Jordanian Journal of Computers and Information Technology

April 2018 | VOLUME 04 | NUMBER 01 | ISSN 2415 - 1076 (Online) ISSN 2413 - 9351 (Print)

JJCIT

www.jjcit.org jjcit@psut.edu.jo

An International Peer-Reviewed Scientific Journal
Financed by the Scientific Research Support Fund

Jordanian Journal of Computers and Information Technology (JJCIT)

The Jordanian Journal of Computers and Information Technology (JJCIT) is an international journal that publishes original, high-quality and cutting edge research papers on all aspects and technologies in ICT fields.

JJCIT is hosted by Princess Sumaya University for Technology (PSUT) and supported by the Scientific Research Support Fund in Jordan. Researchers have the right to read, print, distribute, search, download, copy or link to the full text of articles. JJCIT permits reproduction as long as the source is acknowledged.

AIMS AND SCOPE

The JJCIT aims to publish the most current developments in the form of original articles as well as review articles in all areas of Telecommunications, Computer Engineering and Information Technology and make them available to researchers worldwide. The JJCIT focuses on topics including, but not limited to: Computer Engineering & Communication Networks, Computer Science & Information Systems and Information Technology and Applications.

INDEXING

JJCIT is indexed in:

- CrossRef
  http://search.crossref.org/?q=jjcit
- OCLC WorldCat
  http://www.worldcat.org/search?qt=worldcat_org_all&q=jjcit
- Scilit
  http://www.scilit.net/journals/387088
- Google Scholar
  https://scholar.google.com/citations?user=88ospLoAAAAJ&hl=en

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

NODE DENSITY IMPACT ON ENERGY CONSUMPTION AND CONTACT PROBABILITY OF OPPORTUNISTIC NETWORK

Abubaker Alhutaba[1], Salem Sati[2] and Mohamed Eshtawie[3]

(Received: 19-Sep.-2017, Revised: 28-Nov.-2017 and 17-Dec.-2017, Accepted: 31-Dec.-2017)

ABSTRACT

Opportunistic communication between two encountered nodes is commonly established using a radio technology, such as Wi-Fi or Bluetooth. One issue involved in opportunistic communication is a trade-off between connection time and probability of resource consumption. This paper presents a comprehensive study on density analysis for decentralized distributed opportunistic communication using Wi-Fi technology. In this work, study and analysis of contact probability and energy efficiency of variant density in a particular area are performed. The contribution of this work is the analysis of the impact of density on the connection probability and resources, as well as a simulation study framework to analyze the contact event with a view of energy consumption. The study gave detailed contact information, such as contact probability based on node density and transmission range in a particular area, as well as the beacon exchange process as an element of channel utilization and energy consumption. The influence evaluation of various parameters on each other and finally on the system performance is also presented.

KEYWORDS

Opportunistic Communication, Contact Probability, Beacon Interval, Energy Consumption, Node Density.

1. INTRODUCTION

The opportunistic network typically consists of a large number of devices deployed over a particular area. Opportunistic nodes are capable of exchanging messages with the surrounding environment. This opportunistic environment has no stable end-to-end connectivity. The relay nodes perform the store and carry message to transfer it to the destination. However, in a sparse opportunistic network, the distance between neighbouring nodes is usually bigger than the interface transmission range. Thus, multi-hop forward or routing is unfeasible due to lack of end-to-end connectivity. Message transmission in opportunistic networks is accomplished through hop-by-hop routing. Also, opportunistic nodes have different mobility patterns that vary from deterministic to completely random mobility pattern. The main challenge of spares opportunistic networks is the time of contact together with the energy efficiency of neighbour discovery and message replication. Mobile opportunistic nodes have to discover the neighbours or forwarders in the transmission area. Ideally, each mobile node should be able to discover the next hop to reduce delay and avoid possible message losses at the local buffer of the node in message transmission. Moreover, the mobile opportunistic node discovery process should exploit as much short time available for message exchange and replication as possible. Due to limited energy resources, the discovery process is made difficult by mobile nodes energy constraints. Neighbour discovery is achieved through typical periodic listening or sensing when the node regularly sends a beacon to announce its presence in the area, while other nodes check for possible beacons. Hence, proper node definition ensures timely discovery of all contacts to reduce energy consumption at the mobile node, thus increasing node lifetime but decreasing the capability or delay of detecting contacts and neighbours. In general, to solve the message delivery problem, the opportunistic routing protocol uses broadcast mechanism for message delivery via Store-Carry-Forward fashion through the network. The simplest opportunistic routing protocol is flooding-based routing protocol named Epidemic [1]. In Epidemic routing protocol, the deliverable messages are broadcast to every encountered node that has no buffered copy of the message. The number of

1. A. Alhutaba is with the College of Industrial Technology, Misurata, Libya. Email: Alhutaba_a_abu@yahoo.com
2. S. Sati is with Misurata University, IT Faculty, Misurata, Libya. Email: salem.sati@it.misuratau.edu.ly
3. M. Eshtawie is with the College of Industrial Technology, Misurata, Libya. Email: Eshtawie @yahoo.com

broadcasts or message copies increases through message dissemination process depending on node contact probability and number of neighbours. In addition, the number of message copies in the entire network is limited by the message Time-To-Live TTL and number of network nodes. This paper presents an analysis of the impact of density on contacts and energy. In Section 2, related work is presented. Section 3 discusses opportunistic communications and neighbour discovery for the mobile opportunistic network. The opportunistic communication impact by the network density is discussed in more detail. Section 4 discusses the trade-off between contact probability and energy consumption in mobile opportunistic networks. In Section 5, the metrics and tools used together with the data from different experiment scenarios are presented. The results obtained from experiment scenarios and simulation experiments are investigated and presented in Section 6. Finally, Section 7 gives the conclusion and future work.

2. RELATED WORK

The beacon interval is critical for wireless infrastructure and infrastructure-less opportunistic network communications. There is no specific value defined for beacon interval such as proposed in [2] which sets the beacon interval of the master device in infrastructure-based mobile opportunistic networks. In their work, they suggest the beacon interval to be twice the traditional interval of a commercial Wi-Fi appliance. Namely, the authors term their approach with Double Hundred Beacon Interval 2HKBI. In [3], the scenarios of access-point-based opportunistic network communications are proposed. The authors assume that the node scanning probe is equal to both passive and active scanning times. They also calculate the epidemic dissemination speed regarding the beacon rate in a time slot of one second. Many routing protocols with their benefits have been proposed for opportunistic networks. Commonly, the well-known opportunistic routing protocols do not consider the available energy budget when making routing decisions. Only few researchers have investigated energy-aware protocols. The performance of opportunistic routing protocols which are epidemic are evaluated in [4]. Spray & Wait, PROPHET, MaxProp and Bubble Rap used the ONE simulator presented in [5]. Their analysis show that the most effective routing protocol is using metrics such as delivery ratio/latency and energy consumption. The results illustrate the impact of energy consumption on the routing performance. Other related studies have been conducted in order to evaluate the performance of opportunistic routing protocols under resource constraints. These works proposed Markovian Chain model to address the energy problem in Delay Tolerant Networks DTNs. As an example, in [6], the issue of energy consumption opportunistic forwarding for DTNs by introducing a Markov model and proposed different types of forwarding strategies is considered. Furthermore, the authors in [7] proposed their Energy-Aware Epidemic Routing (EAER), which is an extension of the n-Epidemic routing proposed in [8], which aims to improve the performance in terms of message delivery ratio and energy consumption. The n-parameter policy is proposed for optimizing the possibility of message transmission from a node to its neighbours, known as node degree. By using this strategy, a node will forward its message to the next node only when it is in the range of at least n neighbours as a threshold of message transmission. Furthermore, [12] proposes a routing algorithm which reduces energy consumption and increases delivery probability. The authors achieved this by calculating nodes' remaining energy and available free buffer space for receiving copies of messages. In [13], the authors analyzed the social network model based on a multi-layer detected by encounters of the nodes. Moreover, this paper investigates the relationship between different layers in terms of node degree and population. The authors of [14], suggest an Energy Aware EA-PRoPHET as a new protocol. This protocol considers the limitation of energy budget and physical buffer space for message replication process. The paper shows the simulated results of the suggested protocol, which say that it has better performance. [15] describes the mobility traces based on a campus environment to capture the contacts using Bluetooth. The authors gathered the information based on the Facebook relationships. Their contribution was a way of understanding the human mobility at different social ties.

3. DENSITY IMPACT ON BEACON AND CONTACT

Opportunistic mobile nodes detect whether there is any neighbour device to communicate; i.e., a device that needs to send beacons (active mode) or listen to beacons (passive mode). This task is costly in terms of energy, bandwidth and delay. Due to different mobility models, the network is divided into different groups. However, discussing such a complex situation is useful for routing

protocol design when considering the number of infected nodes in each group that impacts the dissemination speed. Both infrastructure and infrastructure-less opportunistic networks have periodically broadcast their identifiers that can be seen by any Wi-Fi-enabled nodes as control signalling (Beacon). Wi-Fi-enabled nodes typically listen to these announcements in regular intervals. In order to set up a connection, a node must initiate a connection and the other node must accept the connection. Therefore, as contact probe process, the beacon will consume the bandwidth of the radio channel. Furthermore, it consumes the energy resources of the nodes. The beacon interval will have an impact on the contact probability as one hop detection delay of neighbour discovery. The beacon indicates why two encountered nodes disconnect from each other and when they have left the transmission range. The beacon is kept alive signalling, where the transmission range and node density are yields to this event of disconnection, especially when the nodes of the network move in, particular area. The other important reason of the disconnection between encountered nodes is the energy limitation of the mobile opportunistic network. The mobile opportunistic node is suffering from resource limitation in terms of storage, bandwidth and energy. To improve the mechanism of neighbour discovery in a mobile opportunistic network environment, IP Neighbour Discovery (IPND) [9] is implemented and published by IETF in the Internet-Draft. IPND protocol is a method of mobile opportunistic communication for nodes to discover the existence, availability and addresses of encountered nodes as one hop connected. IPND periodically transmits UDP message (broadcast) and receives beacons as a distributed system.

