quad (14)$$
$$I_b = I \cdot \cos(wt - 120^0) = I\angle -120^0 \quad (15)$$
$$I_c = I \cdot \cos(wt + 120^0) = I\angle 120^0 \quad (16)$$

And,

$$I_{a,h} = I \cdot \cos\big(h(wt - 0^0)\big) = I\angle(h \cdot 0^0) = I\angle 0^0 \quad (17)$$

$$I_{b,h} = I \cdot \cos\big(h(wt - 120^0)\big) = I\angle\big(h \cdot (-120^0)\big) \quad (18)$$

$$I_{c,h} = I \cdot \cos\big(h(wt + 120^0)\big) = I\angle\big(h \cdot (120^0)\big) \quad (19)$$

Where,
$I_{a,h}$, $I_{b,h}$, $I_{c,h}$ are the currents of the h harmonic in each phase.

3.1.2 Second step-Finding the direct, inverse and homopolar current sequences

After finding equations (17) to (19), the second step is to find the harmonic sequence using the matrix in equation (7) by replacing equations (17) to (19) into it, thus, the elements of this matrix for a harmonic h are determined as following:

The current of the homopolar sequence is:

$$I_{0,h} = \frac{1}{3}\big(I_{a,h} + I_{b,h} + I_{c,h}\big) \quad (20)$$

$$I_{0,h} = \frac{1}{3}\Big(I\angle 0^0 + I\angle\big(h \cdot (-120^0)\big) + I\angle\big(h \cdot (120^0)\big)\Big) \quad (21)$$

The current of the direct sequence is:

$$I_{1,h} = \frac{1}{3}(I_{a,h} + a \cdot I_{b,h} + a^2 \cdot I_{c,h}) \quad (22)$$

$$I_{1,h} = \frac{1}{3}\Big(I\angle 0^0 + a \cdot I\angle\big(h \cdot (-120^0)\big) + a^2 \cdot I\angle\big(h \cdot (120^0)\big)\Big)$$

$$I_{1,h} = \frac{1}{3}\Big(I\angle 0^0 + I\angle\big(120^0 + h \cdot (-120^0)\big) + I\angle\big(240^0 + h \cdot (120^0)\big)\Big) \quad (23)$$

The current of the inverse sequence is:

$$I_{2,h} = \frac{1}{3}(I_{a,h} + a^2 \cdot I_{b,h} + a \cdot I_{c,h}) \quad (24)$$

$$I_{2,h} = \frac{1}{3}\Big(I\angle 0^0 + a^2 \cdot I\angle\big(h \cdot (-120^0)\big) + a \cdot I\angle\big(h \cdot (120^0)\big)\Big)$$

$$I_{2,h} = \frac{1}{3}\left(I\angle 0^0 + I\angle\left(240^0 + h \cdot (-120^0)\right) + I\angle\left(120^0 + h \cdot (120^0)\right)\right) \quad (25)$$

By calculating equations (21), (23), and (25) we can determine the sequence of the harmonic, we should get two equations equal to zero and one equation different from zero, the equation different from zero gives the correct sequence of the harmonic.

3.1.3 Example

Calculate the sequence of the harmonic 29?

Firstly, one calculates $I_{a,h}$, $I_{b,h}$ and $I_{c,h}$ from equations (17) to (19),

$$I_{a,29} = I \cdot \cos\left(29(wt - 0^0)\right) = I\angle(29 \cdot 0^0) = I\angle 0^0$$

$$I_{b,29} = I \cdot \cos\left(29(wt - 120^0)\right) = I\angle\left(29 \cdot (-120^0)\right) = I\angle(-3480^0) = I\angle(-3480^0 + 9 \cdot 360^0) = I\angle(-240^0)$$

Because 3480/360=9.666, so one needs $9 \cdot 360^0$ to reduce the equation, and because $9 \cdot 360^0 = 0^0$,

$$I_{c,29} = I \cdot \cos\left(29(wt + 120^0)\right) = I\angle\left(29 \cdot (120^0)\right) = I\angle(3480^0) = I\angle(3480^0 - 9 \cdot 360^0) = I\angle(240^0)$$

Secondly, one calculates the homopolar, direct, and inverse current sequences $I_{0,29}$, $I_{1,29}$, and $I_{2,29}$ from equations (20) to (25),

Thirdly, one calculates the homopolar sequence:

$$I_{0,29} = \frac{1}{3}\left(I_{a,29} + I_{b,29} + I_{c,29}\right)$$

$$I_{0,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle\left(29 \cdot (-120^0)\right) + I\angle\left(29 \cdot (120^0)\right)\right)$$

$$I_{0,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-3480^0) + I\angle(3480^0)\right)$$

$$I_{0,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-3480^0 + 9 \cdot 360^0) + I\angle(3480^0 - 9 \cdot 360^0)\right)$$

$$I_{0,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-240^0) + I\angle(240^0)\right)$$

$$I_{0,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(120^0) + I\angle(-120^0)\right) = 0$$

Because the sum of three opposite vectors is equal to 0, as in figure 1.

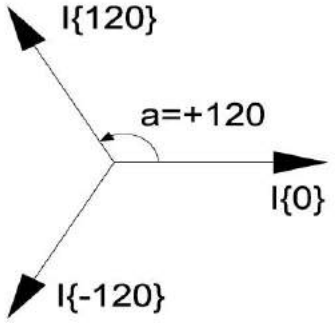

Figure 1: Three opposite vectors with difference in phase equal to 120 and -120 degrees, their sum is equal to zero.

Therefore, the homopolar sequence is equal to zero.

Fourthly, one calculates the direct sequence:

$$I_{1,29} = \frac{1}{3}\left(I_{a,29} + a \cdot I_{b,29} + a^2 \cdot I_{c,29}\right)$$

$$I_{1,29} = \frac{1}{3}\left(I\angle 0^0 + a \cdot I\angle\left(29 \cdot (-120^0)\right) + a^2 \cdot I\angle\left(29 \cdot (120^0)\right)\right)$$

$$I_{1,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle\left(120^0 + 29 \cdot (-120^0)\right) + I\angle\left(240^0 + 29 \cdot (120^0)\right)\right)$$

$$I_{1,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(120^0 - 3480^0) + I\angle(240^0 + 3480^0)\right)$$

$$I_{1,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-3360^0) + I\angle(3720^0)\right)$$

$$I_{1,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-3360^0 + 9 \cdot 360^0) + I\angle(3720^0 - 10 \cdot 360^0)\right)$$

Because 3360/360=9.333, so one needs $9 \cdot 360^0$ to reduce the equation, and because $9 \cdot 360^0 = 0^0$, and 3720/360=10.333, so one needs $10 \cdot 360^0$ to reduce the equation, and because $10 \cdot 360^0 = 0^0$.

Thus,

$$I_{1,29} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-120^0) + I\angle(120^0)\right) = 0$$

Therefore, the direct sequence is equal to zero.

Fifthly, one calculates the inverse sequence,

$$I_{2,29} = \frac{1}{3}(I_{a,29} + a^2 \cdot I_{b,29} + a \cdot I_{c,29})$$

$$I_{2,h} = \frac{1}{3}\Big(I\angle 0^0 + I\angle\left(240^0 + 29 \cdot (-120^0)\right) + I\angle\left(120^0 + 29 \cdot (120^0)\right)\Big)$$

$$I_{2,h} = \frac{1}{3}\left(I\angle 0^0 + I\angle(240^0 - 3480^0) + I\angle(120^0 + 3480^0)\right)$$

$$I_{2,h} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-3240^0) + I\angle(3600^0)\right)$$

$$I_{2,h} = \frac{1}{3}\left(I\angle 0^0 + I\angle(-3240^0 + 9 \cdot 360^0) + I\angle(3600^0 - 9 \cdot 360^0)\right)$$

$$I_{2,h} = \frac{1}{3}(I\angle 0^0 + I\angle 0^0 + I\angle 0^0) = I$$

Therefore, the inverse sequence is different from zero, and one can conclude that the harmonic 29 has an inverse sequence because $I_{2,h} = I$.

3.1.4 Discussion

In order to complete this method for one harmonic one needs,

- Number of operations = 200
- Number of equal symbol "=" = 33
- Number of lines used to write equations = 24
- Time needed to find the harmonic sequence = 7.25 minutes (435 seconds)

3.2 Simplified Method

Because the Mathematical Method is too long, some engineers proposed a simplified method, it consists of calculating the harmonic sequences using simple arithmetic equations as following:

Direct Sequence (or Positive Sequence):

$$h = 3k + 1 \quad (26)$$

and

$$k = \frac{h-1}{3} \quad (27)$$

Where,
h is the harmonic sequence,
k represents an Integer Number.

Inverse Sequence (Negative Sequence):

$$h = 3k - 1 \quad (28)$$

and

$$k = \frac{h+1}{3} \quad (29)$$

Homopolar Sequence (Zero Sequence):

$$h = 3k \quad (30)$$

and

$$k = \frac{h}{3} \quad (31)$$

3.2.1 Example of determining the Harmonic Sequences using the Simplified Method

Consider the harmonic which is proposed in the previous example, h=29. Calculate the harmonic sequence.

For h=29, one uses equations (27), (29), and (31) until obtaining an integer number,

- For the Direct Sequence:

$$k = \frac{h-1}{3} = \frac{29-1}{3} = 9.3333$$

k is not an integer number, thus $h = 29$ is not a Direct Sequence.

- For the Inverse Sequence:

$$k = \frac{h+1}{3} = \frac{29+1}{3} = 10$$

k is an integer number, thus $h = 29$ is an Inverse Sequence.

- For the homopolar Sequence:

$$k = \frac{h}{3} = \frac{29}{3} = 9.666$$

k is not an integer number, thus $h = 29$ is not a homopolar Sequence.

3.2.2 Discussion

In order to complete this method for one harmonic, one needs,

- Number of operations = 10
- Number of equal symbol "=" = 9
- Number of lines used to write equations = 3
- Time needed to find the harmonic sequence = 80 seconds

This method is much practical than the Mathematical Method, but it can also be improved. For a large number of harmonics, this method takes a longer time to determine all harmonic sequences. So a simpler method should be developed to reduce the time of calculation. Practically the Simplified method needs improvement which will be proposed in this paper in section 5.

Currently, such calculation can be found using specific software or a table similar to Table 1. But even, the software can be more simplified in order to reduce the time of simulation and the consuming time to develop such software.

3.3 Determining the Harmonic Sequences using a table of harmonics

Table 1 can also be used to pick out the harmonic sequences, but the disadvantage of this table is that for high harmonics the table will be very large and not practical, and it needs to be always carried.

Table 1, A table presents the Harmonic number for each sequence [16].

Harmonic Rang	Sequence		
	Direct	Inverse	Homopolar
	1	2	3
	4	5	6
	7	8	9
	10	11	12
	13	14	15
	16	17	18
	19	20	21
	22	23	24
	…	…	…

3.3.1 Example of determining the Harmonic Sequences using the Table of harmonics

Consider the harmonic which is proposed in the previous examples, h=29, calculate the harmonic sequence?

The table 1 doesn't contain harmonics more than 24, so one has two choices, the first one is to do a larger table which contains at least the first 200 harmonics, or the second one is to interpret the harmonic sequences using a calculator and by adding +3 for every column until arriving to the desired sequence.

For our example:
22+3=25,
25+3=28,
Thus, 29=28+1, therefore this harmonic has an inverse sequence.

3.3.2 Discussion

In order to complete this method for one harmonic, one needs,

- Number of operations≥1 (it depends on the dimension of the given table), for h=29, one needs 3 operations,
- Number of equal symbol "=" ≥ 1 (it depends on the dimension of the given table), for h=29, one needs 3 equal symbols,
- Number of lines used to write equation≥1 (it depends on the dimension of the given table), for h=29, one needs 3 lines,
- Time needed to find the harmonic sequence≥20 seconds, (it depends on the dimension of the given table), for h=29, one needs 26 seconds to find the answer,

The disadvantage of this method is that for large harmonics such as 60, 90, 120, etc, it becomes more difficult to find the answer if the table is limited for a few number of harmonics, and the time needed to find the answer will become much higher than the Simplified method.

4 INTRODUCING THE NTH FRACTION PART OF A REAL NUMBER

Any Real Number is divided into two parts, the first part is called the "Characteristic", which contains the left part of a Real Number before the Decimal Point, and the second part is called the "Fractional Part" (also called Mantissa), which contains the right part of a Real Number after the Decimal Point [15].

For example: the Real Number "12.584" has two parts:

1- The first Part is "12" located before the Decimal Point, which is called "Characteristic", it is a pure integer number.
2- The second part is "584" which is located after the Decimal Point, and it is called the Fractional Part or Mantissa.

Thus,

$$R = C + F \quad (32)$$

Where,
• R is the Real Number,
• C represents the Characteristic,
• F designs the Fractional Part.

The digits after Decimal Point have names, the first digit is called the "tenths digit", in the above case it is "5", the second digit is called the "hundreds digit", in the above case it is "8", and so on.

4.1 Proposed Extractor of Nth digit

For instance, in Mathematics, and to the best of the author's knowledge, there is no formula that extracts a Fractional Digit. In this paper, it is vital to extract the first Fractional Digit which is the "tenths digit" that will be used in the following sections.

So, this paper proposes an original way to extract the Nth Digit of a Fractional Part, and it is denoted in equation (33)

$$FP(n; A) \quad (33)$$

Where,
• FP is a Function which extracts the Nth Digit of the Fractional Part of a Real Number A.
• n designs the Nth Digit that will be extracted.
• A represents the Real Number.

Its programming is equivalent to

$$FP(n; A) = \mathrm{mod}(\mathrm{Int}(A \cdot 10^n), 10) \quad (34)$$

This is the notation using Microsoft Excel.

Or it can be also written as

$$FP(n; A) = \mathrm{Int}(A \cdot 10^n) \cdot \mathrm{mod}(10) \quad (35)$$

Where,
• $\mathrm{Int}(x)$ rounds a number (x) down to the nearest integer,
• $\mathrm{mod}(10)$ returns the remainder after a number ($\mathrm{Int}(A \cdot 10^n)$) is divided by advisor (10 in our case).

4.2 Example

Let's suppose $A = 13.4768901$

$FP(1; A) = 4$ which is the tenths digit
$FP(2; A) = 7$ which is the tenths digit
$FP(3; A) = 6$ which is the tenths digit
$FP(4; A) = 8$ which is the tenths digit
And so on.

5 PROPOSED METHOD FOR DETERMINING THE HARMONIC SEQUENCES

By using the proposed Infomath function developed in the paper [14], it will be much easier to determine the harmonic sequences with a very simple method comparing to the traditional one,

The equation is:

$$S = \underline{FP\left(1; \frac{h}{3}\right); 3, 6; (D), (I)/(H)} \quad (36)$$

Where,
• S is the sequence of the Harmonic,
• $FP(n; A)$ is a Function which extracts the Nth Digit of the Fractional Part of a Real Number A. In our case, $n = 1$, and $A = h/3$, which means one is interested to extract the Tenths Digit of $h/3$,
• D designs the Direct Sequence of the Harmonic,
• I is the Inverse Sequence of the Harmonic,
• H represents the Homopolar Sequence of the Harmonic.

5.1 How does the InfoMath Function work?

This subsection gives an idea about how the Infomath Function works, but for more details, one can refer to the paper [14].

The form of the InfoMath function is as following:

$$\underline{\text{input; conditions; output}} \quad (37)$$

It is divided into three parts;
1- The first part is the input, which can be a number, equation, or anything else.
2- The second part is the conditions which are applied to the input, if the conditions are verified; the output will give a certain value, expression, equation, or anything else.
3- The third part is the output, which can be a number, equation, or anything else.

Simple example:

$$S = \underline{3; \leq 1, \geq 6; (2), (5)/(8)}$$

In this example, the input is "3", so the conditions must be applied to this input,
1- 3≤1?? If yes the output should be "2"
2- If not, is 3≥6?? If yes the output should be "5"
3- Else, the output should be "8"

In the above case, the input doesn't verify the first two conditions, thus, the Infomath Function is replaced by the output "8", which means

$$S = \underline{3; \leq 1, \geq 6; (2), (5)/(8)} = 8$$

If the first condition is verified, the output will be the first item "2". If not, if the second condition is verified, the output will be the second item "5". If not, Else, the output will be automatically the third item which is "8".

Therefore, if the input is ≤ 1, the output is 2,

If the input is ≥ 6, the output is 5,

Else, the output is 8.