4. ENERGY AND NEIGHBOUR DISCOVERY

The mobile opportunistic network has no reliable end-to-end connectivity, where it is hard to guarantee a stable path due to time-varying network topology. Thus, mobile opportunistic nodes have to replicate messages to relay nodes which are in their communication range. In order to enable such message replication, nodes have to continuously detect the environment to discover neighbour nodes. Obviously, this neighbour discovery is an extreme energy consumption. Therefore, it is important to investigate energy consumption during the contact and neighbour discovery process in the opportunistic network. One strategy for saving node energy is to increase the interval between beacon scans of contact and neighbour discovery. The consequence of this interval increase is that mobile opportunistic nodes may miss the opportunity to contact other nodes and thus opportunities to replicate the messages are lost. Moreover, if nodes scan the environment much frequently, a lot of energy will be consumed in the contact and neighbour discovery process causing it to be inefficient. This valuable reason points to a trade-off between energy consumption and contact or neighbour discovery delay, where the nodes scan their transmission range using the beacon broadcast process. For neighbour discovery which uses a constant contact scan interval, the larger the contact scan interval, the greater the number of missed contacts encountered and vice versa. The reflection of the trade-off between energy consumption and contact probability in opportunistic networks is investigated.

5. TOOLS AND METRICS

5.1 Tools

The ONE simulator is used for simulating various movement models. These movements are generated by synthetic models or real movement models. Furthermore, the ONE simulator is able to forward or replace the messages between nodes through different opportunistic routing protocols. The ONE simulator has four modules; namely, movement, routing, event and report models.

5.1.1 Energy Model

In designing an opportunistic routing scheme, node energy needs to be taken into consideration. This fact reveals the importance of deeply analyzing the core design of routing protocols and message replication issues between encountered nodes. The routing and replication will depend on the energy budgets of both nodes. Therefore, in this work, an energy model for an opportunistic simulator is used. Based on this energy model, the energy efficiency of existing epidemic routing protocol that achieves higher message delivery rates through flooding replication decisions is studied. Furthermore, the available node energy is not considered in the majority of existing opportunistic routing protocols when they make forwarding or replication decisions. This limits both delivery probability of the

message and the network lifetime. The trade-off between energy consumption and epidemic replication efficiency as a dynamic energy optimal control problem is also analyzed. In this scenario, each node decides on its replication probability based on its current energy budget. Since energy decreases with transmissions and receptions of messages, the replication decisions vary with time. Therefore, the replication decision is considered as a criterion to control the evolution of a network that captures the fraction of nodes carrying the copy of the message and the energy budget of the nodes. Each node in our energy model has an energy resource. The energy model which monitors consumed energy for node activities such as replication, transmission or message reception is integrated. Contact and neighbour discovery scanning, the energy model is implemented in the ONE simulator to analyze the impact of energy efficiency of the epidemic routing and IPND discovery protocols. This energy module considers scanning, reception and transmission interface states. To reflect different network density situations, we consider the three scenarios listed as parameters in Table 1 and compare their impact on node density. The comparison regards the different metrics under different radio ranges and node numbers in a particular area. The different three scenarios were simulated with the default settings of the ONE Simulator [5], [10].

5.1.2 Density and Contact Probability

Opportunistic nodes communicate with multiple hops in a store, carry and forward fashion. When a node cannot find any neighbour nodes within its communication range, it should gain a contact opportunity with encountered nodes to replicate the message. The contact is a criterion of finding suitable nodes which can replicate the message to the destination. Clearly, the node with a higher contact probability has a higher priority for replicating the message towards the destination.

Table 1. Simulation settings.

No.	Settings	Map of downtown Helsinki, Finland
1	Simulation time	12 h
2	Number of nodes	60,120,240
3	Group type with speed	Pedestrians (0.5-1.5 km/h) Cars (10-80 km/h) Trains (10-80 km/h)
4	Simulation area	Helsinki, Finland map (4500m , 3400m)
5	Routing protocols	Epidemic
6	Interface type	High speed
7	Transmission range	50,100,150,200,250 m
8	Bandwidth	250 KBps ,10 M
9	Buffer management	FIFO
10	Message size	0.5-1 MB
11	Message creation interval	25-35 s
12	Time-to-live (TTL)	300 min
13	Default buffer size	Pedestrians: 5 MB Cars, trains: 50 MB
14	Mobility model	Pedestrians: Shortest path map based Cars, trains: Map route

Moreover, the higher the node degree, the more likely to act as a relay based on popularity metric. Contact probability and node degree are exploited in existing replication schemes. Essentially, contact probability is an important factor for message replication process of the epidemic to the destination. This contact probability is impacted by transmission range and number of nodes. Furthermore, the node degree has an importance in opportunistic network analysis. The degree of the node indicates the number of nodes connected to this node. As a node's degree increases, it has a chance of contacting with other nodes in the network, the destination may be one of them.

5.2 Metrics

For the simulation scenarios of mobile opportunistic communication, the density has an impact on the

contact probability and the energy consumption. It also has an impact on the available bandwidth. Therefore, we will measure these metrics for the analysis of contact and energy impact by node density.

5.2.1 Contact Probability and Delivery Latency

Two types of latency component simulation scenarios are considered. These are neighbour discovery delay and message transmission delay. Latency is relevant to node density as a component of nodes and transmission range. We further assume that all nodes are cooperative; therefore, they assign the available buffer space across the whole network of N nodes. The expected latency of delivered messages based on the number of message copies can be written as follows:

$$E(t)_{latency} = \frac{1-\exp(-TTLN\lambda_c)}{\lambda_c} \quad (1)$$

where TTL is the message remaining lifetime, c is the Inter-Contact Rate which is the inverse of the average elapsed time of last encounter time for all nodes. The probability of contacts by capturing the number of contacts per hour in different node densities is calculated. We will consider the contact event as the minimum contact C_{min} which is calculated when the transmission range r is equal to one meter. The probability of contact C_{Prob} is calculated as the ratio of maximum contacts of the network that consists of N sets of nodes as vertex and (N-1) of outgoing edges for directed graph G(V, E) by the following equation:

$$C_{prob} = \frac{C_{hour}}{N^2 - N} \quad (2)$$

Equation (2) shows that the contact probability C_{Prob} of the mobile opportunistic network is mostly impacted by the transmission range r regardless of the number of nodes N, where the value of (N^2-N) is assumed as a constant value in the same scenario of each different transmission range r.

5.2.2 Bandwidth and Message Transmission

This metric is conducted with flooding-based DTN routing protocol termed as the greedy uncontrolled epidemic routing protocol. Epidemic routing protocol is evaluated with different number of nodes and radio range. Where the epidemic routing protocol dissemination speed will be impacted by contact probability, contact probability is a function of the number of nodes and transmission range in the bounded area. The routing performance, impacted by node density considered, because node density has a higher impact on the buffer and routing performances. Obviously, the buffer and routing performances are based on the overhead variables which are replication counter and hop counter of the message. The main performance factor is the relayed, successfully transmitted, messages as a metric for resource consumption in terms of bandwidth and channel utilization.

5.2.3 Energy Consumption

Energy consumption is critical in an opportunistic environment. Therefore, we calculated energy consumption of receiving the management traffic. We consider that active scanning in opportunistic communication requires the transmission of requests and adds more energy consumption. The transmission beaconing as UDP message of IPND protocol with 5 second period of interval B_{int} and its energy E_b will be considered. In addition, contact duration C_{dur} will also be considered. Our energy model will conduct different interface states as Send, Receive, Scan states. The energy consumption by beacon in the IPND protocol will be calculated with the following equation:

$$E_B = \sum_{i=1}^{N} \frac{C_{dur}}{B_{\text{int}}} E_b \quad (3)$$

6. NUMERICAL RESULTS

In this work, three different scenarios are applied for the comparison of node density impact on the epidemic performance. As a routing metric, we look at the end-to-end average delay. This delay or

latency metric is used as a performance metric for different DTN applications. Figure 1 shows that epidemic router has a lower delay when the transmission range r is increased by 50 meters for every experiment. This is because the buffer criterion is based on FIFO, so that messages with high buffer delay will be removed. Furthermore, the buffer times have the highest impact on the message TTL, because high buffered messages are either dropped by full buffer space or expired TTL. The delay will be decreased by increasing the transmission range r. As in Figure 1, latency decreases by 500 sec until the range reaches 200 meters. At 250 meter range, latency stays stable,

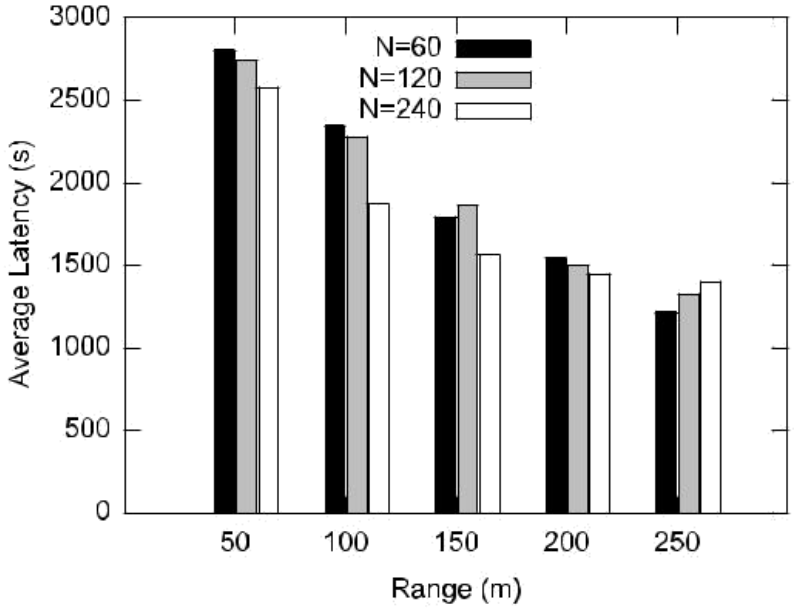


Figure 1 Average latency versus range.

contact probability influence is one of various parameters that have been evaluated. Figure 2 shows that contact probability increases as transmission range increases. Moreover, it shows that the minimum contact probability for different scenarios is achieved when the transmission range is equal to 50 meters. The minimum contact probability of spares network is 0.25. Figure 2 shows that this value is equal for the three scenarios, regardless of the number of nodes. This means that link availability of the three scenarios is equal regardless of the node density values.

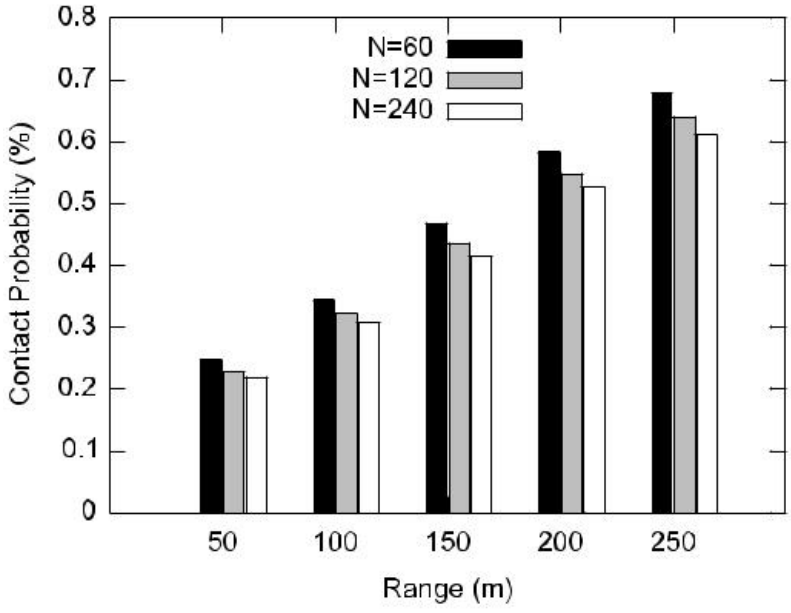


Figure 2. Contact probability versus the range.

To evaluate the epidemic performance with a variety of different node densities, it is also important to measure the relayed or successfully transmitted messages. Figure 3 shows that the number of message copies increases proportionally to the transmission range r and strongly with the number of nodes N. When increasing the number of nodes, the traffic sources of the whole network increase. In addition, the dissemination speed increases by increasing the number of nodes.

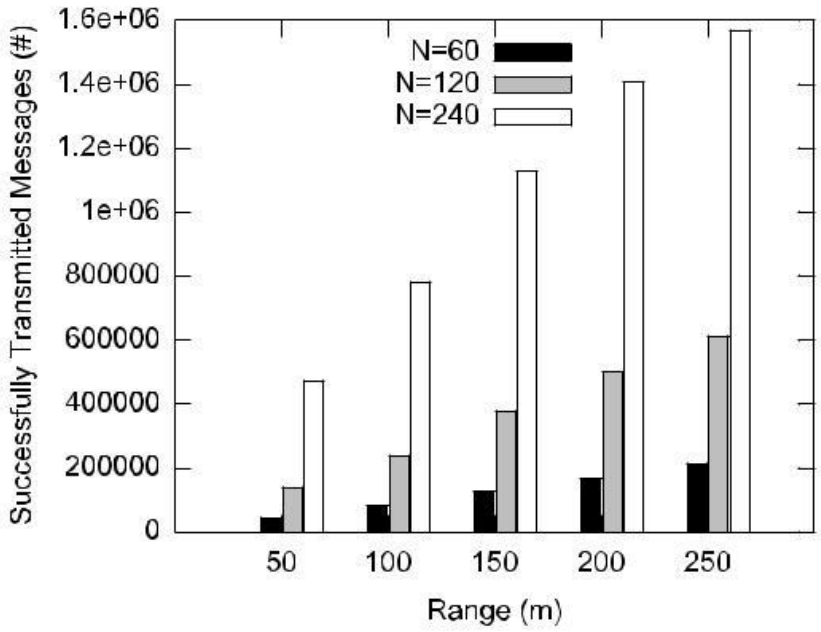


Figure 3. Successfully transmitted messages.

We provide simple energy efficiency. All nodes have an energy model that has the same network constant energy resource as Scan = 1126mW, Send = 1630mW and Receive = 1378mW, with equal values of different states considered in all scenarios. The whole network energy considered for different scenarios is 10^4 Wh. This value of energy guarantees that all nodes will always be on charge. Therefore, node communication interfaces are enabled. The IPND interval is selected as the default value of IPND beacon of IBR-DTN [11] of 5 seconds. Figure 4 shows the energy resource for all nodes after every experiment over simulation time. The Figure also shows the energy of the three different scenarios of 5 second IPND beacon. At 50 meter transmission range, Figure 4 shows that energy consumption increases by 5 % whereas, density is duplicated. The increment in energy consumption becomes 10% when the transmission range equals 250 meter. This is because the epidemic dissemination speed increases as the transmission range increases. Therefore, at a higher number of copies generated by the epidemic router, regarding the scenario of N = 240, we found that the increment of energy consumption was duplicated compared to the scenario of N = 120 nodes. Furthermore, the scenario of N = 60 nodes, which has minimum variety with respect to x-axis, has minimum slope compared with other scenarios. This comes from the fact that low density, low range and low number of nodes give priority to achieving the desired delivery with low energy consumption. From Figure 4, the energy of the network is improved by 10% when the node density is changed at a transmission range of 250 meters. This improvement comes from the fact that when the node density increases, the node degree increases. This is turn leads to improvements in connectivity and contact probability. Therefore, the percentage of delivery increases with short paths.

7. CONCLUSIONS

This paper addresses the problem of mobile opportunistic communication and neighbour discovery using Wi-Fi technology. In this work, we show that energy and bandwidth efficiency of opportunistic communication can be significantly improved. Furthermore, the performance analysis model used to analyze contact and energy of distributed opportunistic communication with different node densities is presented. The performance of epidemic routing with energy consumption, from the aspects of routing

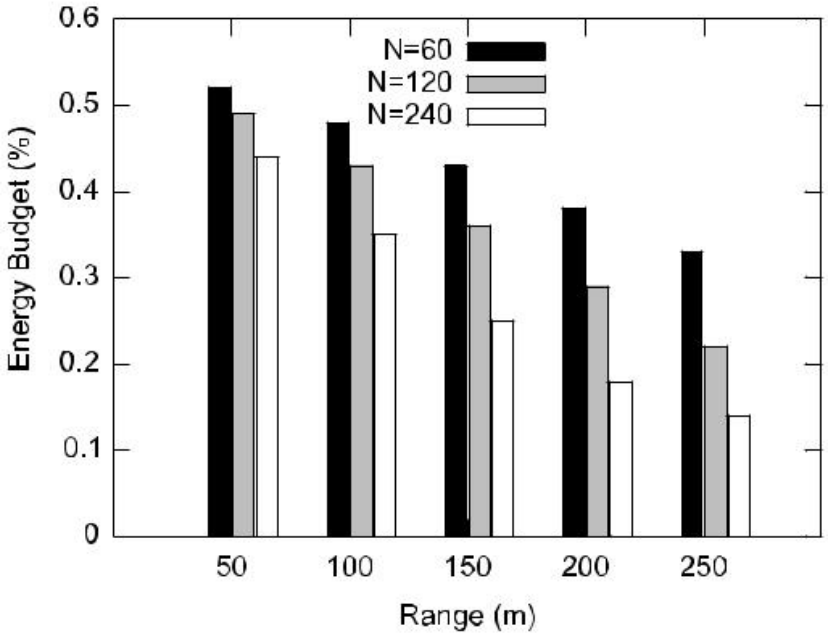


Figure 4. Energy consumption.

metrics in terms of message delivery delay and energy cost as both beacons and relayed messages, is also evaluated. The important parameters in our evaluation include contact probability and energy constraint of a neighbour discovery protocol (IPND). The network density is considered as a function of the number of nodes in the network, in addition to the node transmission range of a particular area. Investigating the impact of dynamic beacons approach on the efficiency of mobile opportunistic communication can be considered as future work for research in this field.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Kalman Graffi for his valuable support and guidelines.

REFERENCES

[1] A. Vahdat and D. Becker, "Epidemic Routing for Partially-Connected Ad Hoc Networks," Duke University, Department of Computer Science, vol. 20, no. 6, 2000.

[2] S. Sati and K. Graffi, "Adapting the Beacon Interval for Opportunistic Network Communications," International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India, pp. 6–12, 10-13 August 2015.

[3] S. Trifunovic, B. Distl, D. Schatzmann and F. Legendre, "WiFi-Opp: Ad-hoc-less Opportunistic Networking, " Proc. of ACM CHANTS '11, pp. 37–42, 2011.