Now, returning to the equation (36),
The input is $FP\left(1; \frac{h}{3}\right)$,

The first condition is (=3), the second condition is (=6),

The outputs are: if the first condition is verified, the output is "D" which means Direct Sequence. If not, if the second condition is verified, the output is "I" which means Indirect Sequence. If not, Else, the output is "H", which means Homopolar Sequence.

5.2 Application of the equation

Now, returning to the main subject, the main idea of the Infomath function is simple, and it is much easier to be applied. Let's consider different values of harmonics. One has 3 harmonics on a transformer, a machine, or a network in which he wants to determine their sequences in order to see which one has a Direct Sequence, an Inverse Sequence, or a Homopolar Sequence. The three harmonics are:
$h_1 = 38, h_2 = 57, h_3 = 67$

By applying equation (36),
• $\frac{h_1}{3} = \frac{38}{3} = 12.6666$

The Tenths Digit is ≠3, but it is =6, thus, it is the Inverse Sequence, and
$S = \underline{6; 3, 6; (D), (I)/(H)} = I$

- $\frac{h_2}{3} = \frac{57}{3} = 19$

The Tenths Digit is ≠3 and ≠6, thus, it is the Homopolar Sequence, and

$$S = \underline{0; 3, 6; (D), (I)/(H)} = H$$

- $\frac{h_3}{3} = \frac{67}{3} = 22.3333$

The Tenths Digit is =3, thus, it is the Direct Sequence, and

$$S = \underline{3; 3, 6; (D), (I)/(H)} = D$$

In this way, the calculation is simplified to only 1 division, and one gets the answer rapidly without doing the whole traditional procedure which is described in section 3.

Now, if one has a large number of harmonics, it is much easier to determine what the sequence of each harmonic is.

5.3 Example

Considering the harmonic h=29 which is discussed in the previous sections. To calculate this harmonic using the proposed method, one uses the equation (36), therefore,

$$S = \underline{FP\left(1; \frac{h}{3}\right); 3, 6; (D), (I)/(H)}$$
$$= \underline{FP\left(1; \frac{29}{3}\right); 3, 6; (D), (I)/(H)}$$
$$= \underline{6; 3, 6; (D), (I)/(H)} = I$$

Therefore this harmonic has an inverse sequence.

5.4 Discussion

In order to complete this method for one harmonic, one needs,

- Number of operations = 1.
- Number of equal symbol "=" = 4,
- Number of lines used to write equation = 1,

Time needed to find the harmonic sequence = 28 seconds.

6 RESULTS AND DISCUSSIONS

In the previous sections, some examples are given and discussed briefly for each method. Therefore, it is necessary to make a table of comparison between these four methods. Table 2 presents a summary about the four used methods in this paper.

Table 2, A comparative table between all used methods in this paper.

Category	Mathematical Method	Simplified Method	Table Method	Proposed method
Needed time to obtain the answer	7.25 minutes	80 seconds	**≥20 seconds** In our example it took 36 seconds	**28 seconds**
Number of operations needed to obtain the answer	200	10	**≥0** In our example it took 3 operations	**1**
Number of lines used to obtain the answer	24	3	**≥1** In our example it took 3 lines	**1**
Programming difficulty	Very complex	medium	Medium	**simple**
Constraints	**Non**	**Non**	Limited by the size of the table	**Non**
Number of used equal symbols "="	33	9	≥1 For h=29 it takes 3 "="	**4**
Possibility of errors by hand calculation	Possible ≥ 0%	**0%**	**0%**	**0%**
Is the method practical	No	**Yes**	No	**Yes**
Complexity of the calculation	Very Complex	Easy	Easy	**Very Easy**
Mental Calculation can be used?	No	Yes but difficult	No	**Yes**

The advantages of each category are presented in bold red color, it is very clear that all advantages are presented in the proposed method in this paper.

The most important factor is the time needed to obtain an answer for one harmonic; hence, if one considers that the mathematical method is the standard method to do a comparison with, therefore, the reduced time ratio for each method is presented as below:

$$\text{Reduced Time Ratio}_{\text{Math method}} = \frac{435}{435} = 1$$

$$\text{Reduced Time Ratio}_{\text{Simplified method}} = \frac{435}{80} = 5.4375$$

$$\text{Reduced Time Ratio}_{\text{Table method}} \leq \frac{435}{20} = 21.75$$

$$\text{Reduced Time Ratio}_{\text{Proposed method}} = \frac{435}{28} = 15.537$$

7 CONCLUSION

In this paper, the harmonic sequences are defined, they are calculated using four methods, the first method is the mathematical method defined by Fortescue, it is considered as the basis of the harmonic decompositions. The second method is a simplified method, in which each sequence is defined by two simple equations; it is much simpler than the first one. The third method is presented in a table, where different harmonics are classified into three categories in the same table, which are the Direct, Inverse, and Homopolar sequences. The problem with this method is that the size of the table is limited to a certain number of harmonics, and it doesn't have a mathematical support. And finally, the fourth method which is proposed in this paper, it has many advantages over the traditional methods which are presented in table 2. The proposed method in this paper is recommended for students, and engineers who are willing to study the harmonics sequences of a certain system or equipment.

8 REFERENCES

1. Fortescue, C. L.: Method of Symmetrical Co-Ordinates Applied to the Solution of Polyphase Networks. Presented at the 34th annual convention of the AIEE (American Institute of Electrical Engineers) in Atlantic City, New Jersy, (28 July 1918).
2. Wakileh, G. J.: Harmonics in rotating machines. Electric Power Systems Research, Elsevier, Volume 66, Issue 1, pp. 31-37 (July 2003).
3. Hamdan, M. I. B.: Impact of rotating machines on harmonic distortion in power distribution systems. Faculty of Electrical & Electronic Engineering, Universiti Malaysia Pahang, pp. 1-25 (2009).
4. Dunfield, J. C., Barton, T. H.: Effect of m.m.f. and permeance harmonics in electrical machines, with special reference to a synchronous machine. Electrical Engineers, Proceedings of the Institution of , Volume 114, Issue 10 , pp. 1443-1450, (October 1967).
5. Robinson, R. B.: Harmonics in a.c. rotating machines. Proceedings of the IEE - Part C: Monographs, Volume 109, Issue 16 , pp. 380 – 387, (September 1962).
6. Fuchs, E. F., Masoum, M. A. S.: Power Quality in Power Systems and Electrical Machines. Elsevier Academic Press, ISBN: 978-0-12-369536-9, (2008).
7. Ludwinek, K., Siedlarz, A.: Harmonic Distortion analysis in armature currents of synchronous machine during co-operation with the power system. Zeszyty Problemowe-Maszyny Elektryczne number 84, (2009).
8. Said, D. M., Nor, K. M.: Effects of Harmonics on Distribution Transformers. Australasian Universities Power Engineering Conference (AUPEC'08), pp. 1-5, (2008).
9. Elmoudi, A., Lehtonen, M., Nordman, H.: Effect of harmonics on transformers loss of life, Electrical Insulation. Conference Record of the 2006 IEEE International Symposium on, Toronto, pp. 408-411, (June 2006).
10. Jayasinghe, N. R, Lucas, J. R., Perera, K. B. I. M.: Power System Harmonic Effects on Distribution Transformers and New Design Considerations for K Factor Transformers. IEE Sri Lanka Annual Sessions, pp. 1-7, (September 2003).
11. Zynal, H. I., Yass, A. A.,: The Effect of Harmonic Distortion on a Three phase Transformer Losses. Canadian Journal on Electrical and Electronics Engineering Vol. 3, No. 5, pp. 255-261, (May 2012).
12. Kocatepe, C., Yumurtacı, R., Arıkan, O., Baysal, M., Kekezoğlu, B., Bozkurt, A., Kumru, C. F.: Harmonic Effects of Power System Loads: An Experimental Study. Chapter 7, of the Book "Power Quality Issues", Edited by A. Zobaa, ISBN 978-953-51-1068-2, InTech, pp. 175-200, (April 17, 2013).
13. Tory K. J., Pope, R.: Eliminating harmonics from the facility power system. Power Transmission Design, pp. 43-46, (April 1997).
14. Bayeh, C. Z.: Introduction to the Infomath function in mathematics and its application in electrical and electronic engineering. WSEAS Transactions on Advances in Engineering Education, Issue 1, Volume 10, pp. 24-36, (January 2013).
15. Graham, R. L., Knuth, D. E., Patashnik, O.: Concrete Mathematics. Addison-Wesley Publishing Company, pp. 1-640, (1989).
16. Bouchard, R. P., Olivier, G. : Electrotechnique. 2e edition, Presses Internationales Polytechnique, ISBN: 978-2-553-00720-0, p. 391, (1999).

International Journal of Digital Information and Wireless Communications (IJDIWC) 6(2): 122-138
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

Modeling Performance of VoIP Traffic over 802.11 Wireless Mesh Network Under Correlated Inter-arrival Times

Barbara Kabwiga Asingwire[1], Okopa Michael[2], Tonny Bulega[3]
Busitema University-Uganda[1], International University of East Africa-Uganda[2],
Makerere University-Uganda[3],
kezabarbie@ymail.com[1], okopamike@gmail.com[2], tbulega@cit.mak.ug[3]

ABSTRACT

Wireless Mesh Networks (WMN) are an attractive communication paradigm for VOIP traffic due to their low cost and relative ease of deployment. In a recent study, the end to end delay of VOIP packets in a WMN were modeled using an M/D/1 queue based model. In this model, the arrival and service patterns is assumed to follow Poisson and deterministic distributions respectively. However, the long held paradigm in the communication and performance studies that voice and data traffic follow Poisson distribution and the service time deterministic is inaccurate and inefficient. In most cases the arrivals are correlated and the sizes of requests are not similar. In this study, we propose to overcome the above challenge by modeling queuing delay and average queue length of VoIP packets in WMN to capture correlated arrivals, general size distribution and multiple channels using the $MMPP/G/m$ queue model. The numerical results obtained from the derived models show that increasing the number of mesh routers, arrival rate and load leads to increase in queuing delay and queue length. In addition, it is observed that correlated arrivals exhibits a higher queuing delay and queue length as compared to Poisson arrivals especially at high values of mesh routers, arrival rates and load.

KEYWORDS

Correlated arrivals, general service time, Wireless Mesh Networks

1 INTRODUCTION

Voice over Internet Protocol (VoIP) is a technology that transports voice packets across packet switched networks using the Internet Protocol [1]. VoIP involves digitization of voice streams and transmitting the digital voice as packets over conventional IP-based packet networks like the Internet.

Wireless Mesh Networks (WMNs) often consist of mesh clients, mesh routers and gateways. The mesh clients may be laptops, cell phones and other wireless devices while the mesh routers forward traffic to and from the gateways which may, but need not, connect to the Internet [2].

WMNs are an attractive communication model because of the low cost and relative ease of deployment [3]. The capability of being self-healing, self-organized, and auto-configurable makes WMNs very reliable and resilient [3]. The Wireless Mesh Network backbone provides alternative paths between each pair of end points which increases reliability and eliminates the single point of failure within the mesh [4]. When one node fails, the rest of the nodes can still communicate with each other, directly or through one or more intermediate nodes.

The quality of a VoIP call in a WMN is affected by many parameters such as delay, packet loss, jitter, delay, etc. The common metric used to measure the performance of Wireless Mesh Networks is delay[5]. Delay is defined as the time taken by the VoIP packet to travel from a mobile node to gateway in a Wireless Mesh Network.

Delay can be measured in either one-way or round-trip delay. A typical VoIP traffic can tolerates up to 150ms end to end delay in a single direction before the quality of the call becomes unacceptable [6].

There are various challenges facing VOIP in WMN. A number of efforts have been carried out to investigate the performance of Void over Wireless Mesh Network [7, 8]. Chhabra et al. [5] modeled average delay for VoIP by varying number of mesh points in WMNs. The authors developed an $M/D/1$ queue based model to estimate the average delay incurred by the VOIP traffic. In an $M/D/1$ model, M stands for Markovian arrival with exponentially distributed inter arrival times, D stands for deterministic service times and 1(one) server.

The expressions for the average queuing delay incurred by VOIP packets from source mesh client to destination (Internet gateway) in a wireless mesh network having n mesh routers can be expressed as [5]:

$$Q_{nd} = s + n\lambda s^2/(2(1 - n\lambda s)) \qquad (1)$$

where

Q_{nd} is the average queuing delay taken by VOIP packets from the source client to the destination via n mesh routers in a network. s is the service rate of each VOIP packet.

However, the long held paradigm that voice traffic and data traffic are adequately characterized by certain Markovian models (e.g., Poisson) has been proven to be inaccurate, inefficient and greatly affects the accuracy of the results since real system workloads exhibit significant correlation patterns [9].

Correlation in traffic arrivals has been observed in compressed digital video streams [10], Ethernet traffic [11] and Web traffic between browsers and servers [12]. In addition, the study of social patterns in online social networks taking place on the World Wide Web shows that the inter-arrival times between subsequent tagging events can only be explained by taking into account correlation in users' behaviors [13].

Similarly, correlation of word inter-arrival times have been detected in texts [13]. Furthermore, the assumption of deterministic service times is unrealistic because the voice packets have varying packet sizes.

In this paper, we investigate the performance of VoIP over WMN using an $MMPP/G/m$ model that captures the correlated arrivals, general service times and multiple channels (servers).

The contribution of this paper is two fold. Firstly, this study proposes an analytical model for evaluating the performance of VoIP traffic in WMN in terms of queuing delay and queue length under correlated inter arrival times using $MMPP/G/m$ queue system.

Secondly, this study evaluated the performance of VoIP traffic in WMN in terms of queuing delay and queue length under correlated arrivals using $MMPP/G/m$ queue system. In this case, the number of mesh routers, average arrival rate, and load in the system are varied to determine the effect of these parameters on the queuing delay and queue length.

2 SYSTEM MODELS

We consider an $MMPP/G/m$ queue system to model the correlations between consecutive inter arrival times. In this case, the arrivals of requests are modeled using Markov Modulated Poisson Process $(MMPP)$. $(MMPP)$ is one of the most used models [14, 15, 16] to capture the characteristics of the incoming traffic such as correlated traffic, burstiness behavior, and long range dependency, and is simply a Poisson process whose mean value changes according to the evolution of a Markov Chain [17]. G stands for general service time distribution, meaning the service time can take on any distribution, e.g exponential or Bounded Pareto, and m denotes the number of servers.

MMPP is normally used for modeling bursty traffic owing to its ability to capture the time-varying arrival rate and the correlation between inter-arrival times while still maintaining

analytical tractability. MMPPs are most frequently observed in queuing theory, however it has also been applied to other applications, such as analysis of Web surfing behavior [18], and for telephone network fraud detection [19]. When there is correlation between inter arrival times as shown in Figure 1, M/M/1 queue system becomes inaccurate, therefore MMPP/G/m queue system is an optimal model to capture the correlation between the inter arrival times.

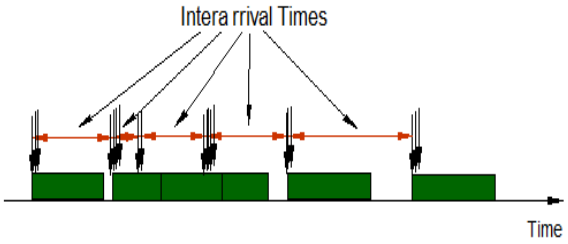

Figure 1. Correlated Inter arrival times (Batch arrivals). Adopted [20]

We model the Wireless Mesh Network using the $MMPP/G/m$ queue. G denotes general service time, and m the number of servers. The arrival process is assumed to follow a Markovian Modulated Poisson Process (MMPP). The service time is assumed to follow a general distribution. In this case the exponential and Bounded Pareto distributions are used as special cases of the general distribution. The System is assumed to consist of multiple servers, m. Two servers are used as a representation of multiple servers. Each mesh router is assumed to be a station in the queuing network representation.
The derived models are used to investigate the effect of the number of mesh routers, average arrival rate, and load in the system on the queuing delay and average queue length of VoIP packets in a WMN.
In the next section, we derive the expressions for average queuing delay and queue length for an $MMPP/G/m$ queue. We approximate the behavior of an $MMPP/G/m$ queue analytically. The method employed consists in approximating the $MMPP/G/m$ queue system using a weighted superposition of different $M/G/m$ queues.