[4] A. Socievole and S. Marano, "Evaluating the Impact of Energy Consumption on Routing Performance in Delay-tolerant Networks," The 8th International Wireless Communications and Mobile Computing Conference (IWCMC), Limassol, Cyprus, pp. 481–486, August 27-31, 2012.

[5] A. Keranen, J. Ott and T. Karkk¨ainen, "The ONE Simulator for DTN Protocol Evaluation," Proceedings of the International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SimuTools), ICST/ACM, 2009.

[6] Y. Li, Y. Jiang, D. Jin, L. Su, L. Zeng and D. Wu, "Energy-efficient Optimal Opportunistic Forwarding for Delay-tolerant Networks," IEEE Trans. Vehicular Technology, vol. 59, no. 9, pp. 4500–4512, 2010.

[7] F. D. Rango, S. Amelio and P. Fazio, "Enhancements of Epidemic Routing in Delay-tolerant Networks from an Energy Perspective," The 9th International Wireless Communications and Mobile Computing Conference (IWCMC), Sardinia, Italy, pp. 731–735, 1-5 July 2013.

[8] X. Lu and P. Hui, "An Energy-efficient n-epidemic Routing Protocol for Delay-tolerant Networks," The 5th International Conference on Networking, Architecture and Storage (NAS), Macau, China, pp. 341–347, July 15-17, 2010.

[9] D. Ellard and R. Altmann, "DTN IP Neighbor Discovery (IPND)," Internet Engineering Task Force, Internet-Draft draft-irtf-dtnrg-ipnd-03, 2016, Work in Progress, [Online], Available: https://tools.ietf.org/html/draft-irtf-dtnrg-ipnd-03

[10] A. Keranen, T. Karkk¨ainen and J. Ott, "Simulating Mobility and DTNs with the ONE (Invited Paper)," Journal of Communications, vol. 5, no. 2, pp. 43–50, 2010.

[11] M. Doering, S. Lahde, J. Morgenroth and L. C. Wolf, "IBR-DTN: An Efficient Implementation for Embedded Systems," Proceedings of the 3rd Workshop on Challenged Networks (CHANTS), San Francisco, California, USA, pp. 117–120, September 15, 2008.

[12] B. B. Bista and D. B. Rawat, "EA-Epidemic: An Energy Aware Epidemic-based Routing Protocol for Delay-tolerant Networks," Journal of Communications, vol. 12, no. 6, June 2017.

[13] A. Socievole, E. Yoneki, F. De Rango and J. Crowcroft, "Ml-sor: Message Routing Using Multi-layer Social Networks in Opportunistic Communications," Computer Networks, vol. 81, pp. 201-219, 2015.

[14] B. B. Bistaand and D. B. Rawat, "EA- PRoPHET: An Energy Aware PRoPHET-based Routing Protocol for Delay-tolerant Networks," IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), pp. 670-677, 2017.

[15] A. Socievole, F. De Rango and A. Caputo, "Wireless Contacts, Facebook Friendships and Interests: Analysis of a Multi-layer Social Network in an Academic Environment," Wireless Days (WD), IFIP, Rio de Janeiro, Brazil, pp.v1-7, Nov. 2014.

ملخص البحث:

يتأسس الاتصال الانتهازي بين غقدتين متواجهتين في العادة باستخدام إحدى تقنيات الراديو مثل "واي فاي" أو "بلوتوث". وتتمثل إحدى القضايا المتعلقة بالاتصال الانتهازي في التسوية بين زمن الاتصال واحتمالية استهلاك الموارد.

تقدم هذه الورقة دراسة شاملة لتحليل الكثافة بالنسبة للاتصال الانتهازي الموزَع اللامركزي باستخدام تقنية "واي فاي". وفي هذا العمل، تمت دراسة وتحليل احتمالية الاتصال وفعالية الطاقة بتغير الكثافة في منطقة معينة. وتتمثل مساهمة هذا البحث في تحليل أثر الكثافة في احتمالية الاتصال وفي استهلاك الموارد، إلى جانب إطارٍ لدراسة محاكاة لتحليل حدث الاتصال من منظور استهلاك الطاقة.

وتعطي هذه الدراسة معلومات مفصلة عن الاتصال، مثل احتمالية الاتصال بناءً على كثافة العُقد ومدى الإرسال في منطقة معينة، إلى جانب عملية تبادل "المرشدات" كعنصر من عناصر الاستفادة من القناة واستهلاك الطاقة. كما تقدم الدراسة تقييماً لأثر متغيرات متنوعة في بعضها البعض وفي أداء النظام.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

HIGH-PERFORMANCE BLOCK MATCHING ALGORITHM FOR HIGH BIT-RATE REAL-TIME VIDEO COMMUNICATION

Nijad Al-Najdawi

(Received: 22-Nov.-2017, Revised: 04-Jan.-2018 and 27-Jan.-2018, Accepted: 15-Feb.-2018)

ABSTRACT

Although the advancements in hardware solutions are growing exponentially along with the communication channels capacity, high quality video encoders for real-time applications are still considered an open area of research. The majority of researchers interested in video encoders target their investigations towards motion estimation and block matching algorithms. Many algorithms that aim to reduce the total number of required mathematical operations when compared to Full Search have been proposed. However, the results often converge to local minima and a significant amount of computations is still required. Therefore, in this research, a hierarchy-based block matching method that facilitates the transmission of high bit-rate videos over standard communication methods is proposed. The proposed algorithm is based on the frequency domain, where the algorithm examines the similarities between a chosen frequency subset, which significantly reduces the total number of comparisons and the total mathematical computations required per block.

KEYWORDS

Video Communication, Video Coding, Video Conferencing, Motion Estimation, Block Matching.

1. INTRODUCTION

Digital videos consist of successive frames sampled over a period of time. Those successive frames carry high data redundancy. Therefore, eliminating bits of redundant data can be extremely helpful in reducing the size of digital video and compressing the video. Several types of compression techniques have been proposed in the past few decades. Those compression techniques are classified as being either lossless or lossy. The former type is achieved by eliminating redundant bits and reproduces exact original dataset. The latter is achieved by eliminating least important bits and reproduces a similar copy that might be indistinguishable by the human visual system from the original. Lossy compression techniques achieve better compression and are more applicable to digital videos; while, lossless techniques are in their nature more applicable for digital images. Lossless image compression allows the use of human visual system limits, by producing data that is sufficient to be classified as "good enough". The latest compression standards have set the architectures for video codecs as consisting of the following basic blocks: prediction, transform and entropy coding. Prediction includes estimates for the position of a current block inside a video frame. Transform process converts a block of pixels into frequency domain. Entropy coding involves encoding video data into a compressed bit stream.

Usually, consecutive frames have the same still or moving objects, creating a high correlation between consecutive frames. Therefore, researchers have investigated the use of methods that examine the object movements in a video sequence in order to produce motion vectors that represent the estimated motion. On the other hand, those estimated vectors are forwarded to the proper motion compensation methods that use those vectors to simulate the object movement, achieving data compression. Motion estimation and compensation methods are considered the most important techniques that eliminate temporal redundancy in successive video frames. However, those techniques are more applicable to translational motion and still have their limits when applied to rotational motion which is difficult to estimate and requires other techniques for processing. Therefore, motion estimation algorithms usually

This paper is an extended version of a short paper that was presented at the International Conference "New Trends in Computing Sciences (ICTCS) 2017", 11-13 October 2017, Amman, Jordan.

N. Al-Najdawi is with Computer Science Department, Al-Balqa Applied University, Salt, Jordan. Email: n.al-najdawi@bau.edu.jo

assume the following: objects movement is translational, illumination is uniform across spatial and temporal domains, occlusions of objects by others are neglected, and finally uncovered background is not to be considered.

Various methods for coding have been proposed for video compression. Those coding techniques include intra-frame and inter-frame coding which are used to minimize the total number of bits required to transmit or store videos. In intra-frame coding, each frame is separately coded and this type of coding includes: transformation quantization and frame encoding. Inter-frame coding investigates the temporal redundancy and is usually applied in video coding in order to achieve the actual compression. In this type of coding, motion estimation and compensation algorithms are normally applied to eliminate the temporal redundancy that exists between successive frames. Various motion estimation approaches were proposed in literature; amongst those approaches, block matching algorithms were proven to be more suitable because of their reliability and simplicity. Block matching algorithms are used to estimate the object's motion in successive frames on the basis of rectangular blocks. These algorithms assume that all the pixels within a block have the same motion behaviour [8].

In block matching algorithms, frames are divided into $N \times N$ blocks; where all blocks in the current frame are matched with candidate blocks within a search area (window) on the reference frame (considering that candidate blocks have a translation movement in other frames) and the displacement motion vector is recorded for the best matched candidate. In inter-frame coding, the motion vector and the residual frame (resulting from subtracting input frame from the prediction of the reference frame) are usually transmitted. At the receiver side, the decoder builds the frame difference and adds it to the reconstructed reference frame. Therefore, data compression is achieved by eliminating inter-frame redundancy. This demonstrates the fact that better prediction methods give smaller error signals and a reduced transmission bit-rate [19].

In this paper, in addition to the introduction section, section 2 provides an up-to-date literature review of motion estimation algorithms. Section 3 introduces the transformation process. In section 4, the proposed hierarchical search algorithm is described along with the proposed matching criterion. Section 5 provides the experimental results and analysis. Finally, section 6 concludes this research.

2. LITERATURE REVIEW

A large number of block matching algorithms have been proposed over the last decades, such as the traditional methods found in [2]-[5], [7], [11], [15]-[16] and [24]. Amongst the available block matching algorithms, full search leads to the best possible match of the block in the reference frame with a block in another frame by calculating the cost function at each possible location in the search window. The resulting motion compensated frame has the highest peak signal-to-noise ratio when compared to any other block matching algorithm. However, this is the most computationally extensive block matching algorithm [8].

Optimized block matching algorithms speed up the exhaustive search required by full search algorithms based on fixed search patterns. Researchers in this domain have investigated the use of many algorithms in order to enhance the traditional search algorithms. In [33], Diaz Cortes et al. proposed a block matching algorithm that combines harmony search with a fitness approximation model. The authors considered the motion vectors in search window as potential matches. The authors applied a fitness function in order to evaluate the matching quality of each motion vector in addition to a strategy to decide which motion vectors can be estimated amongst the rest of the motion vectors. In [34], the authors proposed a hierarchy-based motion estimation algorithm using Gaussian image pyramid and unidirectional estimates of motion vectors at the top level. In their work, the authors proposed the use of five candidates for each motion vector. At the bottom level of the hierarchy, the motion vectors are corrected based on the sum of absolute difference values of the blocks. Moreover, in their work, the unidirectional motion vectors are assigned to bidirectional motion vectors.

In [35], Abdelazim et al. proposed the use of cross search algorithm in the H.265 standard that deals with high-efficiency video coding. In their work, the authors proposed a speed optimization technique in the frequency domain phase-correlation that enables compressing the videos rapidly while maintaining the video quality. In [36], Jia and Ding proposed a fast sub-pixel motion estimation algorithm. In their work, the authors proposed a scheme to skip sub-pixel search process in smooth

prediction units. Moreover, the authors proposed a fast sub-pixel search algorithm based on texture direction analysis in order to reduce the computational complexity. In [37], the authors presented a low computational complexity systolic hardware architecture for full search block matching algorithm. In their work, the proposed architecture is based on one-bit transform-based full search algorithm. The proposed motion estimation hardware architecture performs full search for four macro-blocks in parallel, where the proposed architecture was implemented in VHDL. In [38], the authors presented a three-step searching method in order to estimate the motion vectors of high-resolution image sequences using low number of computations. The searching strategy of this algorithm is carried in three steps, where the first search is performed in the large areas, the second is performed in the adaptive directional search and the last is performed in the small area search.

In [39], Arora et al. proposed a dynamic zero motion pre-judgment technique along with an adaptive diamond pattern search-based algorithm in order to enhance the search efficiency and accuracy of motion estimation. The dynamic zero motion pre-judgment is used for early identification of the stationary blocks. However, for the rest of the stationary blocks, an initial search center is used which has a high probability to be near actual motion vector. The variable size diamond pattern is used to obtain the global minima. In [40], Kovacevic et al. presented a motion estimation technique that combines recursive block-matching and customized phase plane correlation. In [41], Kamble et al. developed an approach for video coding using a modified three-step search block matching algorithm and weighted finite automata coding. In their work, the proposed block matching algorithm is based on the combination of rectangular and hexagonal search patterns and is used to compute motion vectors. The proposed weighted finite automata are used for the coding with a focus on reducing the encoding time. In order to reduce the encoding time, the authors in [47] proposed another approach for fractal coding using the weighted finite automata. The authors of [42] proposed a motion estimation method for image stabilization, integrating the speeded up robust features algorithm, modified random sample consensus and the Kalman filter. The authors achieved video stabilization with filtered motion parameters using the modified adjacent frame compensation.

In [43], the authors presented an enhanced version of the dynamic pattern search algorithm by means of reducing the search point computation. In their work, the algorithm starts by identifying the stationary blocks; then, the search points within the search area were evaluated for minimum distortion. The proposed work has been compared with other techniques like full search, diamond search and hexagon search. In [44], the authors proposed a two-step approach for enhancing the accuracy of initial search center prediction that is applied in the H.264 standard, in order to improve the motion estimation speed in video encoding. In their work, candidate blocks are identified in the first step for initial search center prediction. In the next step, the search is refined to obtain best possible initial search center.

In [45], the authors presented a hybrid approach for motion estimation. The hybrid method combines the dynamic zero motion pre-judgment technique with the initial search centers technique. In their work, calculating the initial search centers shifts according to the process of zero motion pre-judgment. In [46], the authors analyzed various tools involved in fast motion estimation algorithms. Moreover, the authors proposed a number of improvements in order to achieve a fast hybrid algorithm.

However, fewer researchers have investigated applying motion estimation algorithms in the frequency domain, such as the work of Young and Kingsbur [22] who proposed an alternate block matching method by applying a motion estimation technique based on overlapped transforms. Argyriou and Vlachos [1] in their work, proposed the use of gradient correlation in the frequency domain. Edrem et al. [6] estimated the motion parameters using a harmonic retrieval approach. Tzimiropoulos et al. [21] proposed a method for detecting symmetries in real images in the frequency domain. In their approach, the authors used motion estimation techniques to sequentially determine associated parameters. Pingault and Pellerin [17] tested motion transparency phenomena in video sequences based on the frequency domain. Their method contains an algorithm that introduces a new statistical model.

Hierarchical motion estimation algorithms are widely used for their accuracy; where in such algorithms several searching methods at different levels are applied. These types of algorithms are widely used due to their accuracy. However, applying those algorithms in the frequency domain has not yet been investigated properly. In hierarchical block matching techniques, the reliability of motion

vectors is related to the block size, where large blocks are more likely to converge on local minima. Moreover, in such algorithms, the advantages of selecting large blocks with small blocks at different levels are combined. Various research topics on the hierarchical search algorithms have been tackled in literature [25]-[29], [31].

In this work, a motion estimation algorithm based on a two-level hierarchy is proposed with a new block matching criterion to be applied at both levels of the hierarchy, as can be seen in Figure 1. The next section introduces the transformation method applied in this research.

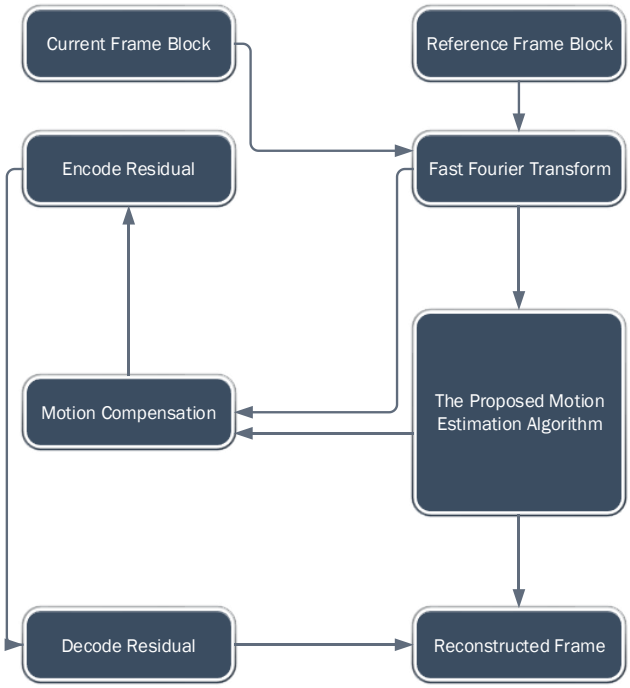


Figure 1. The architecture of general video encoders (the right side of the figure highlights the contribution in this research).

3. TRANSFORM DOMAIN

Video compression reduces the spatio-temoral redundancy that exists in the frame data and between consecutive frames using intra-frame and inter-frame coding methods. Intra-frame coding involves spatial to frequency transformation of the video frame and quantizing the frame frequencies by means of removing high frequencies that represent insignificant visual details in a given frame. Regardless of the transformation method that has been applied, it should be computationally acceptable and revertible [18]. In inter-frame coding, compression is achieved by utilizing the temporal redundancy using proper motion estimation and compensation algorithms. Various spatio-temporal transformation methods have been proposed in literature and are either image-based or block-based methods [9].

Block-based transformation methods are most applicable for use in video coding, since motion estimation algorithms are based on block matching methods. In this work, the Discrete Fourier Transform (DFT) is chosen, as it allows working in the frequency domain comparable to other transformation methods available in literature.

3.1 The Discrete Fourier Transform

Based on the Fourier theory, a complex signal can be decomposed into infinite series of cosine and sine terms and a group of coefficients that can be determined. The original function $f(t)$ can be decomposed into a series of basis states, based on (1).

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(\frac{2\pi nt}{T}) + \sum_{n=1}^{\infty} b_n \sin(\frac{2\pi nt}{T}) \quad (1)$$

where $a_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\cos(\frac{2\pi nt}{T})dt$, $b_n = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\sin(\frac{2\pi nt}{T})dt$.

The relationship above can be simplified as shown in (2)

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \exp(-i\frac{2\pi nt}{T}) \tag{2}$$

where $c_n = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\exp(i\frac{2\pi nt}{T})dt$, such that $c_0 = \frac{1}{2}a_0$, $c_n = \frac{1}{2}(a_n + ib_n)$ and $c_{-n} = \frac{1}{2}(a_n - ib_n)$.