3 EXPRESSIONS FOR THE PERFORMANCE METRICS

Consider an $MMPP/G/m$ queue, and let the $MMPP$ that models the incoming data traffic be composed by H states $(S_1...S_H)$.
We use the notation $M_i/G/m$ to refer to an $M/G/m$ queue whose average arrival rate is λ_i observed in state S_i and the service rate is μ and is constant among all the S_i states. The analytical approximations is based on the observation that if the MMPP stays in state S_i long enough without transiting to another state, the average waiting time at time t reach the same steady state observed for the corresponding $M_i/G/m$ queue. The values are pinned on the same steady state value of $M_i/G/m$ as long as the $MMPP$ does not change its state from S_i. Similar approximations have been used to approximate mean queue length and average response time for an $MMPP/M/1$ queue system [21].
The average packet queuing delay is an average evaluated over the number of incoming requests which are not distributed equally over time (during the state S_i the arrival rate of requests is λ_i, while during the state S_j the arrival rate of requests is a different amount λ_j. Therefore, the average packet queuing delay of the $MMPP/G/m$ is a weighted sum of the average packet queuing delays of the $Mi/G/m$ queues, expressed as:

$$Q = \sum_{i=1}^{H} Q_i r_i \qquad (2)$$

The weights (r_i) are not simply composed by the asymptotic probabilities of the $MMPP$ (as in the case of the average queue length) but it is scaled to keep into account the different arrival rate per each state. Hence;

$$r_i = \frac{p_i \lambda_i}{\sum_{j=1}^{H} p_j \lambda_j}.$$

3.1 Expression for average packet queuing delay under MMPP/M/m

The expression for average packet queuing delay for $M/M/m$ which is a special case of $M/G/m$ is derived as follows: When there are more than one server, a new customer arriving may enter any of the free servers. Let λ and μ be the rate of the Poisson process for the arrivals and the parameter of the exponential distribution for the service times respectively.
The balance equations are:

$$\lambda P_{i-1} = \begin{cases} i\mu P_i, & \text{for } i \leq m \\ m\mu P_i, & \text{for } i > m \end{cases} \quad (3)$$

$$P_i = \begin{cases} \frac{P_o (m\rho)^i}{i!}, & \text{for } i \leq m \\ \frac{P_o m^m \rho^i}{m!}, & \text{for } i > m \end{cases} \quad (4)$$

where $\rho = \frac{\lambda}{m\mu}$. The probability that an arriving customer is forced to join the queue is given by
$P_Q = \sum_{i=m}^{\infty} P_i = \sum_{i=m}^{\infty} \frac{P_o m^m \rho^i}{m!}$
$= \frac{P_o m^m \rho^m}{m!} \sum_{i=m}^{\infty} \rho^{i-m} = \frac{P_o (m\rho)^m}{m!(1-\rho)}$
where

$$P_o = [\sum_{i=o}^{m-1} \frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!(1-\rho)}]^{-1}$$

The average number of customers in the queue is given by
$N_Q = \sum_{i=m}^{\infty} (i-m) P_i = \sum_{i=o}^{\infty} i P_{i+m}$
$= \sum_{i=o}^{\infty} i P_o \frac{m^m \rho^{i+m}}{m!} = \frac{P_o (m\rho)^m}{m!} \sum_{i=o}^{\infty} i\rho^i = \frac{P_o (m\rho)^m}{m!} \cdot \frac{\rho}{(1-\rho)^2}$
Finally obtaining the average queue length as:

$$N_Q = \frac{P_o (m\rho)^m}{m!} . \frac{\rho}{(1-\rho)^2} \quad (5)$$

Hence, the average packet queuing delay is given using little's law [22].

$$Q_d = \frac{N_Q}{\lambda} = \frac{P_o (m\rho)^m}{m!} . \frac{\rho}{\lambda(1-\rho)^2} \quad (6)$$

The expression for average packet queuing delay for two states for $MMPP/M/m$ is given as:

$$Q_d = \sum_{i=1}^{2} r_i . \left(\frac{P_o (m\rho)^m}{m!} . \frac{\rho}{\lambda(1-\rho)^2} \right) \quad (7)$$

where $r_i = \frac{p_i \lambda_i}{\sum_{j=1}^{2} p_j \lambda_j}$ and $P_o = [\sum_{i=o}^{m-1} \frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!(1-\rho)}]^{-1}$.
The expression for total average packet queuing delay for two states for $MMPP/M/m$ is given as:

$$Q_d = E[x] + \sum_{i=1}^{2} r_i . \left(\frac{P_o (m\rho)^m}{m!} . \frac{\rho}{n\lambda(1-\rho)^2} \right) \quad (8)$$

where $\rho = \frac{n\lambda}{\mu}$, n is the number of mesh clients and $E[x]$ the service time for each packet.

3.2 Expression for average packet queuing delay under MMPP/BP/m

The expression for average packet queuing delay for $MMPP/BP/m$ can be derived from the general expression for $MMPP/G/m$. However, the average packet delay for $MMPP/G/m$ queue system can be got by approximating the $MMPP/G/m$ queue system using a weighted superposition of different $M/G/m$ queues. There are numerous approximations for the average delay a job experiences under the $M/G/m$ queue system [25, 23, 24]. A naturally refined heavy-traffic approximation exploiting the exact $M/M/m$ results is given in [28] as:

$$E[W^{M/G/m}] = \frac{C^2 + 1}{2} E[W^{M/M/m}] \quad (9)$$

where C^2 is the coefficient of variation of the service time distribution, $E[W^{M/G/m}]$ is the mean delay under M/G/m queue system, and $E[W^{M/M/m}]$ is the mean delay under M/M/m queue system.
The expression for the average packet delay under $MMPP/G/m$ can be deduced as

follows: From equation 6, the average packet delay under M/G/m is given as:

$$Q_d = \frac{(C^2+1)}{2}\frac{P_o(m\rho)^m}{m!}.\frac{\rho}{\lambda(1-\rho)^2} \quad (10)$$

The expression for average packet queuing delay for two states for $MMPP/G/m$ is given as:

$$Q_d = \sum_{i=1}^{2} r_i.\frac{(C^2+1)}{2}\frac{P_o(m\rho)^m}{m!}.\frac{\rho}{\lambda(1-\rho)^2} \quad (11)$$

where $r_i = \frac{p_i\lambda_i}{\sum_{j=1}^{2} p_j\lambda_j}$ and $P_o = [\sum_{i=o}^{m-1}\frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!(1-\rho)}]^{-1}$.
The expression for total average packet queuing delay for two states for $MMPP/G/m$ is given as:

$$Q_d = E[x] + \sum_{i=1}^{2} r_i.\frac{(C^2+1)}{2}\frac{P_o(m\rho)^m}{m!}.\frac{\rho}{\lambda(1-\rho)^2} \quad (12)$$

where $\rho = \frac{n\lambda}{\mu}$, n is the number of mesh clients and $E[x]$ the service time for each packet.
In the next section, we derive the expressions for average packet queue length for $MMPP/M/m$

3.3 Expression for average packet queue length under MMPP/M/m

The evolution of the average packet queue length value of the $MMPP/G/1$ can be described as follows: each time a state transition occurs there is a transient phase (t_{12} for a transition from S_1 to S_2 and t_{21} for a transition from S_2 to S_1) after which the mean queue length of the $MMPP$ reaches the steady state of the corresponding $Mi/G/1$. Denote;

(i) Q_i to be the steady state queue length of $Mi/G/1$ queue.

(ii) p_i to be the asymptotic probability for the $MMPP$ to stay in state S_i.

The mean queue length of the $MMPP/G/1$ queue can be approximated as:

$$Q = \sum_{i=1}^{2} p_i Q_i \quad (13)$$

The expression for the average queue length under MMPP/G/1 can be expressed as:

$$E[N_q] = \sum_{i=1}^{2} p_i \frac{\lambda_i^2\overline{x^2}}{2(1-\rho_i)} \quad (14)$$

The expression for the average queue length of MMPP/M/m can then be derived as follows:
The average queue length for an M/M/m queue system is given in equation 5 as:

$$N_q = \frac{P_o(m\rho)^m}{m!}.\frac{\rho}{(1-\rho)^2} \quad (15)$$

Therefore, the expression for average packet queue length for two states for $MMPP/M/m$ is given as:

$$N_q = \sum_{i=1}^{2} p_i.\left(\frac{P_o(m\rho)^m}{m!}.\frac{\rho}{(1-\rho)^2}\right) \quad (16)$$

where $P_o = [\sum_{i=o}^{m-1}\frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!(1-\rho)}]^{-1}$.

3.4 Expression for average packet queue length under MMPP/BP/m

The expression for average packet queue length for $MMPP/BP/m$ can be derived from the general expression for average queue length of $MMPP/G/m$. Since queuing delay and queue length change directly proportionately, we can derive the expression for average queue length using a similar approach as in equation 9. Hence, the average queue length under $M/G/m$ can be expressed as:

$$E[N^{M/G/m}] = \frac{C^2+1}{2}E[N^{M/M/m}] \quad (17)$$

where C^2 is the coefficient of variation of the service time distribution. Hence the expression for the average queue length under $M/G/m$ can be deduced as follows:

$$N_q = \frac{(C^2+1)}{2}\frac{P_o(m\rho)^m}{m!}.\frac{\rho}{(1-\rho)^2} \quad (18)$$

Therefore, the expression for average packet queue length for two states for $MMPP/G/m$ is given as:

$$N_q = \sum_{i=1}^{2} p_i . \frac{(C^2+1)}{2} \frac{P_o(m\rho)^m}{m!} . \frac{\rho}{(1-\rho)^2} \quad (19)$$

where $P_o = [\sum_{i=o}^{m-1} \frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!(1-\rho)}]^{-1}$.

4 PERFORMANCE EVALUATION

In this section, we evaluate the performance of M/D/1 and MMPP/G/1 using MATLAB and present numerical results. The analysis is extended to the comparison of M/D/m, MMPP/M/m and MMPP/BP/m. The results nicely illustrate the impact of correlated arrivals as compared to Poisson arrivals.

4.1 Job Size Distributions under Consideration

Two job size distributions are considered, that is, the exponential and Bounded Pareto distribution. Exponential distribution is used to model the service time. Exponential distributed has a low variability since its coefficient of variation is 1. Exponential distribution models service time of requests with similar sizes. The pdf (probability density function) of an exponential distribution is given as: $f(x) = \mu e^{-\mu x}, \quad x \geq 0, \quad \mu \geq 0$, where μ is the mean service rate. Bounded Pareto (BP) distribution is commonly used in analysis because it has a high Coefficient of Variation and therefore it can exhibit the high variability property as observed in the Internet traffic and also because the maximum job size can be set to mimic the largest observed Internet flow sizes [22]. Denote the Bounded Pareto distribution by $BP(k, P, \alpha)$ where k and P are the minimum and the maximum job sizes and α is the exponent of the power law. The probability density function of the Pareto is given in [22] as

$$f(x) = \frac{\alpha k^\alpha}{1-(k/P)^\alpha} x^{-\alpha-1}.$$

$$k \leq x \leq P, \quad 0 \leq \alpha \leq 2.$$

4.2 Model Parameters

There are many codecs available for digitizing speech. The three commonly used codec in Internet telephony are G.711, G.723.1 and G.729 [26]. We consider parameters used for G.711 codec because G.711 codec gives the best voice quality, introduces the least delay, and is less sensitive than other codecs to packet loss. G.711 codec uses service time of 0.01 seconds. In addition G.711 uses a 100ms sample period corresponding to a maximum arrival rate of 10 packets per second. The $MMPP$ parameters are set on the basis of the results reported in [27], which has shown, via real traces analysis, the feasibility to model incoming traffic to a GRID server. According to the data reported in [27], the incoming data traffic of the analyzed GRID server can be modeled by a 2-state MMPP model. The transition probability p_1 from state S_1 to state S_2, is 0.17, while the reciprocal transition probability p_2, from state S_2 to state S_1, is 0.08.
The hypothetical parameters used in the analysis is consistent with parameters used in literature [5, 28, 26, 27]. Number of mesh routers used is 10, service time is 0.1 seconds, transition states is two, arrival rate varies from 5 to 10 packets/second, $p_1 = 0.17, p_2 = 0.08$, $\lambda_1 = 22.10$ requests/second, $\lambda_2 = 7.16$ requests/second, Squared coefficient of variation for BP distribution, $C^2 = 0.53$ [28]. Mean for Bounded Pareto distribution, $E[X] = 2.56$ [22].

4.3 Evaluation of queuing delay for M/D/1 and MMPP/M/1

We compare the performance of M/D/1 against MMPP/M/1 in terms of queuing delay. The results nicely illustrate the impact of correlated arrivals as compared to Poisson arrivals in terms of queuing delay as the performance metric.

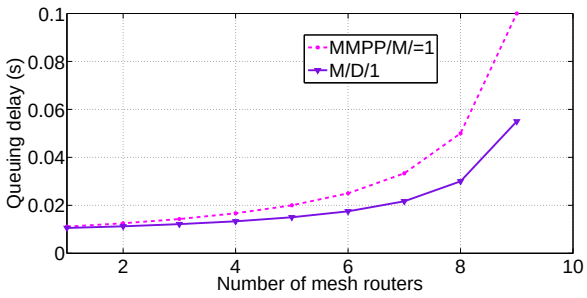

Figure 2. Queuing delay as a function of number of mesh routers

Figure 2 shows a graph of queuing delay as a function of number of mesh routers. We used equations 1 and 8 to plot the graph. It can be observed from figure 2 that queuing delay increases with increase in the number of mesh routers. The increase in queuing delay as the number of mesh routers increase can be explained by the fact that at each mesh router packets experience queuing delay and this accumulates as the number of mesh routers increase. We further observe that for low number of mesh routers, the queuing delay for MMPP/M/1 queue system and M/D/1 queue system are almost the same, however as the number of mesh routers increase the queuing delay for MMPP/M/1 queue system is higher than for M/D/1 queue system. The difference in queuing delay between MMPP/M/1 queue system and M/D/1 queue system is more pronounced for higher numbers of mesh routers.

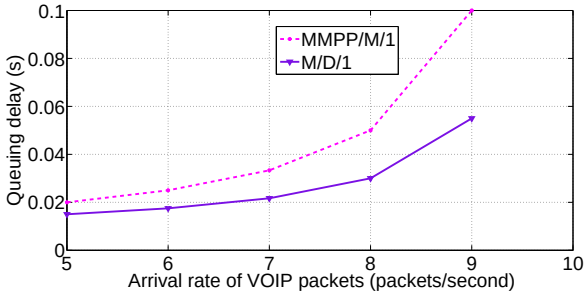

Figure 3. Queuing delay as a function of arrival rate of VoIP packets

Figure 3 shows a graph of queuing delay as a function of arrival rate of VoIP packets. We used equations 1 and 8 to plot the graph in figure 3. It can be observed from figure 3 that queuing delay increases with increase in arrival rate of VoIP packets regardless of the queue system. The increase in queuing delay as the arrival rate of VoIP packets increase can be explained by the fact that increase in arrival rate leads to increase in the number of packets at each mesh router hence increasing the queuing delay. We also observe that queuing delay for MMPP/M/1 queue system is higher than the queuing delay for M/D/1 queue system for all considered arrival rates. The difference in queuing delay between MMPP/M/1 queue system and M/D/1 queue system is observed to be higher for higher arrival rates of VoIP packets.

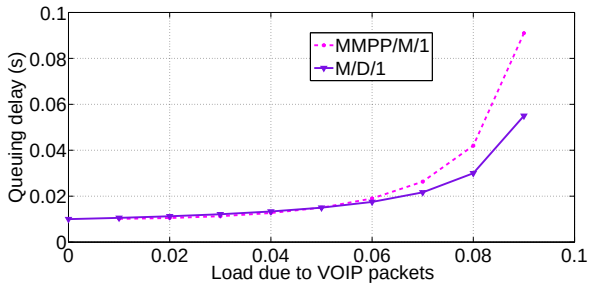

Figure 4. Queuing delay as a function of load due to VoIP packets

Figure 4 shows a graph of queuing delay as a function of load due to VOIP packets. We used equations 1 and 8 to plot the graph in figure 4. It can be observed from figure 4 that queuing delay generally increases with increase in load due to VOIP packets regardless of the queue system. Increase in load leads to increase in the number of packets in the system which in turn leads to increased queuing delay. We also observe that queuing delay for MMPP/M/1 queue system is slightly lower than the queuing delay for M/D/1 queue system for load values less than 0.05, however for load values higher than

0.05, the queuing delay for MMPP/M/1 queue system is higher than for M/D/1 queue system. In addition, the difference in queuing delay between MMPP/M/1 queue system and M/D/1 queue system is observed to be higher for higher load values of VOIP packets.

5 EVALUATION OF QUEUING DELAY FOR MMPP/M/1 AND MMMPP/M/2

In this section, we compare the performance of MMPP/M/1 against MMPP/M/2 in terms of queuing delay. The results nicely illustrate the impact of increasing the number of servers on the performance of VOIP packets under correlated arrivals in terms of queuing delay as a performance metric.

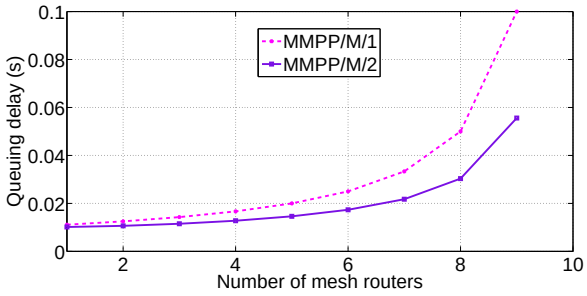

Figure 5. Queuing delay for MMPP/M/m as a function of number of mesh routers

Figure 5 shows a graph of queuing delay as a function of number of mesh routers for MMPP/M/m queue system. We used equation 8 to plot graph 5. We observe from figure 5 that queuing delay generally increases with increase in number of mesh routers regardless of the number of servers. We also observe that initially queuing delay for MMPP/M/1 and MMPP/M/2 queue systems are the same, however as the number of mesh routers increase, the queuing delay for MMPP/M/1 is higher than for MMPP/M/2. Queuing delay for MMPP/M/1 is higher than for MMPP/M/2 due to the fact that MMPP/M/2 has higher number of servers as compared to MMPP/M/1. The more the number of servers, the higher the processing power and hence the lower the queuing delay. The difference in queuing delay for MMPP/M/1 and MMPP/M/2 queue systems is more pronounced as the number of mesh routers increase.

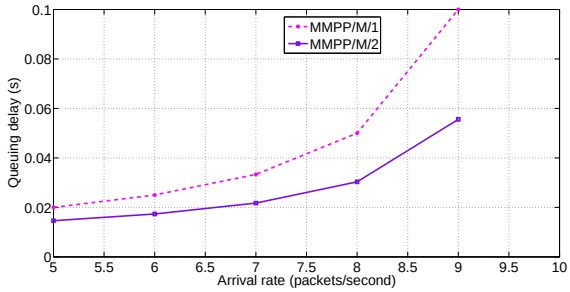

Figure 6. Queuing delay for MMPP/M/m as a function of arrival rate of VOIP packets

Figure 6 shows a graph of queuing delay as a function of arrival rates for MMPP/M/m queue system. We used equation 8 to plot graph 6. It can be observed from figure 6 that queuing delay generally increases with increase in arrival rate irrespective of the number of servers. We also observe that queuing delay for MMPP/M/1 queue system is higher than for MMPP/M/2 queue system for all considered arrival rates. Queuing delay for MMPP/M/1 is higher than for MMPP/M/2 due to the fact that MMPP/M/2 has higher number of servers as compared to MMPP/M/1. The more the number of servers, the higher the processing power and hence the lower the queuing delay. The difference in queuing delay for MMPP/M/1 and MMPP/M/2 queue systems increase as the arrival rate increases.

Figure 7 shows a graph of queuing delay as a function of load for MMPP/M/m queue system. We used equation 8 to plot graph 7. It can be observed from figure 7 that queuing delay generally increases with increase in load due to VOIP packets regardless of the number of servers. We also observe that for low load values the queuing delay for MMPP/M/1 and MMPP/M/2 queue systems are the same,

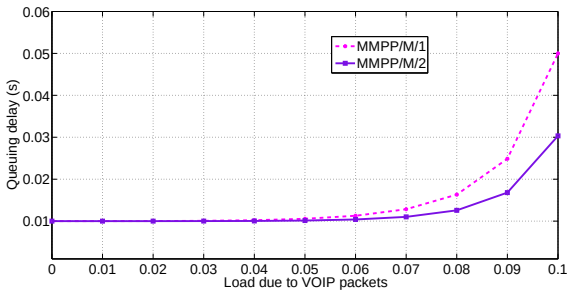

Figure 7. Queuing delay for MMPP/M/m as a function of load due to VOIP packets

however as the load increases, queuing delay for MMPP/M/1 queue system is higher than for MMPP/M/2 queue system. There is a marked difference in queuing delay for MMPP/M/1 and MMPP/M/2 queue systems as the load increases. Queuing delay for MMPP/M/1 is still higher than for MMPP/M/2 due to the fact that MMPP/M/2 queue system has higher number of servers as compared to MMPP/M/1 and hence higher processing power thereby lowering the queuing delay.

6 EVALUATION OF QUEUING DELAY FOR M/D/1 AND MMPP/BP/1

In this section, we compare the performance of M/D/1 against MMPP/BP/1 which is a special case of MMPP/G/1 in terms of queuing delay. The results illustrate well the impact of correlated arrivals as compared to Poisson arrivals in terms of queuing delay.

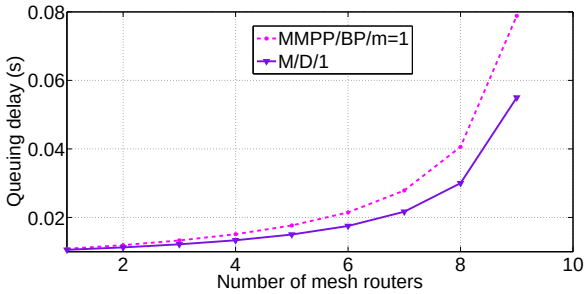

Figure 8. Queuing delay as a function of number of mesh routers

Figure 8 shows a graph of queuing delay as a function of number of mesh routers for MMPP/BP/1 and M/D/1 queue systems. We used equations 1 and 12 to plot graph 8. It can be observed that queuing delay generally increases with increase in number of mesh routers regardless of the queue system. We also observe that for low values of number of mesh routers, the queuing delay for MMPP/M/1 and M/D/1 queue systems are much closer, however as the number of mesh routers increases, queuing delay for MMPP/M/1 is higher than for M/D/1 queue system. Furthermore, we observe that there is a marked difference in queuing delay for MMPP/M/1 and M/D/1 queue systems as the number of mesh routers increase.

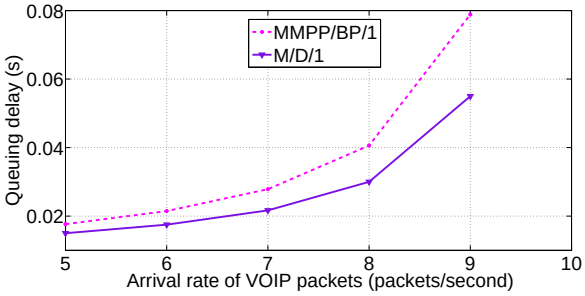

Figure 9. Queuing delay as a function of arrival rate of VoIP packets

Figure 9 shows a graph of queuing delay as a function of arrival rate of packets for MMPP/BP/1 and M/D/1 queue systems. We used equations 1 and 12 to plot the graph in figure 9. It can be observed from figure 9 that queuing delay generally increases with increase in arrival rate of packets regardless of the queue system. We also observe that for low arrival rates, the queuing delay for MMPP/BP/1 and M/D/1 queue systems are much closer, however as the arrival rate increases, queuing delay for MMPP/BP/1 is higher than for M/D/1 queue system. Furthermore, we observe that there is a marked difference in queuing delay for MMPP/BP/1 and

M/D/1 queue systems as the arrival rate increase.

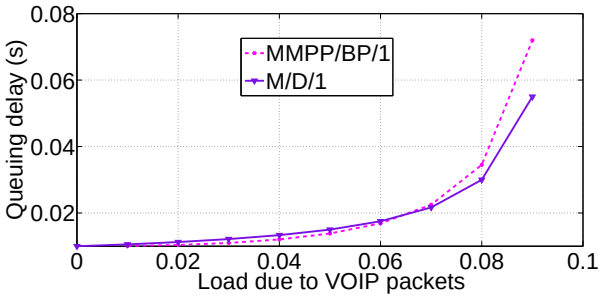

Figure 10. Queuing delay as a function of load due to VOIP packets

Figure 10 shows a graph of queuing delay as a function of load due to packets for MMPP/BP/1 and M/D/1 queue systems. We used equations 1 and 12 to plot the graph in figure 10. We observe from figure 10 that queuing delay increases with increase in load without regard to the queue system. We also observe that for low load values, the queuing delay for MMPP/BP/1 is lower than the queuing delay for M/D/1 queue system, however after the load of approximately 0.065 the queuing delay for M/D/1 queue system is lower than for MMPP/BP/1 queue system. It can further be noted that the difference in queuing delay between MMPP/BP/1 and M/D/1 queue systems are much higher as the load increases.
Next, we evaluate the average queuing delay for MMPP/BP/1 and MMMPP/BP/2 queue systems.

7 EVALUATION OF QUEUE DELAY FOR MMPP/BP/1 AND MMMPP/BP/2

We evaluate average queuing delay for MMPP/BP/1 and MMMPP/BP/2 queue systems to investigate the effect of increasing number of servers on the MMPP/BP/m, queue system.
Figure 11 shows a graph of queuing delay as a function of number of mesh routers for MMPP/BP/m queue system. We used

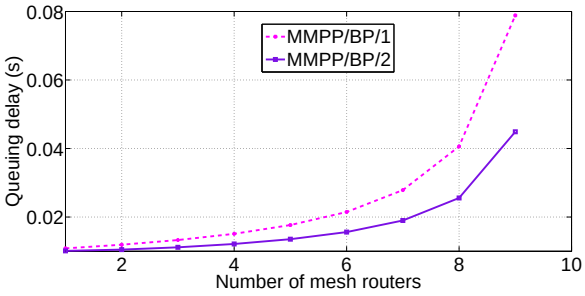

Figure 11. Queuing delay for MMPP/BP/m as a function of number of mesh routers

equation 12 to plot graph 11. We observe from figure 11 that queuing delay generally increases with increase in number of mesh routers irrespective of the queue system. We also observe that for low number of mesh routers the queuing delay for MMPP/BP/1 and MMPP/BP/2 queue systems are closer, however as the number of mesh routers increase, queuing delay for MMPP/BP/1 queue system is higher than for MMPP/BP/2 queue system. There is a marked difference in queuing delay between MMPP/BP/1 and MMPP/BP/2 queue systems as the number of mesh routers increase. Queuing delay for MMPP/BP/1 is higher than for MMPP/BP/2 due to the fact that MMPP/BP/2 has higher number of servers as compared to MMPP/BP/1 and hence higher processing power thereby lowering the queuing delay.

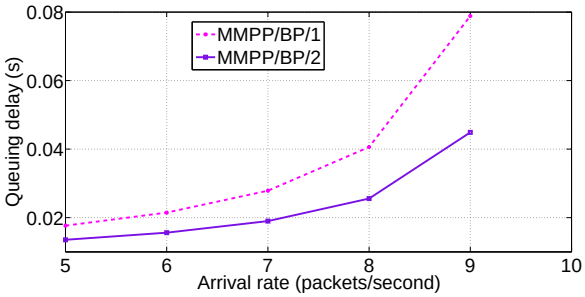

Figure 12. Queuing delay for MMPP/BP/m as a function of arrival rate of VoIP packets

Figure 12 shows a graph of queuing delay as a function of arrival rate of VoIP packets for MMPP/M/m queue system. We used equation 12 to plot graph 12. We observe from figure 12 that queuing delay increases with increase in arrival rate of packets into the system. We also observe that for low arrival rate values, the queuing delay for MMPP/BP/1 and MMPP/BP/2 queue systems are closer, however as the arrival rate increases, queuing delay for MMPP/BP/1 queue system is higher than for MMPP/BP/2 queue system. This trend is due to the fact that MMPP/BP/2 has a higher number of servers than MMPP/BP/1 and therefore higher processing rate that results in lower queuing delay. We further observe that as the arrival rate increases, the difference in queuing delay between MMPP/M/1 and MMPP/M/2 increases.

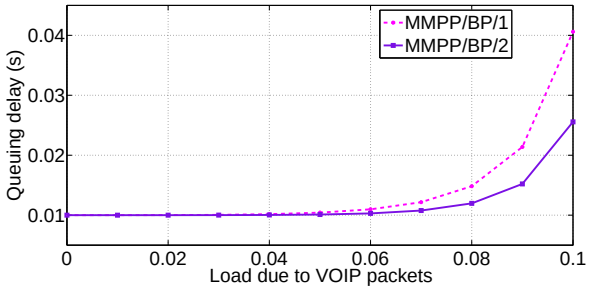

Figure 13. Queuing delay for MMPP/BP/m as a function of load due to VoIP packets

Figure 13 shows a graph of queuing delay as a function of load due to VOIP packets for MMPP/M/m queue system. We used equation 12 to plot graph 13. We observe from figure 13 that queuing delay increases with increase in load. We also observe that at low load the queuing delay for MMPP/BP/1 and MMPP/BP/2 queue systems are the same, however as the load increases, queuing delay for MMPP/BP/1 queue system is higher than for MMPP/BP/2 queue system. We further observe that as the load increases, the difference in queuing delay between MMPP/BP/1 and MMPP/BP/2 increases.

In the next section, we evaluate the average queue length for M/D/1 and MMPP/M/1 queue systems.

8 EVALUATION OF AVERAGE QUEUE LENGTH FOR M/D/1 AND MMPP/M/1

In this section, we compare the performance of M/D/1 against MMPP/M/1 in terms of queue length. Queue length is one of the performance metrics that are commonly used to evaluate the performance of systems.

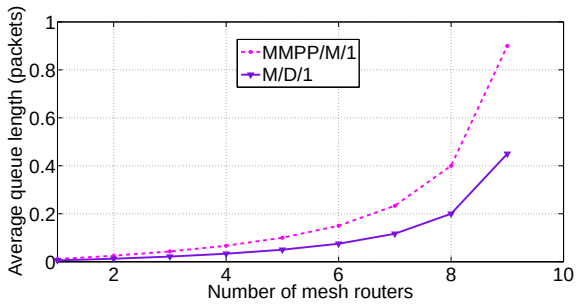

Figure 14. Average queue length as a function of number of mesh routers

Figure 14 shows a graph of queue length as a function of number of mesh routers. We used equations 1 and 19 to plot the graph. It can be observed from figure 14 hat average queue length increases with increase in the number of mesh routers regardless of the queue system. The increase in queue length as the number of mesh routers increase can be explained by the fact that at each mesh router packets queue hence increasing the queue length as the number of mesh routers increase. We further observe that for low number of mesh routers, the average queue length for MMPP/M/1 queue system and M/D/1 queue system are almost the same, however as the number of mesh routers increase the average queue length for MMPP/M/1 queue system is higher than for M/D/1 queue system. The difference in average queue length between MMPP/M/1 queue system and M/D/1 queue system is more pronounced as the number of mesh routers increase.

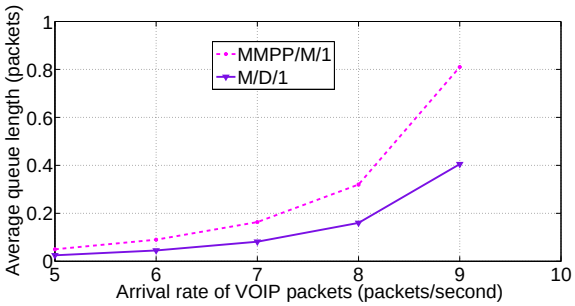

Figure 15. Average queue length as a function of arrival rate of VoIP packets

Figure 15 shows a graph of queue length as a function of arrival rates for VOIP packets. We used equations 1 and 19 to plot the graph. We observe from figure 15 that average queue length generally increases with increase in arrival rate of VOIP packets irrespective of the queue system. The increase in queue length as the arrival rate of VOIP packets increase can be explained by the fact that at each mesh router packets queue hence increasing the queue length as the number of mesh routers increase. We further observe that the average queue length for MMPP/M/1 is higher than for M/D/1 queue system regardless of the average arrival rate. The difference in average queue length between MMPP/M/1 queue system and M/D/1 queue system is more pronounced as the average arrival rate increases.