In order to use Fourier transform with discrete input data such as the data available in digital videos and images, integrals are replaced by sums, T is replaced by N, $f(t)$ changes to x_n and c_n is replaced by X_n, which represents the Digital Fourier Transform shown in (3) and its inverse shown in (4). The DFT reveals periodicities in input data as well as the relative strengths of any periodic components [32].

$$X_n = \sum_{k=0}^{N-1} x_k \exp(i\frac{2\pi nk}{N}) \tag{3}$$

$$x_k = \frac{1}{N}\sum_{n=0}^{N-1} X_n \exp(-i\frac{2\pi nk}{N}) \tag{4}$$

Using (3), the N input samples (pixels) in a given block are converted into N frequency samples. The DFT is a coefficient matrix multiplication as shown in (5).

$$\begin{pmatrix} X_0 \\ .. \\ X_n \end{pmatrix} = \begin{pmatrix} W_{00} & .. & W_{0n} \\ .. & .. & .. \\ W_{n0} & .. & W_{nn} \end{pmatrix} \begin{pmatrix} x_0 \\ .. \\ x_n \end{pmatrix}, \; W_{nk} = \exp\left(i\frac{2\pi nk}{N}\right) \tag{5}$$

The above calculation is of order N^2. In order to reduce the DFT complexity, a number of researchers investigated the use of different patterns in W_{nk}. Amongst those approaches, the Fast Fourier Transform or FFT is proposed as a computational method in the order of $N \log N$. In this research, the Cooley–Tukey algorithm is used, as it is the most common FFT algorithm available in this domain [32] to transform the video frames with different block sizes at different levels of the two-level hierarchy. Moreover, only part of the frequencies is considered in the block matching criterion to get the best match as will be demonstrated in the next section.

4. THE PROPOSED HIERARCHICAL SEARCH ALGORITHM

Hierarchical block matching algorithms normally start the search process with small blocks and use their motion vectors as starting points to search for larger blocks in next hierarchies (the selected block sizes at each hierarchy affect the reliability of the produced motion vectors, where large blocks result in local minima.). Generally, in the spatial domain, three level hierarchical searches are used, as the data in its original form (pixel domain) is highly correlated. However, given the fact that data in the frequency domain is decorrelated, this facilitates the reduction of hierarchical levels needed to perform the matching process. Therefore, in this research, a two-level hierarchy in the frequency domain is used and proven to be sufficient. The proposed algorithm contains well-known algorithms in each level of the hierarchy with a new matching criterion (described in section 4.1.1) to be used at each level. The steps of the proposed algorithm are summarized in Algorithm-1 and visually represented in Figure 2.

4.1 The Proposed Matching Criterion

In order to compare algorithms in this domain, the standard Sum of Absolute Differences (SAD) shown in (6) is applied over all the possible searching positions.

$$SAD_{(m,n)} = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} |A - B| \tag{6}$$

Algorithm-1

Step 1: *sub-sampling level-1 (the lowest level that consists of the video frame at its full resolution) by a factor of 2 in vertical and horizontal directions to produce level-2.*

Step 2: *transforming the frames at level-1 and level-2 into the frequency domain using the FFT with* 4×4 *block size at level-2 and* 8×8 *at level-1.*

Step 3: *the search process starts from level-2 with* 4×4 *block sizes, with the TSS search algorithm (described in section 4.1.2) to get a coarse motion vector that will be passed to level-1, based on the proposed matching criterion described in section 4.1.*

Step 4: *the two-dimensional logarithmic search algorithm (described in section 4.1.3) with* 8×8 *block sizes is applied, based on the proposed matching criterion described in section 4.1. in order to get the final motion vector.*

Step 5: *the resulting motion vectors from step-4 are added to the previous image in order to obtain the next predicted image frame.*

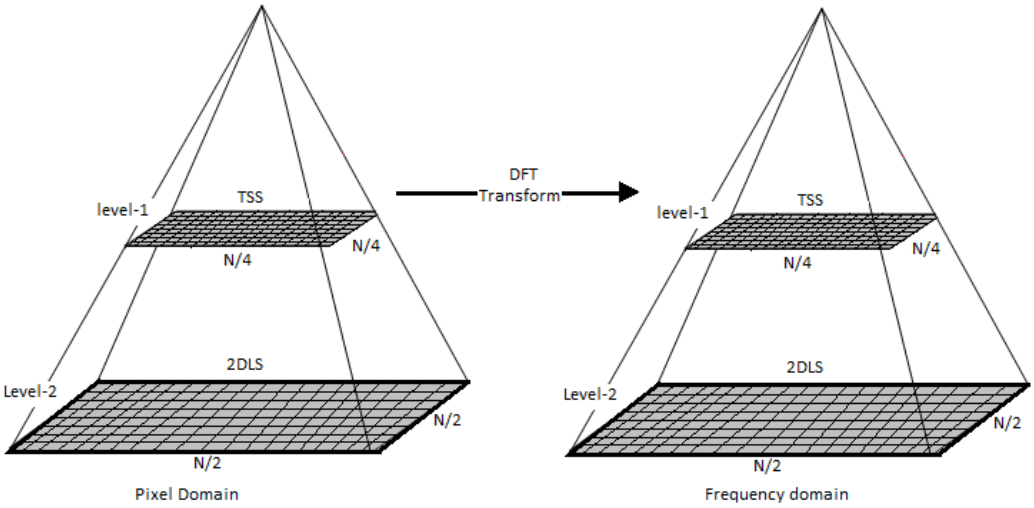


Figure 2. The proposed hierarchical search algorithm.

where $A = f_t(\lambda + x, \beta + y)$ is the block location coordinates, (λ, β) is the current block coordinates at the reference frame, $B = f_{t-1}(\lambda + m + x, \beta + n + y)$ is defined as the candidate block in the previous frame within the window size $-W \le m,n \le W$. The performance of the algorithm is highly dependent on the matching criterion. However, when applying the matching criterion in the spatial domain, the number of required computations cannot be reduced, as this will directly affect the matching results, since frame pixels are highly correlated. Therefore, in the frequency domain (where frame data is highly decorrelated), reducing the number of required computations is more appropriate.

In full-search algorithms, $SAD_{(m,n)}$ is computed at all $(2W+1)^2$ block positions within the search window. This results in a huge number of subtractions, additions and comparisons for each reference block. This massive number of computations can be reduced with fast search motion estimation techniques, where comparison criterion shown in (6) can be effectively reduced for search location.

The FFT produces frequency coefficients arranged in a pattern where the corners of coefficients block contain the lowest frequencies that describe the general vertical and horizontal information in the pixel block. However, the rest of the coefficients in the block include high frequencies that describe vertical and horizontal details in the pixel block. In this research, the coefficients at the four corners of the transformed block are only considered in the SAD matching criterion. Therefore, the total number of computations is reduced to a constant of 4 subtractions and 4 additions for each candidate block at each search position, instead of N^2 operations required by other algorithms in the spatial domain.

Information in these parts of the block is adequate to distinguish the desired block from the rest of the surrounding blocks as can be seen later in the experimental results section.

4.2 The TSS Algorithm

The steps of the TSS algorithm are applied at level-1 of the hierarchy in both of the current and previous frames as shown in Algorithm-2.

Algorithm-2

Step 1*:* Set the window size (W) to 2^N+1, where $N=2$ (N= number of levels in the hierarchical *search).*

Step 2*:* Set the step size (S) to 2^N size, where $N=2$.

Step 3*:* Start with search location at the center and apply the following:

a) *The eight locations at +/- S around location (0, 0) are to be searched and the one with the minimum SAD is selected based on the matching criterion described in section 4.1.*

b) *The search origin is set to the above selected location and the step size is reduced by a factor of 2*

c) *The search repeats until S = 1 and the location with minimum SAD is considered as the best match in level-1.*

Step 4*: Pass the obtained coarse motion vector to level-2.*

In the TSS algorithm, the total number of computations is reduced when compared to the full search algorithm by a factor of 9. Instead of evaluating 225 blocks, the TSS only evaluates 25 blocks.

4.3 The Two-Dimensional Logarithmic Search Algorithm

The two-dimensional logarithmic search algorithm is closely related to the three-step search algorithm. This algorithm requires more steps than the three-step search; however, it has a better accuracy. The two-dimensional logarithmic search algorithm is described in Algorithm-3:

Algorithm-3

Step 1: Set the window size to 4^N+1, $N=2$ (N= number of levels in the hierarchical search).

Step 2: Set the step size to 4^N, where $N=2$.

Step 3: Start with search location at the center and apply the following:

a) *Search the 4 locations at S distance from the center on the vertical and horizontal directions.*

b) *Amongst the searched locations, select the one with minimum SAD based on the matching criterion described in section 4.1.*

c) *If a point other than center has the minimum SAD, then this point is considered as the new center.*

d) *Repeat steps b and c*

e) *If the minimum SAD is given by the center point, then set S = S/2.*

f) *If S = 1, search the eight locations surrounding the center at S distance.*

g) *Set the motion vector as the point with the minimum SAD.*

The resulting motion vectors from this step will be added to the previous frame in order to obtain the next predicted image frame. The TDLS algorithm is related to the TSS algorithm; however, it is used for a large search window size.

5. EXPERIMENTAL RESULTS AND DISCUSSION

In order to test the efficiency of the proposed methods, two sets of standard testing videos are used in this research (shown in Table 1 and Table 2). The first set comprises a total of six standard videos of type CIF with an aspect ratio of 4:3. The CIF (Common Interchange Format) is a video format initially proposed in the H.261 standard. This video format is used to standardize the horizontal and vertical resolutions of YCbCr sequences in video signals. This type of video is used in standard video teleconferencing systems. CIF defines a video sequence with a resolution of 352 × 288 and a frame rate of 30 frames per second with color encoding using the YCbCr 4:2:0 standard, where the selected video sequences consist of 300 frames in each sequence. The second set of videos comprises a total of 3 High Definition (HD) videos (1080p) with an aspect ratio of 16:9 and color encoding using the YCbCr 4:4:4 standard, where the selected video sequences consist of 500 frames in each sequence.