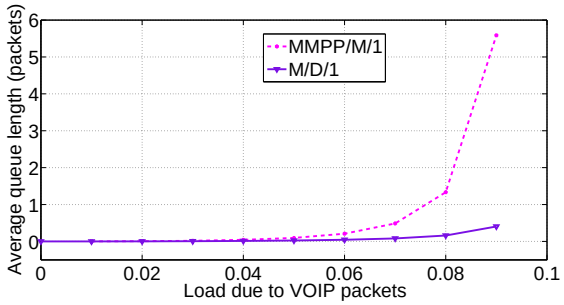

Figure 16. Average queue length as a function of load due to VoIP packets

Figure 16 shows a graph of queue length as a function of load due to VOIP packets. We used equations 1 and 19 to plot the graph. We observe from figure 16 that average queue length generally increases with increase in load due to VoIP packets regardless of the queue system. The increase in queue length as the load due to VoIP packets increase can be explained by the fact that if load increases more time is taken processing packets and this in turn leads to an increase in the queue length. We further observe that for low load values the queue length for MMPP/M/1 and M/D/1 queue systems are the same, however as the load increases, the queue length for MMPP/M/1 queue system is higher than for M/D/1 queue system. There is an observed marked difference in queue length for MMPP/M/1 and MMPP/M/2 queue systems as the load increases.

9 EVALUATION OF AVERAGE QUEUE LENGTH FOR MMPP/M/1 AND MMPP/M/2

In this section, we compare the performance of MMPP/M/1 against MMPP/M/2 in terms of queue length to investigate the effect of increasing number of servers on the average queue length.

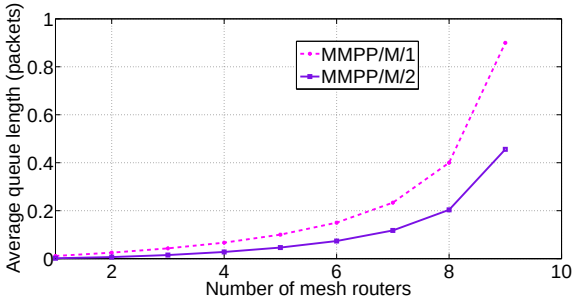

Figure 17. Average queue length as a function of number of mesh routers for MMPP/M/1 and MMPP/M/2

Figure 17 shows a graph of queue length as a function of number of mesh routers for MMPP/M/1 and MMPP/M/2 queue systems. We used equation 19 to plot the graph. We observe from figure 17 that average queue length generally increases with increase in number of

mesh routers irrespective of the queue system. The increase in queue length as the number of mesh routers increase can be explained by the fact that at each mesh router packets queue hence increasing the queue length as the number of mesh routers increase. We also observe that for low number of mesh routers the queue length for MMPP/M/1 and MMPP/M/2 queue systems are much closer, however as the number of mesh routers increase, the queue length for MMPP/M/1 queue system is higher than for MMPP/M/2 queue system. The difference in queue length between MMPP/M/1 and MMPP/M/2 is even higher as the number of mesh routers increase.

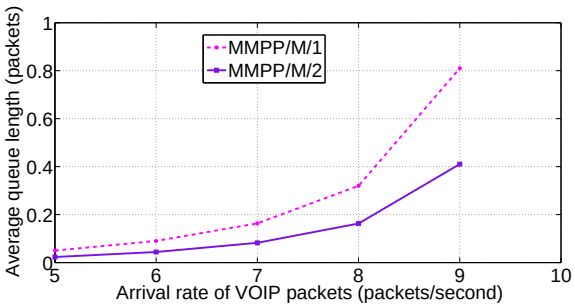

Figure 18. Average queue length as a function of arrival rate of VoIP packets for MMPP/M/1 and MMPP/M/2

Figure 18 shows a graph of queue length as a function of arrival rate for MMPP/M/1 and MMPP/M/2 queue systems. We used equation 19 to plot the graph. We observe from figure 18 that average queue length generally increases with increase in arrival rate irrespective of the queue system. The increase in queue length as the number of mesh routers increase can be explained by the fact that at each mesh router packets queue hence increasing the queue length as the number of mesh routers increase. We further observe that the average queue length for MMPP/M/1 queue system is higher than the average queue length for MMPP/M/2 queue system for all considered arrival rates. The difference in queue length between MMPP/M/1 and MMPP/M/2 is even higher as the arrival rate increases.

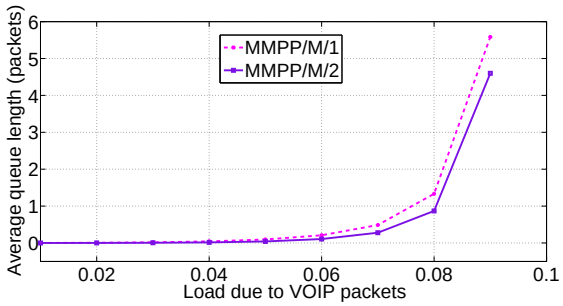

Figure 19. Average queue length as a function of load due to VoIP packets for MMPP/M/1 and MMPP/M/2

Figure 19 shows a graph of queue length as a function of load due to VoIP packets for MMPP/M/1 and MMPP/M/2 queue systems. We used equation 19 to plot the graph. We observe from figure 19 that average queue length generally increases with increase in load due to VoIP packets. At low load values, the average queue length for MMPP/M/1 and MMPP/M/2 are the same, however as the load increases, the average queue length of MMPP/M/1 is higher than average queue length for MMPP/M/2. The difference in queue length between MMPP/M/1 and MMPP/M/2 increases with increase in load after attaining a load of 0.04.

In the next section, we evaluate the average queue length for M/D/1 and MMPP/BP/1.

10 EVALUATION OF AVERAGE QUEUE LENGTH FOR M/D/1 AND MMPP/BP/1

In this section, we compare the performance of M/D/1 against MMPP/BP/1 in terms of queue length.

Figure 20 shows a graph of queue length as a function of number of mesh routers. We used equations 1 and 12 to plot the graph in figure 20. We observe from figure 20 that average queue length generally increases with increase in number of mesh routers. We further

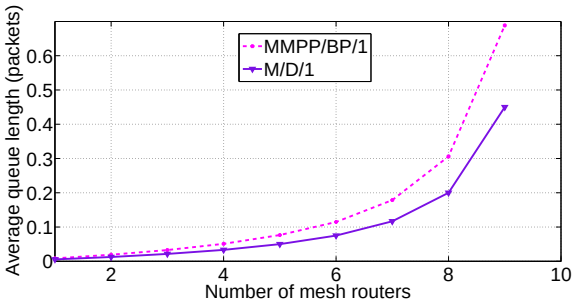

Figure 20. Average queue length as a function of number of mesh routers

observe that for lower number of mesh routers, the average queue length for MMPP/BP/1 and M/D/1 are closer, however as the number of mesh routers increase, the average queue length of MMPP/BP/1 is higher than average queue length for M/D/1.

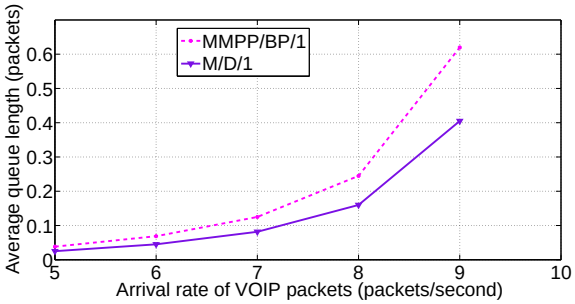

Figure 21. Average queue length as a function of arrival rate of VoIP packets

Figure 21 shows a graph of queue length as a function of arrival rate of VOIP packets. We used equations 1 and 12 to plot the graph in figure 21. We observe from figure 21 that average queue length generally increases with increase in arrival rate of VoIP packets. We further observe that at lower arrival rates of packets into the system, the average queue length for MMPP/BP/1 and M/D/1 are closer, however as the arrival rate increases, the average queue length of MMPP/BP/1 is higher than average queue length for M/D/1.
Figure 22 shows a graph of queue length as a function of load due to VoIP packets. We used

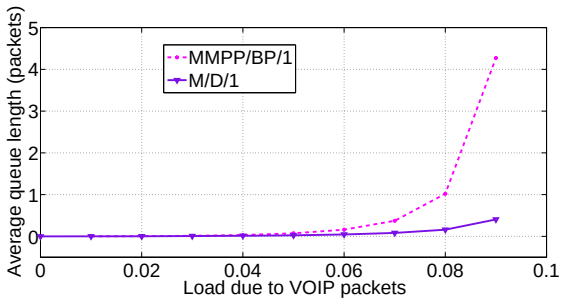

Figure 22. Average queue length as a function of load due to VoIP packets

equations
refaba and 12 to plot the graph in figure 22. We observe from figure 22 that average queue length increases with increase in load due to VOIP packets. We further observe that at lower load values, the average queue length for MMPP/BP/1 and M/D/1 are the same, however as the load increases, the average queue length of MMPP/BP/1 is higher than average queue length for M/D/1.

11 EVALUATION OF AVERAGE QUEUE LENGTH FOR MMPP/BP/1 AND MMPP/BP/2

In this section, we compare the performance of MMPP/BP/1 against MMPP/BP/2 in terms of queue length to investigate the effect of increasing number of servers on the average queue length.

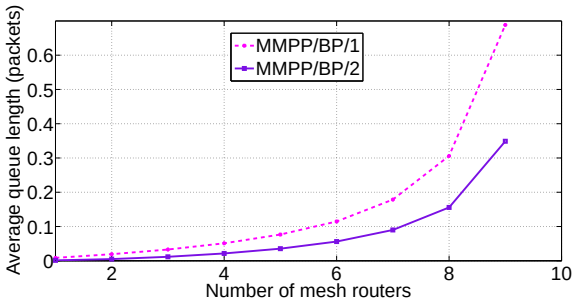

Figure 23. Average queue length as a function of number of mesh routers for MMPP/BP/1 and MMPP/BP/2

Figure 23 shows a graph of queue length as a function of number of mesh routers. We used equation 12 to plot graph 23. We observe from figure 23 that average queue length generally increases with increase in number of mesh routers. We further observe that for lower number of mesh routers, the average queue length for MMPP/BP/1 and M/D/1 are closer, however as the number of mesh routers increase, the average queue length of MMPP/BP/1 is higher than average queue length for M/D/1.

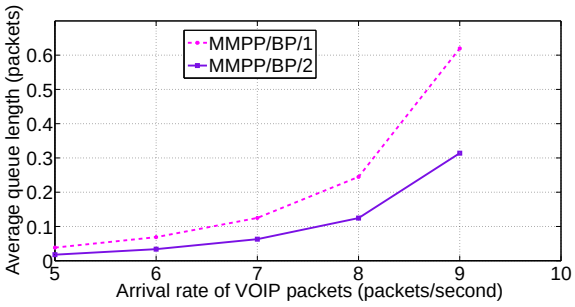

Figure 24. Average queue length as a function of arrival rate of VoIP packets for MMPP/BP/1 and MMPP/BP/2

Figure 24 shows a graph of queue length as a function of arrival rate of VoIP packets for MMPP/BP/1 and MMPP/BP/2. We used equation 12 to plot graph 24. We observe from figure 24 and that average queue length generally increases with increase in arrival rate of VoIP packets. We further observe that for lower number of mesh routers, the average queue length for MMPP/BP/1 and MMPP/BP/2 are closer, however as the arrival rate increases, the average queue length of MMPP/BP/1 is higher than average queue length for MMPP/BP/2.
Figure 25 shows a graph of queue length as a function of load due to VoIP packets for MMPP/BP/1 and MMPP/BP/2 queue systems. We used equation 12 to plot graph 25. We observe from figure 25 that average queue length increases with increase in load due to VoIP packets. We further observe that for lower load values, the average queue length for MMPP/BP/1 and MMPP/BP/2 are the same,

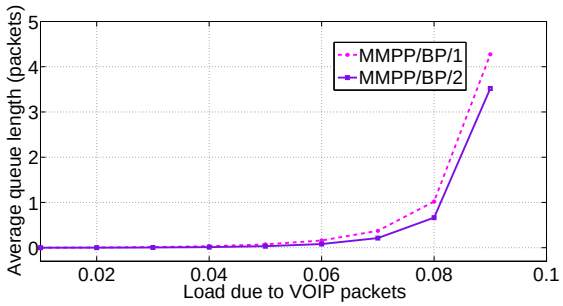

Figure 25. Average queue length as a function of load due to VoIP packets for MMPP/BP/1 and MMPP/BP/2

however as the load increases, the average queue length of MMPP/BP/1 is higher than average queue length for MMPP/BP/2.

12 DISCUSSIONS

In a recent study, the end to end delay of VOIP packets in a WMN were modeled using an M/D/1 queue based model. In this model, the arrival pattern is assumed to follow Poisson distribution. However, the long held paradigm in the communication and performance communities that voice traffic and, data follow Poisson distribution is inaccurate and inefficient. To overcome the above challenge, we modeled queuing delay and queue length of VoIP packets in WMN using MMPP/G/m queue system with arrivals that follow Markov Modulated Poisson Process, and general service time. The numerical results obtained from the derived models show that increasing the number of mesh routers, arrival rate and load leads to increase in queuing delay and queue length. Increase in queuing delay and queue length as a result of increasing the number of mesh routers can be explained by the fact that at each mesh router packets experience delay implying the more the mesh routers, the more the delay. Furthermore, increase in load and arrival rate lead to increase in the number of packets being processed and this in turn leads to increased queuing delay and queue length. In addition, it is observed that correlated

arrivals modeled by MMPP/M/m and MMPP/BP/m queue systems exhibit a higher queuing delay and queue length as compared to Poisson arrivals modeled by M/D/1 queue system especially at high values of mesh routers, arrival rates and load.

13 CONCLUSION

We presented an analytical model for evaluating the performance of VoIP traffic in WMN under correlated inter arrival times and general service time using MMPP/G/m queue system. The model is used to evaluate the performance of VoIP traffic in WMN in terms of queuing delay and queue length. The numerical results obtained from the derived models show that increasing the number of mesh routers, arrival rate and load leads to increase in queuing delay and queue length. In addition, it is observed that correlated arrivals modeled by MMPP/M/m and MMPP/BP/m queue systems exhibit a higher queuing delay and queue length as compared to Poisson arrivals modeled by M/D/1 queue system especially at high values of mesh routers, arrival rates and load.

ACKNOWLEDGMENT

This work was partially funded by the International University of East Africa, Kampala, Uganda Research Fund.

REFERENCES

[1] B. Goode, ”Voice over Internet Protocol (VoIP),” in Proc.IEEE, vol. 90, no. 9, Sep. 2002.

[2] I.F.Akyildiz and X.Wang, ”A Survey on Wireless Mesh Networks,” in Proc.IEEE Communication Magazine, vol. 43, no. 9, 2005, pp. 23-30.

[3] R. Bruno, M. Conti, and E. Gregori, ”Mesh Networks: Commodity Multihop AdHoc Networks,” inProc.IEEE Communication Magazine,vol. 37, March, 2005, pp. 123-131.

[4] S. Mahmud and S. Khan, ”A comparison of MANETs and WMNs: commercial feasibility of community wireless networks and MANETs,” in Proc.the 1st international conference on Access networks , 2006, pp.36-54.

[5] A.Chhabra and G. Singh, ”Performance Evaluation and Delay Modelling of VoIP Traffic over 802.11 Wireless Mesh Network,” in Proc.International Journal of Computer Applications, vol. 21, no.9, May 2011.

[6] I. T. Union, ”ITU-T Recommendation G.114.”

[7] H. B. Kekre, V. A. Bharadi, R. S. Bansode, and V. Kaul, ”Performance Evaluation of VOIP over Multi Radio Multi channel Network,” in Proc. International Journal of Computer Trends and Technology (IJCTT), vol. 5, no. 4, Nov. 2013.

[8] H. Wei, K. Kim, A. Kashyap, and S. Ganguly, ”On Admission of VoIP Calls over Wireless Mesh Network,” inProc.IEEE ICC, 2006.

[9] V. Gupta, M. Burrough, and H. Balter, ”Analysis of Scheduling Policies under Correlated Job Sizes,” School of Computer Science Carnegie Mellon University Pittsburgh, PA, CMU-CS-10-107, 2010.

[10] M. E. Crovella and A. Bestavros, ”Self-similarity in World Wide Web traffic: Evidence and possible causes.” in Proc. ACM SIGMETRICS, May 1996, pp. 160-169.

[11] L. Will, E. Taqqu, M. Willinger, and D. Wilson, ”On the self-similar nature of Ethernet traffic,” in Proc. IEEE/ACM Transactions on Networking, vol. 2, no. 1, 1994, pp. 1-15.

[12] Z. Sahinoglu and S. Tekinay, ”On Multimedia Networks: Self-Similar Traffic and Network Performance,” New Jersey Institute of Technology.

[13] P. A. Moro, ”Statistical Properties of Inter-arrival Times Distribution in Social Tagging Systems,” Andrea Capocci University.

[14] T. Yoshihara, S. Kasahara, and Y. Takahashi, ”Practical Time-Scale Fitting of Self-similar traffic with Markov Modulated Poisson process,” in Proc. Telecommunication Systems , vol. 17, no. 2, 2001, pp. 185211.

[15] A. Nogueira, P. Salvador, R. Valadas, and A. Pacheco, ”Fitting self-similar traffic by a superposition of MMPPs modeling the distribution at multiple time scales,” in Proc. IEICE Transactions, vol. 87, no. 3, 2004, pp. 678-688.

[16] P. Rodriguez, C. Spanner, and E. W. Biersack, ”Analysis of Web Caching Architectures: Hierarchical and Distributed Caching,” in Proc. IEEE/ACM Transactions on Networking, vol. 9, no. 4, 2001, pp. 404-418.

[17] W. Fischer and K. Meier-Hellstern, ”The Markov-modulated Poisson process (MMPP) cookbook,” in Proc. Performance Evaluation, vol. 18, no. 2, Sep. 1993, pp. 149-171.

[18] Guizani, M. Rayes, A. Khan, and B. Al-Fuqaha, ”Network Modeling and Simulation, A practical Perspective.” John Wiley&Sons, 2010.

[19] Scott and S. Smyth, The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling. Oxford University Press, 2003.

[20] D. P. Bertsekas, ”Session 1813 Traffic Behavior and Queuing in a QoS Environment,” Department of Electrical Engineering, M.I.T, 2002.

[21] B. Ciciani, A. Santoro, and P. Romano, ”Approximate Analytical Models for Networked Servers Subject to MMPP Arrival Processes,” Dipartimento di Informatica e Sistemistica, Universita di Roma La Sapienza.

[22] L. Kleinrock, Queueing Systems, Volume I& II. Computer Applications. John Wiley & Sons, 1976.

[23] D. Yao, ”Refining the Diffusion Approximation for the M/G/m Queue”, Proc. Operations Research, vol.6, no.33, 1985.

[24] O. J Boxma, J. W Cohen, and N. Huffels, ”Approximations of the Mean Waiting Time in an M/G/s Queueing System”, Proc. Operations Research (INFORMS), vol.6, no.33, 1979.

[25] N. Gans, G. Koole, and A. Mandelbaum, ”Telephone Call Centers: Tutorial, Review, and Research Prospects”, Proc. Manufacturing & Service Operations Management, vol.2, no.5, 2003.

[26] L. Carvalho, E. Mota, R. Aguiar, and A. Barreto, ”An E-Model Implementation for Speech Quality Evaluation in VoIP Systems,” IEEE Symposium Computers and Communications (ISCC), 2005, pp. 1530-1346.

[27] H.Li, M.Muskulus, and L.Wolters, ”Modeling Job Arrivals in adata-intensive Grid,” in Proc. 12th Workshop on Job Scheduling Strategies for Parallel Processing, 2006.

[28] A. O. Allen, Probability, Statistics and Queuing Theory, with computer Science Applications, 2nd edition. Academic Press, Boston, 1990.

International Journal of Digital Information and Wireless Communications (IJDIWC) 6(2): 139-152
The Society of Digital Information and Wireless Communications, 2016 ISSN: 2225-658X (Online); ISSN 2412-6551 (Print)

Standardizing Sustainability Benefits of Cloud Computing for Non-Expert Decision-Makers

Karim Mualla *(Author)*
School of Energy, Geoscience, Infrastructure and Society
Heriot-Watt University
Edinburgh, United Kingdom
kjm33@hw.ac.uk

Gareth Pender *(co-Author)*
School of Energy, Geoscience, Infrastructure and Society
Heriot-Watt University
Edinburgh, United Kingdom
g.pender@hw.ac.uk

David Jenkins *(co-Author)*
School of Energy, Geoscience, Infrastructure and Society
Heriot-Watt University
Edinburgh, United Kingdom
d.p.jenkins@hw.ac.uk

ABSTRACT

Cloud computing is currently being provisioned and utilized by organisations in various and unstandardized approaches depending on the contract between the service provider and the end-user. The problem in this approach has been observed in the long-term cost, management overheads, and the environment sustainability associated with ICT usage. One method is to treat cloud computing services as a commodity via a metered billing approach, similarly to the way users pay for water, gas and electricity. This approach has shown a potential for sustainability and environmental benefits. These advantages have ranged from cost reductions, power optimisation, and simplifying in-house management processes. The objective of this paper is to examine the sustainability and environmental aspects of cloud computing services for non-expert users. The paper conducts an ICT energy consumption analysis and concludes a decision-making framework for the purpose of standardizing cloud computing sustainability aspects for non-expert decision-makers. The paper argues that certain cloud characteristics such as resource pooling, elasticity, multi-tenancy and service consolidating, can have a positive impact on any organisation with a medium or low-level ICT capacity in order to reduce energy consumption and gain future sustainability advantages.

KEYWORDS

Cloud Computing, Non-Expert, Sustainability, ICT

1. INTRODUCTION

Non-expert decision-makers have been observed to struggle when asked to evaluate cloud computing sustainability benefits or trade-offs before making decisions on whether to outsource core ICT components of their organisations into the cloud or not. This paper addresses this issue from a sustainability perspective by evaluating the current improper standardization of cloud computing models and features offered by top service providers.

At present, the audience of cloud computing is not limited to ICT specialists who operates in organisations with complex ICT work nature. Non-experts in almost all types of organisations are also benefiting today from different features of cloud computing. A basic definition of cloud computing for non-expert clients is the use of the

Internet for the tasks performed on computers. The Cloud here represents the Internet [1]. Non-experts can outsource in-house ICT components onto the cloud, or integrate newly added cloud components with existing in-house systems. The potential benefits to organisations from adopting cloud services were argued to illuminate capital investments for the ICT infrastructure in an organisation, reduce long-term expenses on ICT maintenance and upgrade, and limit the human interaction for provisioning and supporting ICT in the organisation. In addition, other benefits has been observed for being improperly standardised from the non-expert end-user perspective, whereas most research has been carried out on large datacentres and the service provider's perspective. Some of which is the sustainability and energy optimisation of cloud computing services. This paper addresses this benefit from a decision-making perspective for non-experts in accordance to the different cloud computing characteristics, hosting, and service delivery models.

The structure of this paper is divided as follows: Section 2 will introduce a brief background on the sustainability principles of cloud computing from the viewpoint of non-expert users. In Section 3, a literature review will be discussed briefly with reference to the current cloud computing energy-management services and market standards offered by academics and service providers. Section 4 will evaluate the main characteristics of cloud computing in accordance with the relevant environmental aspects. In Section 5, these aspects will be analysed through selected case studies in contrast to the cloud computing key architectural layers of Infrastructure, Platform, and Service. Following this, the paper will analyse results and evaluate the findings of the previous case studies in relation to the illustrated sustainability aspects. Section 6 will discuss the development of a cloud computing sustainability framework for non-expert decision-makers for the purpose of potentially overcoming the improper standardisation of environmental benefits and trade-offs, which accompanies different types of cloud computing features. At the end, conclusions and future works are listed.

2. BACKGROUND

Organisations and clients are using cloud computing services on a daily basis whether they realise it or not [2]. For example, internet email accounts, social networks, GPS locations, and numerous other forms of online data storage and sharing are constantly being accessed by millions of users worldwide. These services are supplied by ICT providers that own virtualized datacentres for end-users to access through the Internet. In general terms, cloud-computing is a ubiquitous platform which provides on-demand ICT services through either the public Internet, or other privately-managed and secure tunnelling networks like Virtual Private Networks (VPN) [3]. The Cloud concept came to life mainly because of the growing ICT requirements in almost each industry, which were not being fulfilled through previous models due to costly services and complex management procedures. Furthermore, another reason for the cloud features to take over conventional methods was due to the highly energy-consuming hardware, which is adopted in the traditional ICT deployments. This was more highlighted for large datacentres where a big number of servers and racks are adopted and using great amounts of energy including the cooling, ventilation, and other power-consuming associated tasks. Nevertheless, although organisations such as hospitals, shopping malls, government agencies and universities are not as ICT-dependant as the large datacentres in terms of the number of servers and other infrastructure, nevertheless, adopting cloud computing services can have impacts on the energy bill of those portfolios.

Several cloud computing scientists and organizations have identified different characteristics, service-delivery models, architectural types, and legal aspects of a system necessary to support cloud computing. According to the NIST definition of cloud computing concepts, five essential characteristics were necessary: On-Demand Self Service, Broad Network Access, Resource Pooling, Rapid Elasticity, and Measured Services [4]. In addition, experts from The Cloud Security Alliance have identified a sixth cloud characteristic and named it

Economy of Scale [5]. Furthermore, other client-cloud computing characteristics were widely discussed by many organizations, such as Multi Tenancy [6], which indicates the distributed manner of computing access and sharing of resources across the cloud. Cloud computing hosting models were divided into four interrelated models as follows:

- Public: Cloud providers offer a full range of computing services via online means, which enables organizations to outsource the entire ICT infrastructure into the cloud.

- Private: Organizations operate either on-site, exclusively managed, or via a third-party outsourced cloud, or a combination of both.

- Community: Multiple organizations with similar operational goals and security policies, share the same virtual ICT services and platform, which can be managed by one of the above, a third-party, or a combination of all.

- Hybrid: Often the most preferable cloud deployment method for end-users, as it ensures additional management flexibilities regarding security, risk elimination, information systems portability, and better standardization. The hybrid solution offers a mixture of various sub-components from previous deployment approaches. In particular, this model irrespectively combines the technical and nontechnical aspects from Private, Public and Community models [7].

Moreover, client-cloud computing potential benefits extend beyond obtaining cost reductions and management flexibility. On this note, multiple energy saving characteristics were pointed out by academics and service providers given that ICT virtualization can have a significant potential for eliminating plugged-in equipment, thus minimizing associated electricity consumption, space and management. The Green and sustainability characteristics of cloud computing are discussed and analysed in this paper from the perspective of non-expert decision-makers. A brief summary of those aspects are presented as follows: [8] [9]:

- Dynamic Provisioning: The ability to reduce unwanted cloud computing components through better matching of server capacity with actual clients' demand.

- Multi-Tenancy: The ability to normalize and flatten unmeasured peak loads by serving large numbers of clients on a shared hosting infrastructure.

- Server Utilization: The ability to operate servers at higher utilization rates via virtualization techniques.

- Datacentres Efficiency: The ability to use advanced datacentre features which reduce the overall power loss through improved methods of power conditioning, air cooling, and other methods.

3. LITERATURE REVIEW

This section will discuss the literature on existing sustainability approaches for organisations using cloud computing and virtual networking technologies, and current market solutions for cloud-based energy management in organisations with different sizes and work natures.

One of the major issues facing this planet today is pollution due to greenhouse gases. ICT-related pollution came to attention given the swift development of new technologies which led to the dumping of large amounts of unused and outdated hardware without a proper recycling strategies or waste management. The relevance of that to this paper is explained in what cloud-computing services can potentially benefit the environment if utilized properly in terms of eliminating outdated hardware through virtual methods. These fast advancements in both the industrial and digital fields have raised many concerns regarding different environmental aspects such as greenhouse gas emissions, waste management, the output of raw material, and the availability and consumption of energy which is especially witnessed in third world countries.

Moreover, carbon monoxide's high rates were observed by experts to reach unprecedented levels, especially in developed urban cities where almost half of the world's population resides [10]. These highly developed cities can be currently portrayed as the battlefield ground between different organizations which include the environmental side on one hand that strives for sustainability, and the winning side which only seeks economic prosperity [11]. With respect to the disappointing history of achieving ICT sustainability despite the massive amount of literature published on this subject, not much was offered in terms of how to effectively balance both economic growth and sustainability in an ICT infrastructure strategy.

One of the major potential benefits from implementing fully, or partially on-demand cloud-computing solutions in organisations, is the ability to acquire an easily maintainable energy saving, and self-healing cable-free infrastructure [12]. Whilst the logic behind this statement arises due to the properties of virtualized techniques achieved through online dependent cloud-computing concepts, the general statement assumes that accessing and controlling the entire organisation's internal systems requires nothing more than a simple, reliable, and secure internet connection. These tasks outsource using such an approach correspond with internal functions, including IT systems, HVAC equipment, sensors, elevators, lighting control, CCTV, fire alarms, and other implemented building devices.

Following this through easily attainable online access by a secure WIFI network as an example, a large-scale of permission management, administration, and heavy daily support can to a certain degree be outsourced to external datacentres owned and operated by cloud providers [13]. In consequence, a high number of connected organisations can be managed simultaneously using the same ICT infrastructure. As a result, several sustainability objectives can be considered achieved from such migration procedures, as earlier attempts to acquire a cable-free virtualized building solution were unsuccessful due to complex networking hardware and wiring infrastructure.

In relation to energy efficient ICTs for smart applications, whether related to organisations in buildings, transportation, agriculture or any other smart principle; it can be acknowledged from previous published work that cloud-computing techniques have not been standardized and applied as a fully operating IT platform. The reasons behind this are due to performance, administration, and security vulnerabilities. Although similar topics have been the target of numerous computer science studies concerning virtual information benefits for companies' IT solutions, only a few papers have discussed the energy efficient advantages from cloud-computing utilization as will be listed next. In addition, it can be concluded from previous literature that cloud-computing benefits with regard to sustainable management and decision-making approaches are, in most cases, presented as a secondary topic in a broader energy consumption study [14] [15] [16].

According to a 2009 study by the British Computer Society and Oxford University Press, energy efficient cloud-computing has examined several Low carbon footprint approaches for IT datacentres and communication services [17]. For the primary aim of reducing Green House Gas Emissions (GHG) from computation and the physical space occupied by associated hardware, the paper significantly portrays the cloud approach as an inherently power saving technology that has recently attracted the large-scale of attention of non-expert managers. However, it has been pointed out that despite the fact that most literature has focused on hardware aspects in relation to usage, optimization, and energy efficient performance, the information and communication services for potential Green solutions has not been fully implemented as an ICT infrastructure. In particular, cloud-computing solutions were mainly deemed at that time inapplicable for potential power consumption reduction.

Moreover, the study discussed various benefits to be gained from implementing an IT solution based on cloud concepts. These services, which to a considerable extent, are categorized Green in different operational tasks, performance, and energy-aware aspects, are fundamentally

concerned with dumping heavy computational workload on an online virtually-managed system. In theory, this workload is only required either infrequently, or on a scheduled basis. For example, a certain datacentre processing function might be needed for only 30 minutes on a Sunday night, such as crunching a large number of data as part of a weekly backup. Although this particular task requires a hundred parallel servers, the normal building operation only requires 50 servers to operate on a normal workday basis.

The paper has also analysed Amazon's cloud-computing monthly costs regarding a datacentre's energy distribution over a 3-year period [18]. Furthermore, the study argued that an estimation of 30% savings can be obtained from unnecessary cooling power. In addition, 20% of energy emitted from networking infrastructure in a sizable building could also be dispensed with [19]. Regarding the potential benefits of cloud-computing for the environment, a 2008 study by the Accenture has argued that energy consumption from networked-based servers alone can be reduced by 20% from using cloud services [20]. In addition, HP stated that savings from cloud-computing deployments can reach up to 30% with regard to the energy spent on cooling for heavy-duty hardware. Furthermore, it was estimated by the same study that the carbon exhaust of these equipment is currently reaching around 70% of the datacentre's total power exhaust.

The main conclusions were centred on achieving virtualized, energy efficient solutions while providing insights on how to best manage the approach in large-scale infrastructures. These environments have a high demand for information and communication services as well as various other nontechnical requirements, which can also be integrated onto a single virtualized platform [21].

Microsoft published a report on cloud-computing smart applications, which discussed potential possibilities for cloud approaches to achieve power efficient resource management [22]. According to a 2011 Microsoft Corporation report on making organisations smart, control over the cloud has recently been one of the centrally debated topics. Further, smart transportation, and a new generation of grid systems were both considered essential platforms for achieving sustainability.

Case studies concluded that accurate decisions to enhance energy performance and management in Smart Buildings could not be effectively executed in real-time circumstances, as it was simply impossible to make sense of events, reports, and data analytics captured from IP systems. This was argued as one of the problems cloud-computing can solve via the Infrastructure as a Service layer (IaaS). The study argued that these recently innovated cloud approaches are transforming the way energy consumption, in both buildings, and cities will occur in the long-term. Although full IT transparency is being offered for networking and processing infrastructure, contributions from several Microsoft partners like Hitachi, Stanford and California University, are comprehensively examining methods to enhance current models on Smart Building energy management. For instance, the previous model suggests connecting a network of organisations into a Smart Grid, which to some extent can potentially be deployed across the world.