The selected videos from both sets are well-known standard videos that are used to test the efficiency and compare the work with other benchmark algorithms. These video sequences from both sets are selected with increasing motion complexity ranging from slow to high motion complexity. More than 3300 video frames from the different sequences were used in the experiments and are listed in Table 1 and Table 2. The well-known PSNR is used to evaluate the proposed motion estimation algorithm performance. The $PSNR = 10\log_{10}(L^2 / MSE)$, where L represents the range of pixel values. The Mean Square Error comparison criterion measures the similarity between the frame pixels and is measured as follows: $MSE = 1/N\sum_{i=1}^{N}(x_i - y_i)^2$, where N represents the number of pixels inside the frame, x_i, y_i represent the pixel value inside the original and predicted frames, respectively. High PSNR values indicate better quality. A PSNR result above 30dB means that changes caused by compression algorithms cannot be visually recognized.

Applying the PSNR between the original and reconstructed frames measures the efficiency of the proposed work. Therefore, Table 1 compares the obtained PSNR values of the proposed algorithm with those from other state-of-the-art algorithms in this domain using videos from the first set. Figure 3 and Figure 4 visually represent the results in Table 1. Table 2 shows and compares the PSNR results of the proposed work with those of state-of the-art algorithms using HD videos (1080p) in the second set. Using the first set of test videos (CIF), and as can be seen in Table 1, the proposed work outperforms the standard three-step search [13], two-dimensional logarithmic search [10] and the diamond search algorithm [23] with 22%, 28% and 23% average enhancement, respectively. Moreover, when compared to the well-know KSHS algorithm [20], ETSS [12], CDMHS [14] and FHFS [30], the proposed algorithm outperforms those algorithms with 12%, 16%, 7% and 2%, respectively.

Using the HD (1080p) videos set, the proposed work outperforms the standard three-step search [13], two-dimensional logarithmic search [10] and the diamond search algorithm [23] with 12%, 26% and 24% average enhancement, respectively. Moreover, when compared to the well-know KSHS algorithm [20], ETSS [12], CDMHS [14] and FHFS [30], the proposed algorithm outperforms those algorithms with 11%, 10%, 9% and 5%, respectively. Figure 5 provides a visual representation of the average PSNR values shown in Table 1 and Table 2 and represents the results of the proposed algorithm compared to the rest of the state-of-the-art algorithms when applied to HD (1080p) high-resolution videos and normal CIF standard videos. Figure 6 shows samples of the reconstructed frames from the HD (1080p) videos listed in set 2. Figure 7, Figure 8 and Figure 9 provide samples of the reconstructed frames listed in set 1 that contains the standard CIF well-known videos.

The complexity of the proposed work is compared against the benchmark block-based motion estimation algorithms. Let w be the search window size, N represents the block size (hierarchy-based algorithms use various block sizes at each level of the hierarchy), the full search algorithm requires $(2w+1)^2 \times (2N^2 - 1)$ additions and $(2w+1)^2 \times N^2$ absolute differences, a. The cross diamond modified hierarchical search requires $(2w+1)^2 \times (2N^2 - 1)$ additions at level-1, $8 \times (2N^2 - 1)$ additions at level-2 and $23 \times (2N^2 - 1)$ additions at level-3; in addition to $(2w+1)^2 \times N^2$ absolute differences at level-1, $8 \times N^2$ and $8 \times N^2$ at level-2 and level-3, respectively. The Kalman simplified hierarchical search requires $(2w+1)^2 \times (2N^2 - 1), 8 \times (2N^2 - 1)$ and $8 \times (2N^2 - 1)$ at level-1, level-2 and level-3,

respectively. In terms of the number of absolute differences, the algorithm requires $(2w+1)^2 \times N^2$, $8 \times N^2$ and $8 \times N^2$ at each level, respectively. The proposed algorithm requires $15 \times ((N^2/2)-1)$ and $23 \times ((N^2/2)-1)$ additions at level-1 and level-2, respectively and $8 \times (N^2/4)$, $23 \times (N^2/4)$ comparisons at both levels. Using an appropriately sized search window (w) and an appropriate block size (N) for each algorithm, the proposed algorithm requires less than 1% of the total number of additions and the total number of absolute differences compared to the full search algorithm. When compared to the rest of the algorithms, the algorithm requires less than 5% of the total complexity required by the cross-diamond modified hierarchical search algorithm and less than 15% of the total complexity required by the Kalman simplified hierarchical search. This complexity reduction can be attributed to the substantial reduction in the total number of operations required in the proposed matching criterion.

Generally, in order to evaluate the use of motion estimation algorithms in the frequency domain, a performance comparison is conducted that evaluates the standard full search algorithm when applied in both pixel and frequency domains. Table 3 presents the resulted PSNR values of the full search algorithm implemented in both domains based on the standard set of HD (1080p) test videos (first 50 frames of each video are included in the test). As shown in this table (Table 3), the resulting average PSNR in the frequency domain is slightly better than that in the pixel domain. This small enhancement does not cover the cost of the extra complexity caused by the transformation process. However, it can achieve far better results (in terms of complexity reduction) when accompanied with proper search and matching techniques.

The direct implementation of Discrete Fourier Transform (DFT) requires $O(N^2)$ operations. However, when using Fast Fourier Transform (FFT), this can be reduced to $O(N \log(N))$, resulting in a substantial difference in the tractability of the DFT. The fact that the transition between the domains can be computed efficiently allows for more efficient implementations of the DFT.

Table 1. The resulting PSNR values of the proposed algorithm and the rest of the standard algorithms when applied to the standard set of CIF test video sequences from set-1.

Standard Videos (CIF 352x288)	TSS [13]	2DLS [10]	DS [23]	KSHS [2]	ETSS [12]	CDMHS [14]	FHFS [30]	**Proposed work**
Akiyo	31.9	29.5	30.8	34.2	32.6	36.3	35.2	**36.4**
Mother and Daughter	32.0	30.8	31.7	34.7	31.2	36.2	37.5	**36.5**
News	30.2	29.4	30.2	33.9	32.6	34.3	36.2	**37.2**
Hall	30.9	28.8	30.0	32.4	32.9	33.8	35.9	**37.7**
Flower Garden	28.3	28.0	29.2	30.8	30.5	32.4	36.1	**36.8**
Football	27.1	25.7	27.5	31.9	30.7	33.0	35.7	**36.0**
Average PSNR	**30.1**	**28.7**	**29.9**	**33.0**	**31.8**	**34.3**	**36.1**	**36.8**

Table 2. The resulting PSNR values of the proposed algorithm and the rest of the standard algorithms when applied to the standard set of HD test video sequences from set-2.

Standard Videos HD (1080p)	TSS [13]	2DLS [10]	DS [23]	KSHS [2]	ETSS [12]	CDMHS [14]	FHFS [30]	***Proposed work***
Park_Joy	42.8	39.1	39.8	40.5	43.2	43.4	42.1	**43.8**
In-To-Tree	38.6	35.9	36.1	42.8	39.6	32.9	44.6	**45.2**
Station	36.7	32.4	33.3	38.4	36.9	42.3	43.7	**45.7**
Blue_Sky	41.2	34.7	35.2	39.0	42.2	45.2	40.4	**44.1**
Average PSNR	**39.8**	**35.5**	**36.1**	**40.2**	**40.1**	**41.0**	**42.7**	**44.7**

Table 3. The PSNR values of the full search algorithm implemented in pixel and frequency domains based on the standard set of HD (1080p) test videos.

Full Search Algorithm Implementation	*Park_Joy*	*In-To-Tree*	*Station*	*Blue_Sky*	***Average PSNR***
Pixel domain	45.19	47.78	46.31	46.46	**46.44**
Frequency domain	46.24	46.93	45.70	47.81	**46.67**