4. MARKET SOLUTIONS FOR CLOUD-BASED ENERGY MANAGEMENT

According to a revenue chart created by the Pike study, the market of energy systems for Smart building management has had a growing and almost consistent rate of revenues since the introduction of cloud-based services. For example, while revenues have gone as far as $ 2 billion in 2011, it has been estimated that by 2020 a return profit of $ 6 billion will occur from using the Panoptix service by Smart Buildings in the US alone [23].

According to Fujitsu, a smart energy management service referred to as Enetune was set to be launched in June 2013 as part of an energy optimization process for businesses and organisations located over multiple locations [24]. This service will employ the online Cloud as a

data capturing, storing, and processing platform from different energy consuming sources [25]. Cloud-based services arrive with a bill at the end of each month. For example, the Enetune EMS service costs about 400 US dollars per project (location) on a monthly basis. This is excluding support, upgrades or any other bespoke features.

It was pointed out by the development manager at Open General that the migration process from conventional web-enabled technologies in a building energy management system, into a transparent cloud-based solution, is considered essential to data integration methods within organisations [26]. In particular, with the employment of open communication protocols such as BACnet, Zig-Bee, and Mod-Bus, two levels within the system architecture has been identified with regard to data integration: Software level and the Controller level.

According to an article by Automated Buildings Enterprise, the current market of cloud energy management for organisations is leaning strongly towards a Hybrid interconnected approach [27]. This connection is expected to take place with several related industries such as Smart Grids and others. An example of such services to support this application is the use of Virtual Real-time Information Systems (VRIS) [28] [29].

A research at Microsoft has carried out a cloud-computing energy performance study with respect to selected applications from the ICT organization such as Word, Excel and Outlook exchange [30]. Whereby the deployment of these tools is considered almost a given in each Smart Building ICT environment, the main objective of the study was to highlight greenhouse gas emissions from utilizing a Microsoft cloud-based alternative.

Other studies have identified the cloud-computing energy optimization factor via mobile platforms. This was particularly discussed in a study by Purdue University where the main objective was focused on enhancing computing capabilities and applications across mobile devices [31]. The ultimate solution was to ensure maximum battery life for ad-hoc ICT systems. Although cloud utilization was debated as a potential solution for a low-power ICT lifecycle, multiple challenges were addressed.

In reference to power consolidation via cloud approaches, another study was deemed significant to this research given multiple Smart Buildings' ICT services [32]. The paper addressed the mutual liaisons between ICT utilization on one hand, and associated energy consumption on the other while taking into account execution performance obtained from strengthened workloads. The main focus was highlighting complexities in achieving energy consideration by identifying both performance barriers and benefits gained from energy consolidation across different smart environments where a certain degree of system integration is accomplished.

5. SUSTAINABLY ASPECTS OF CLOUD COMPUTING

The dynamic scaling ability offered by on-demand cloud services provides the opportunity for energy saving in organisations, this is now discussed. It was argued that ICT investments are the most influential factor for attaining 'Green', low carbon organisations [33], also, numerous projects have recently been carried out on adopting emerging ICT solutions for contributions to energy saving [34] [35]. For instance, EU Commission standards, initiated in 2009, investigated the significance of in-depth relationships between ICTs' technical administration, and energy-intensive industries such as Smart Buildings and Transportation.

According to the Accenture Group and in response to intensive virtualization and economy of scale techniques applied by cloud datacentres, cloud-computing solutions have the ability to reduce a company's carbon emissions by approximately 30% per IT user [36]. This was argued as being a result of outsourcing applications, networking bandwidth, and processing units into cloud-hosted datacentres. This indicated that some ICT components deployed in a non-virtual manner are responsible for energy consumption in terms of different portfolio sizes and workload. The following will present cloud-computing

contribution to power usage minimization with reference to the previous literature analysis. This is argued in accordance with key cloud management attributes discussed earlier.

In a normal organisations in a building environment, users (e.g. people or IP devises) access the internet through either a local area network (LAN) cable (e.g. RJ 45), or via a direct wireless connection. Then, a cloud service request, which follows IaaS, PaaS, or SaaS, is sent as IP packets from the internal on-premises router to the internet provider's main router, to eventually reach the cloud service provider's gateway router. These requests are then subsequently dispatched to a shared pool of distributed virtual machines (VMs), which host a diverse scope of ICT resources, covering software applications, development platforms, processors, and networking bandwidth. Each step of the previous process consumes a certain amount of energy. Other tasks/services that are not directly involved in the cloud service delivery process (e.g. cooling, lighting, and electrical equipment needed to support the ICT lifecycle in Smart Buildings) were argued to consume the largest part of energy [37].

Research found during the Literature Review chapter has particularly highlighted four key areas of cloud characteristics by the Accenture group, which demonstrated positive impacts on reducing greenhouse gas emissions and refining ICT energy usage in organisations. These aspects have covered:

- Dynamic provisioning
- Multi-tenancy
- Virtualization
- Server capacity utilization
- Cloud provider's large-scale datacentres

Other sub-factors were identified in the overall ICT energy consumption process in organisations, in which cloud computing has a strong potential to optimize (Table 1).

To date, cloud experts claim that there is not a clear consensus towards classifying cloud-computing as a Green ICT [39]. Generally, management awareness was considered arguably misperceived towards cost minimization on one hand, and reducing ICT electricity usage on the other. While many features in organisations aim to automate as much end-user tasks as possible, in-house ICT infrastructure is mostly over-implemented as deliveries and capacity exceed what is actually needed, through the use of costly systems. This was termed in the conservative and traditional service methodology as Over Provisioning. This approach would result in minimizing energy efficiency and maximizing carbon savings, as this indicates that multiple versions of each system or networking process, is replicated, installed, and supported separately on each site.

In most cases when unpredictable heavy user access occurs, it is very difficult to predict the amount of bandwidth required to install a specific system. Non-expert decision-makers might adopt several frameworks for Green ICT operation; however, another crucial aspect must be taken into consideration. This highlights analysing resource minimization of internally hosted alternatives. Accordingly, cloud providers' energy efficient datacentres are mostly run next to massive renewable energy sites in order to maximize energy usage in the best way possible [40]. Organisations can rely on these heavily-burdened structures for obtaining resource-efficient ICT systems with minimum on-site equipment installed. However, this should be performed prior to taking into consideration all energy consuming attributes within the ICT environment. These attributes were defined by several academics throughout different frameworks via ICT power-usage parameters as [41]:

- COP (Coefficient of Performance) average
- The Carbon intensity of the electricity being used by each ICT component (kg/kWh)
- Electricity prices per ICT component
- Networking (next-hop) cost per GB, for data transfer (up/download)
- CPU uptime, downtime, quantity, frequency ratio, and power required

Cloud computing has a significant potential for eliminating plugged-in equipment, thus, minimizing associated electricity consumption, space, and management effort. More, this was assumed to reinforce Green utilization for Smart Building ICT applications. In addition, the fact that in most cloud hosting cases energy is being displaced from onsite to offsite, this displacement only saves energy if these processes can be run more efficiently due to economy of scale (e.g. large datacentres) [42]. These datacentres could be situated in geographically favourable locations such as cooler climates, which will have lower cooling loads than buildings located elsewhere which are using those off site servers.

In concern, multiple energy efficient aspects were concluded in response to previously discussed cloud characteristics as clarified next.

• Enabling Resource Virtualization

Regardless of the deployment method, the core concept behind cloud-computing is the ability to run several operating systems on one machine. Therefore, adopting virtual machines with relatively similar capabilities can be acquired either on, or off-premises within a single or multi-branched ICT environments. With that in mind, the e-waste footprint of each ICT element such as servers with high CPU power can be substituted by VMs, thus, reducing electricity spending and energy of physical plugged-in units.

• Strengthening Consolidation

Although virtualization is considered the primary cloud energy-efficient aspect, a crucial reliance on software automation for scaling up/down as workload demands, forms a key benefit behind virtual ICT implementation. This criterion allows non-expert Smart Building decision-makers to fully utilize rented cloud resources in contrast to resource ratios via conservative physical ICT methods. The conventional approach was noted as more costly and power consuming in reference to unhandled high rates of server utilization, which can be minimized through virtualization and a solid backup of software automation.

• Enabling Energy-Efficient Behaviour

The pay-as-needed billing concept of cloud-computing has a significant impact on energy end-user behaviour for enhancing lifecycle administration and service oriented expenditures. More, while Smart Buildings' heavy ICT dependence mostly includes the utilization of plugged-in, off-site, and third party managed infrastructure, this is accomplished - in cloud terms – following an as-needed approach. Each unwanted ICT element will simultaneously be switched off, and these resources are then pushed back into the shared pool as previously explained in the cloud characteristics section.

• Applying Multi-branched Demand Patterns

In relation to a multi-branched set of Smart Buildings such as Banks, Hotels, or Hospitals, operating on a single networking platform, the multi-tenancy attribute of cloud-computing is considered a major energy saving aspect for several reasons in accordance with each cloud deployment model as follows:

- Public clouds mobility standards allow differently located users to access ICT services and applications from anywhere via the Internet, while taking into account several security, reliability, and data integrity considerations. This saves energy in Smart Buildings because these users are relying more on privately owned end-systems from off-premises locations, which takes the load off the organization's ICT infrastructure.

- Private clouds, which are installed on one site such as the main headquarters of the organization, can act as a cloud provider datacentre for other Smart Building branches. These are able to access, utilize, and release ICT resources by following similar public cloud techniques.

- Hybrid clouds, whether deployed on or off-premises have a significant role in reinforcing both the security and performance of previous approaches as discussed earlier in the deployment models analysis.

Previous points indicate that each location with a certain workload peak rate, have a strong potential for saving energy. For instance, peak rates for networking, processing, or application access are widely reduced when distributed between multiple end-users via shared demand patters. Although less physical infrastructure is required as a result, the economy-of-scale aspect of cloud-computing plays a considerable role in maximizing energy efficiency and resource troughs.

6. CASE STUDY ANALYSIS

This paper interviewed a senior manager at [a popular ICT provider] in the UK. This interview has followed a semi-structured approach and covered potential environmental and sustainability benefits or trade-offs of this company's client records from the perspective of non-expert decision-makers.

It was first confirmed by the interviewee that SaaS services are currently the most popular among the majority of the provider's clients. The Hybrid Hosting was identified by the interviewee as the most popular deployment model across the current clients. The following figure shows how cloud services have almost achieved the ICT demand level of end-users, while the classic capacity line is fixed and does not always correspond with the ICT demand of end-users. As a result, end-users will end-up with either (Figure 1):

- Over Capacity: This will occur when organizations use owned conventional ICTs that have more capacity than required in real-life. As a result, these users will be forced to continue managing and paying for system runtime and maintenance of unneeded ICT infrastructure.

- Under Capacity: This will occur when organizations use their owned conventional ICTs that have less capacity than required in real-life. As a result, these users will be forced to fully purchase new systems to meet this demand, and provide support for this new infrastructure, which is only needed for a short period of time.

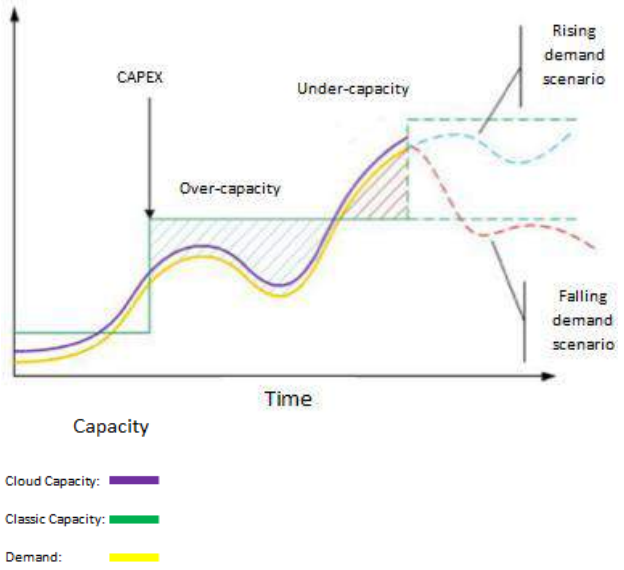

Figure 1. Capacity comparison of Cloud vs. Classic Demand

This According to the provider's client records which were provided exclusively to this paper, consumers with unstable ICT demands are complaining from the rising cost of their existing in-house systems. The Over Capacity is observed more in big organizations where unexploited resources are purchased, which leads to additional expenses on support and upgrade. However, the Under Capacity aspect is observed more in small organizations where less ICT demand is usually required in most applications.

In relation to the energy-efficient cloud-computing factor, it can be concluded from the interview that big provider names in the ICT industry are focused on service delivery in terms of support, availability, and customer satisfaction, rather than empowering organisations with energy-efficient features of cloud applications. This conclusion came primarily as a result of service requesters' demands towards eliminating in-house ICT maintenance, upgrades concerns, and staff salaries. It was observed by this paper that the majority of clients over the past 5 years are not particularly interested in the energy-efficient benefits gained from cloud services. Their main interest is obtaining cost reductions and decreasing time-consuming management efforts. The interviewee explained this by stating that obtaining considerable energy cuts from cloud computing is still a debatable argument depending on multiple

Table 1. Energy consuming Elements against Cloud-Computing Contribution for Organisations

Energy consuming ICT Elements	Cloud-Computing Contribution	Example
Core Applications	Eliminating extensive CPU power	Internally installed, long-running software consume a considerable amount of CPU power (e.g. CRM tools for banks, security monitoring tools for shopping malls, etc.). Although this is not usually a concern during the development of these applications, electricity consumption can be reduced through outsourcing core applications through utilizing SaaS resources, which are virtually run over distributed machines at the cloud provider's infrastructure.
PCs & Servers' Response Time	Reducing end-user response time by relying on distributed VMs instead of high performance on-premises servers	Non-expert managers in different types of organizations have a tricky task of weighing cloud-computing QoS and energy saving features on one hand, with branching limitations on the other. For example, it is well known that any physical server would perform, to a large extent, better than a virtual substitute [38]. However, enhancing response time is nonetheless a major energy saving attribute, as cloud providers mostly employ a large number of VMs assigned over globally located datacentres, hence, ensuring reliability via data replication, rapid provisioning, and availability rates.
Networking Hierarchy Systems	Minimizing capacity bandwidth from no-more-necessary internal networking devices (e.g. topology design, wired networking awareness)	Utilizing cloud solutions could relatively increase networking processes regarding number of hops between source and destination. However, employing a dynamically scalable infrastructure as a service IaaS, will only consume energy on the basis of delivering packets to the in-house router according to peak workloads. This is done in Smart Building networking systems, which are mostly structured to deal with worst case scenario such as throttle-neck periods. As a result, this can eliminate internal connection complexities as opposed to implementing conventional on-premises datacentres that require power-burdened networking devices (e.g. switches, cables, signal power points, hubs, conditioning equipment, etc.).

ICT attributes related to the specific organisation involved.

7. SUMMARY OF FINDINGS

One of the key examples presented by the interviewee was the VisitBritain agency, where the cloud sustainability factor has played a significant role in forming the client's ICT strategy. In this example, a large amount of hardware, and networking infrastructure was required to support a heavy communication processes and ICT capacity peaks.

This demand was only required for the 2012 London Olympic games, which only cover one month of uptime ICT utilization. Therefore, cloud-computing features were a great solution for this scenario, avoiding having both over capacity and under capacity at the same time. Furthermore, cloud-computing sustainable techniques played a significant role in that respect, where ICT virtualization, migration, and support, provided large scale virtual machines, server components, networking bandwidth, and 24/7 contingency maintenance of an entirely outsourced infrastructure.

In terms of the feasibility to outsource and host an entire building's networking infrastructure on a cloud platform, end-users will only be required to use thin-client Graphical User Interfaces (GUIs) as an on-site ICT infrastructure. Examples of these GUIs are screens with ad-hoc ports Ethernet access and minimum buffering power such as IBM pure-systems, Google ChromeBox Cloud-based PCs, which are an optimized private cloud with a self-service user interface.

The paper simulated costs and potential energy savings in a major [higher education facility] in the UK, which is currently using conventional ICT hosting methods. Following a semi-structured interview with the university's Information Director, this paper gathered the following date. The university employs 25 ICT personnel in-house, and the total cost assigned to the ICT infrastructure per year is £0.5 million. This was divided as follows:

- £ 100,000 for Information Systems upgrade
- £ 250,000 for Networking and Communication Systems upgrade
- £ 100,000 for Hardware Maintenance (e.g. core networks, remote monitoring, etc) (20% of the total budget each year)
- £ 50,000 for Software support from various vendors (excluding fixed contract costs)

This paper used the tool: SBCE, which has been developed as part of a PhD project at Heriot-Watt University in 2016. The tool offers dynamic and elastic cost estimation, simulation and management consultancy features in addition to energy saving calculations of the ICT infrastructure for organisations. After selecting the substituted ICT components that were estimated if cloud computing was adopted in contrast to the current conventional methods, the following savings were calculated which compares between thin and thick clients, and the total cost of the ICT infrastructure.

Estimated Total Cost for the first year cloud-computing simulation: Deployment costs (£ 73,322.40) + Support costs (£ 22,889.22) = £96,211.62

Estimated Total Cost for the three years cloud-computing simulation: Deployment costs (£ 220,106.44) + Support costs (£ 68,707.82) = £288,814.26

It can be concluded that applying a cloud solution seems cheaper than the on-going multi-vendor, in-house solution. This was demonstrated from the £0.5 million pounds spent by the university on the ICT infrastructure per year, in contrast to the £96,211.62 required for the first year from applying the cloud alternative. The previous results are excluding any additional elasticity service demand patters, or any fixed service contracts with specific vendors such as Blackboard, and others, which costs Heriot-Watt University around £ 50,000 per year as explained earlier.

The university acquires about 5,000 computers covering school labs and staff offices. These in addition to the main library are administered privately by the in-house ICT support team. In that context, this study previously proposed the purchase and utilization of light-weight thin-clients, instead of the currently utilized thick-client devices. The former will soon become obsolete resulting in thick-client hardware being dumped and replaced on a regular basis. Therefore, expenses related to purchasing, upgrading, managing, and licensing, are enormous as the university's DIS has acknowledged in the interview earlier. In addition, with regard to hardware acquisition and associated power consumption for the entire infrastructure, the Green aspect of operating in an environmentally friendly manner can be drastically improved from employing thin-client equipment.

For instance, Google and Samsung offer ChromBox, a light weight PC that only consumes 8-15 watts instead of the 250 watts per each regular thick-client device. As a result, end-user device costs can reach around £ 269 instead of a £ 600 average for an HP desktop computer. With accordance to the higher education case study, the number of watts approximately consumed by end-user PCs only can be measured approximately as follows (Table 2).

Table 2. Watts approximately consumed by end-user PCs: Thick-client vs. Thin-client

Following the existing thick-client approach:	*5,000 PCs: each PC consumes 250 watts ↔ 250 x 5,000 = 1.25MW (Total Consumption)* *5,000 PCs: each PC costs £ 600 ↔ 600 x 5,000 = £ 3,000,000 (Total PC Infrastructure Cost)*
Following the potential thin-client approach:	*5,000 thin PCs: each PC consumes 12 watts ↔ 12 x 5,000 = 60,000 watts (Total Consumption)* *5,000 PCs: each thin PC costs £ 269 ↔ 269 x 5,000 = £ 1,345,000 (Total PC Infrastructure Cost)*

By default, any thick-client device will exclude costs related to any operating system licenses, anti-virus protection, and other required software, given that devices like ChromeBox are online-based, self-healing with automatic built-in system upgrade. Moreover, other desktop computers were also classified under the thin-client category. These have also been argued to optimize energy usage, minimize hardware possession, and ensure efficient remote utilization of resources given that the operating system is already hosted on the manufacturer's cloud environment [43]. Some examples of today's ICT market, this light-weight hardware can range from the HP MultiSeat PC, to the Wyse computer by Dell, in addition to other networking storage systems such as Sun MicroSystems, KronosSystem, and ReadyNAS by NetGear.

8. CONCLUSION AND FUTURE WORK

The objective of this paper is to standardize the sustainability aspects of cloud computing from the perspective of non-expert decision-makers. The paper introduced a brief background on the sustainability principles of cloud computing from the viewpoint of non-expert users. Following that, a literature review was discussed briefly with reference to the current cloud computing energy-management services and market standards offered by academics and service providers. Furthermore, the main characteristics of cloud computing in accordance with the relevant environmental aspects were discussed and analysed through selected case studies in contrast to the cloud computing key architectural layers of Infrastructure, Platform, and Service. The paper then analysed results and evaluate the findings of the previous case studies in relation to the illustrated sustainability aspects.

As discussed, studies such as the UN Habitat indicated that developed cities with high population such as London and Beijing, are accountable for nearly 85% of greenhouse gas emissions [44]. According to other previously reviewed studies, this number classified these cities, in carbon terms, as unsuitable places to live in the future. It was also stated that organisations in buildings are responsible for around 45% of energy consumption in Europe alone. In particular, ICT in a normal building with medium-capacity datacentres is currently

responsible for over 10% of the total cost of this structure. Furthermore, the overall global CPU power and storage capacity was observed to double every 18 months, and the global ICT consumption growth was noticed to rise from 123 billion kWh in 2005, to 246 billion kWh in 2010. This reflected a 2% increase of the worldwide CO2 emissions. As a result, it was estimated that a set of server racks, which include around a thousand servers, would currently cost around $ 4.5 million, mainly due to its power consumption in a normal capacity datacentre. Cloud-computing was introduced to help mitigate this issue, not only from ease-of-management and economical perspectives, but also in relation to various associated environmental factors. This was argued to have a strong potential to minimize software and hardware physical acquisition and usage in different types of organisations.

The main conclusions can be summarized as follows:

- An observation was made that cloud experts currently claim that there is not a clear consensus towards classifying cloud computing as a Green ICT.

- In almost organisations, Servers occupy the biggest percentage in cost and energy consumption as opposed to all other ICT components and associated attributes. In addition, 42% of the power consumption of an ICT-burdened organisation is designated to the cooling infrastructure.

- The main conclusion argued that for a Green installation in a heavily ICT dependent organization, adopting cloud-computing was in most of the client cases -from top service providers- more cost efficient in terms of hardware, datacentre costs, and management. In the long-term, in some cases the savings got lower when a client moves from a legacy environment into a cloud one due to extra costs such as support expenses, and non-planned hardware upgrade.

- With regard to the higher education facility cost and energy spend simulation, cloud-computing was observed to be cheaper and more energy efficient than the conventional existing approach for a 3-year deployment.

Future work is suggested to highlight the energy use of specific industries with accordance to ICT, and evaluate the potential energy reductions from adopting cloud computing services. This would offer a better understanding of each industry individually, hence, the research would assist non-expert decision-makers in identifying the trade-offs, benefits, and areas of concern that are only specific to their organisations' work nature, operational objectives, and other administrative attributes such as size, employees, and budgets.

REFERENCES

[1] K. Mualla, D. Jenkins, (2015). 'Evaluating Cloud Computing Challenges for Non-Expert Decision-Makers'. International Journal of Digital Information and Wireless Communications (IJDIWC): Vol 5, Issue 4.

[2] J. Rubner, "When the Sky's the limit". Collective Intelligence: Cloud Computing, 2011.

[3] R. Bernnat, W. Zink, N. Bieber, J. Strach, "Standardizing the Cloud: A Call to Action". Booz and Company Inc, 2012.

[4] P. Mell, T. Grance, "The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology", 2011.

[5] K. Mualla, D. Jenkins, (2015). 'Evaluating Cloud Computing Management Challenges for Non-Expert Clients'. Proceedings of the Second International Conference on Data Mining, Internet Computing, and Big Data, Reduit, Mauritius 2015.

[6] R. Buyya. J. Broberg. A. Goscinski, "Cloud Commuting: Principles and Paradigms". Wiley Press, New York, USA, 2011.

[7] BizCloud Corp. "Defining Cloud Deployment Models", http://bizcloudnetwork.com/defining-cloud-deployment-models, 2010

[8] C. Mines, "4 Reasons Why Cloud Computing is Also a Green Solution", GreenBiz, Web: http://www.greenbiz.com/blog/2011/07/27/4-reasons-why-cloud-computing-also-green-solution?page=0%2C1, 2011.

[9] A. Kofmehl, A. Levine, G. Falco, K. Schmidt, "Energy-Smart Buildings: Demonstrating how information technology can cut energy use and costs of real estate portfolios". Accenture Corporation, 2011.

[10] P. Simon, (2012). 'IBM, Smarter Buildings: A Smarter Planet'. Web Link: http://www.ibm.com/smarterplanet/uk/en/green_buildings/ideas/index.html

[11] C. McKinsey. 'Siemens'. (2008). Sustainable Urban Infrastructure London Edition – a view to 2025: A research project sponsored by Siemens.

[12] K. Weldon, (2012). 'M2M Evolution: What are the Enablers of Future Growth?'. M2M Evolution, Advisory Report.

[13] Graybar Service Enterprise. (2013). 'In-Building Wireless Solutions'. Web Link: http://www.graybar.com/applications/intelligent-buildings/mobility/in-building-wireless.

[14] W.Roth, W. Kurt,. Y.Feng, Michael. P. LIana, L. Quartararo, (2005). 'Energy Impacts of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential'. TIAX LLC for U.S. Department of Energy.

[15] B. Tung, H. Tsang, K. Lai, L. Lam, K. Tung, H. (2011). 'Hybrid Energy Management Solution for Smart Building'. IEEE International Conference on Consumer Electronics (ICCE).

[16] J. Reed, (2000). 'The Structure and Operation of the Commercial Building Market'. ACEEE Study on Energy Efficiency in Buildings.

[17] A. Berl, E. Gelenbe, M. Di Girolamo, Giuliani, Giovanni. De Meer, Hermann. Dang, Minh Quan. Pentikousis, Kostas. (2009). 'Energy-Efficient Cloud Computing'. Oxford University Press on behalf of The British Computer Society.

[18] Amazon: Elastic Compute Cloud (EC2). (2013). 'Monthly On-Demand Instance Price Calculator'. Web Link: http://calculator.s3.amazonaws.com/calc5.html

[19] Accenture Report: Silicon Valley Leadership Group. (2008). 'Data Centre Energy Forecast Report'.

[20] Group & WSP Report. (2010). 'Cloud Computing and Sustainability: The Environmental Benefits of Moving to the Cloud'. Web Link: http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture_Sustainability_Cloud_Computing_TheEnvironmentalBenefitsofMovingtotheCloud.pdf

[21] M. Grajek, (2012). 'ICT approach: A Targeted Approach'. Bruegel Policy Contribution.

[22] K. Willson, B. Mitchel, J. Gimenez, (2011). 'The Central Role in Cloud Computing in Making Cities Energy Smart'. Microsoft Corporation.

[23] E. Bloom, B. Gohn, (2012). 'Smart Buildings: Ten Trends to Watch in 2012 and Beyond'. Pike Research: CleanTech Market Intelligence.

[24] Enetune-Fujitsu. (2012). 'Fujitsu Launches New Enetune Cloud-based Energy Management System'. Public and Investor Relations Division, Fujitsu Limited. Web Link: http://www.fujitsu.com/global/news/pr/archives/month/2012/20120515-04.html

[25] Fareastgizmos. (2012). 'Fujitsu Enetune Cloud-based Energy Management System helps to optimize energy usage'. Web Link: http://www.fareastgizmos.com/other_stuff/fujitsu-enetune-cloud-based-energy-management-system-helps-to-optimize-energy-usage.php

[26] N. Munasinghe, (2010). 'The Future of Cloud Connectiviity for BAS'. Open General.

[27] R. Lavelle, K. Onuma, (2010). 'Virtual Building Energy Management: Moving to Cloud-based Building Energy Management'. Open General.

[28] G. Zucker, F. Judex, C. Hettfleisch, R. Schmidt, P. Palensky, D. Bascıotti, (2012). 'Energy aware building automation enables Smart Grid-friendly buildings'. Elektrotechnik & Informationstechnik, Springer-Verlag.

[29] M. Younis, M. Youssef, K. Arisha, (2003). 'Energy-aware management for cluster-based sensor networks'. Department of Computer Science and Electrical Engineering, University of Maryland.

[30] D. Williams, Y. Tang, (2013). 'Impact of Office Productivity: Cloud Computing on Energy Consumption and Greenhouse Gas Emissions'. University of Reading, United Kingdom.

[31] K. Kumar, Lu. Yung-Hsiang. (2010). 'Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?'. Purdue University, Published by the IEEE Computer Society.

[32] S. Srikantaiah, A. Kansal, F. Zhao, (2008). 'Energy Aware Consolidation for Cloud Computing'. Pennsylvania State University and a Microsoft Research.

[33] A. Peltomäki, (2009). 'ICT for a Low Carbon Economy: Smart Buildings'. Findings by the High-Level Advisory Group and the REEB Consortium on the Building and Construction sector, EU Commission.

[34] M. Pérez Ortega, (2012). 'FIT4Green ICT Project', Website: http://events.networks.imdea.org/content/e-energy-2012/e2dc-workshop.

[35] Project Earth: Energy Aware Radio and neTwork tecHnologies. (2013). 'Driving the Energy Efficiency of Wireless Infrastructure to its Limits'. Website: www.ict-earth.eu.

[36] A. Kofmehl, A. Levine, Falco, S. Gregory. (2011). 'Energy-Smart Buildings: Demonstrating how information technology can cut energy use and costs of real estate portfolios'. Accenture Corporation.

[37] A. Berl, E. Gelenbe, M. Di Girolamo, G. Giuliani, H. De Meer, M. Dang, K. Pentikousis, (2009). 'Energy-Efficient Cloud Computing'. Oxford University Press on behalf of The British Computer Society.

[38] L. Cherkasova, R. Gardner, (2005). 'Measuring CPU overhead for I/O processing in the Xen virtual machine monitor'. ATEC '05 Proceedings of the annual conference on USENIX Annual Technical Conference.

[39] L. Younge, L. Wang, S. Lopez-Alarcon, S. Carithers (2010). 'Efficient resource management for Cloud computing environments'. Green Computing Conference, 2010 International

[40] G. Kumar, B. Saurabh, L. Rajkumar. (2012). 'Green Cloud computing and Environmental Sustainability'. Cloud computing and Distributed Systems (CLOUDS) Laboratory, Dept. of Computer Science and Software Engineering, University of Melbourne, Australia.

[41] C. Mines, (2011). '4 Reasons Why Cloud Computing is Also a Green Solution'. GreenBiz, Web Link: http://www.greenbiz.com/blog/2011/07/27/4-reasons-why-cloud-computing-also-green-solution?page=0%2C1

[42] P. Costello, R. Rathi, (2012). 'Data Center Energy Efficiency, Renewbale Energy and Carbon Offset Investment Best Practices'. Real Energy Writers.

[43] A. Lagar-Cavilla, N. Tolia, R. Balan, E. Lara, D. O'Hallaron, (2006). 'Dimorphic Computing'. School of Computer Science - Carnegie Mellon University, Pittsburgh.

[44] J. Zhao, (2012). "Climate Change Mitigation in Beijing, China". Case study prepared for Cities and Climate Change: Global Report on Human Settlements 2011.

International Journal of
DIGITAL INFORMATION AND WIRELESS COMMUNICATIONS

The *International Journal of Digital Information and Wireless Communications* aims to provide a forum for scientists, engineers, and practitioners to present their latest research results, ideas, developments and applications in the field of computer architectures, information technology, and mobile technologies. The *IJDIWC* publishes issues four (4) times a year and accepts three types of papers as follows:

1. **Research papers**: that are presenting and discussing the latest, and the most profound research results in the scope of IJDIWC. Papers should describe new contributions in the scope of IJDIWC and support claims of novelty with citations to the relevant literature.
2. **Technical papers**: that are establishing meaningful forum between practitioners and researchers with useful solutions in various fields of digital security and forensics. It includes all kinds of practical applications, which covers principles, projects, missions, techniques, tools, methods, processes etc.
3. **Review papers**: that are critically analyzing past and current research trends in the field.

Manuscripts submitted to IJDIWC should not be previously published or be under review by any other publication. Original unpublished manuscripts are solicited in the following areas including but not limited to:

- Information Technology Strategies
- Information Technology Infrastructure
- Information Technology Human Resources
- System Development and Implementation
- Digital Communications
- Technology Developments
- Technology Futures
- National Policies and Standards
- Antenna Systems and Design
- Channel Modeling and Propagation
- Coding for Wireless Systems
- Multiuser and Multiple Access Schemes
- Optical Wireless Communications
- Resource Allocation over Wireless Networks
- Security; Authentication and Cryptography for Wireless Networks
- Signal Processing Techniques and Tools
- Software and Cognitive Radio
- Wireless Traffic and Routing Ad-hoc Networks
- Wireless System Architectures and Applications