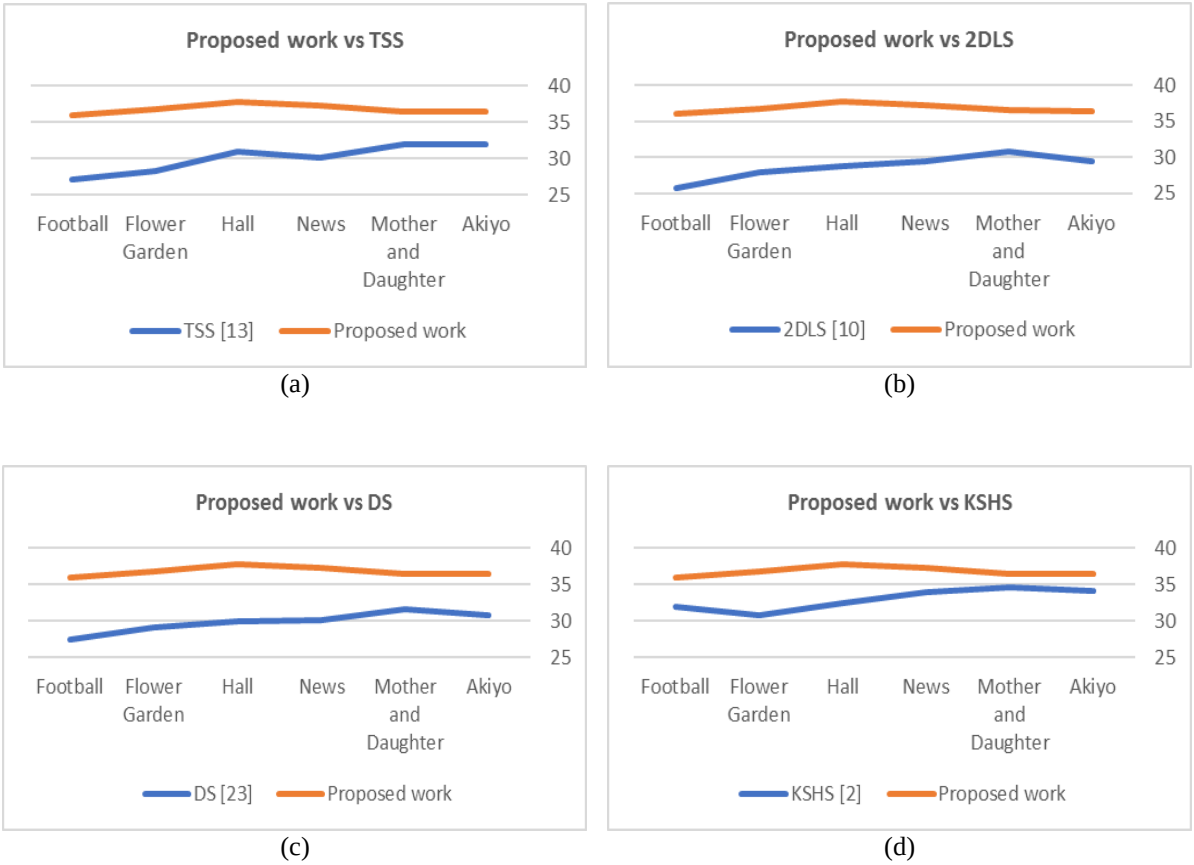
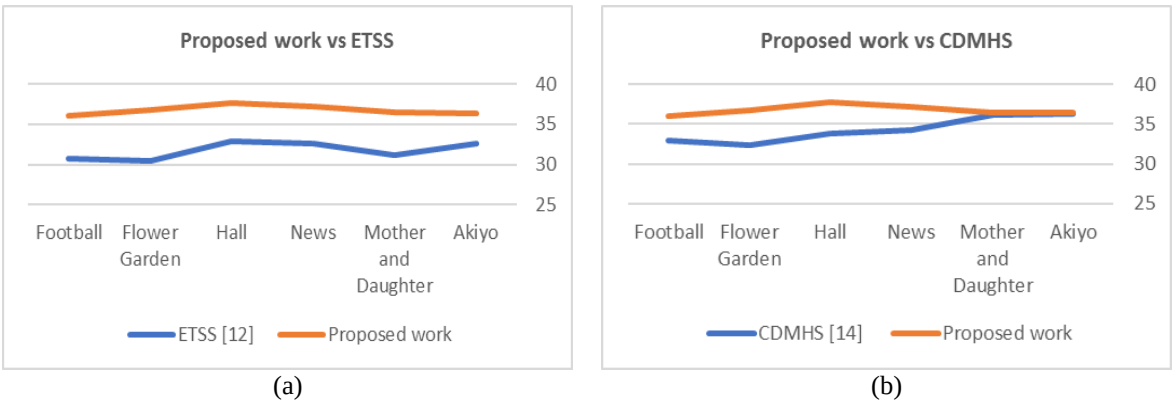
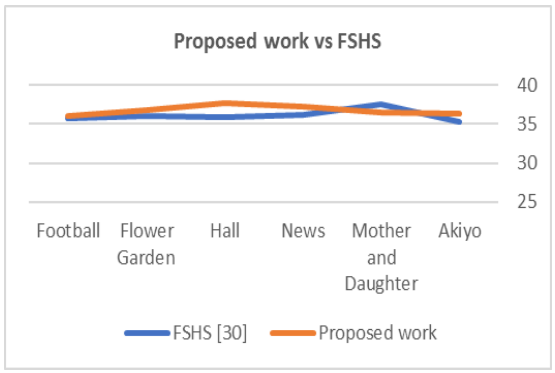


Figure 3. Visual representation that compares the proposed work with well-known algorithms as follows: (a) three-step search, (b) two-dimensional logarithmic search, (c) diamond search, (d) Kalman simplified hierarchical search.



Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.



(c)

Figure 4. Visual representation that compares the proposed work with well-known algorithms as follows: (a) enhanced three-step search, (b) cross diamond modified hierarchical search, (c) frequency-based fast hierarchical search.

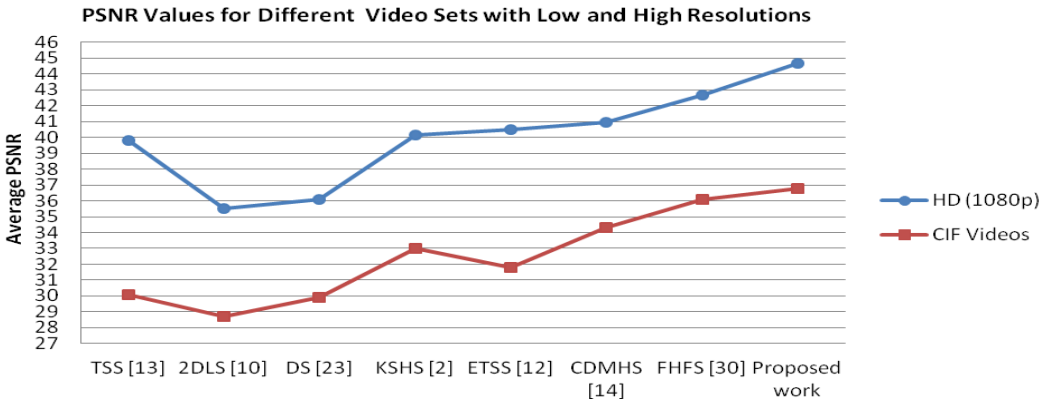


Figure 5. Visual representation of the average PSNR values shown in Table 1 and Table 2, that represents the results of the proposed algorithm compared to the rest of the state-of-the-art algorithms when applied to HD (1080p) high-resolution videos and normal CIF standard videos.

Figure 6. From top to bottom and from left to right, the reconstructed frames from "Park_Joy", "In-To-Tree", "Station" and "Blue_Sky" video sequences.


Figure 7. Sample of standard videos consisting of low motion activities, from left to right, reconstructed “Akiyo” and reconstructed “Mother and Daughter” video frames.

Figure 8. Sample of standard videos consisting of moderate motion activities, from left to right, reconstructed “News” and reconstructed “Hall” video frames.


Figure 9. Sample of standard videos consisting of high complex motion activities, from left to right, reconstructed “Flower Garden” and the reconstructed “Football” video, frames.

6. CONCLUSIONS

Digital videos consist of successive frames sampled over a period of time and carry high data redundancy. Digital video sizes can be massively reduced by eliminating redundant bits which can be achieved by proper compression methods. Various types of compression methods have been proposed in literature and during the last few years, many algorithms have been proposed to compress the massive amount of data available in digital videos while maintaining as much of the visual quality as possible. Motion estimation techniques based on block matching algorithms have been widely used for this purpose. In block matching techniques, each video frame is divided into blocks of similar sizes that contain frame pixels. Object movements successive video frames are searched and investigated on

block basis. In this work, block matching is applied in the frequency domain, where a group of carefully chosen frequencies that correctly identify each block distinctively is tested. The algorithm proposed in this research has reduced the total number of required operations, significantly reducing the algorithm's complexity. The proposed algorithm has been tested using standard test videos and has proven to outperform other state-of-the-art algorithms. Two sets of standard test videos were used in this work, the first set is comprised of the standard CIF videos and the other set is comprised of the standard HD (1080p) videos. Using the standard set of CIF videos, the proposed work outperforms the standard three-step search, two-dimensional logarithmic search and the diamond search algorithms with 22%, 28% and 23% average enhancement, respectively. Moreover, when compared to the well-know Kalman simplified hierarchical search algorithm, the enhanced three-step search algorithm, the cross diamond modified hierarchical search and the frequency-based Hierarchical fast search, the proposed algorithm outperforms those algorithms with 12%, 16%, 7% and 2%, respectively. Moreover, using the standard HD (1080p) videos set, the proposed work outperforms the standard three-step search, two-dimensional logarithmic search and the diamond search algorithms with 12%, 26% and 24% average enhancement, respectively. When compared to the well-know Kalman simplified hierarchical search algorithm, the enhanced three-step search algorithm, the cross diamond modified hierarchical search and the frequency-based hierarchical fast search, the proposed algorithm outperforms those algorithms with 11%, 10%, 9% and 5%, respectively. The complexity of the proposed work is compared against the benchmark block-based motion estimation algorithms. Results show that the proposed algorithm requires less than 1% of the total number of additions and the total number of absolute differences when compared to the full search algorithm. Moreover, when compared to the rest of the algorithms, the proposed work requires less than 5% of the total complexity required by the cross-diamond modified hierarchical search algorithm and less than 15% of the total complexity required by the Kalman simplified hierarchical search.

REFERENCES

[1] V. Argyriou and T. Vlachos, "Quad-Tree Motion Estimation in the Frequency Domain Using Gradient Correlation," IEEE Transactions on Multimedia, vol. 9, no. 6, pp. 1147-1154, 2007.

[2] J. Cai and W. Pan, "On Fast and Accurate Block-based Motion Estimation Algorithms Using Particle Swarm Optimization," Elsevier Information Sciences, vol. 197, pp. 53-64, 2012.

[3] E. Cuevasa, D. Zaldívara, M. Pérez-Cisneros and D. Oliva, "Block-matching Algorithm Based on Differential Evolution for Motion Estimation," Elsevier Engineering Applications of Artificial Intelligence, vol. 26, no. 1, pp. 488-498, 2013.

[4] E. Cuevasa, D. Zaldívara, M. Pérez-Cisneros, H. Sossa and V. Osuna, "Block matching algorithm for motion estimation based on Artificial Bee Colony (ABC)," Elsevier Applied Soft Computing, vol. 3, no. 6, pp.3047-3059, 2013.

[5] Z. Cui, G. Jiang, S. Yang and C. Wu, "A New Fast Motion Estimation Algorithm Based on the Loop–epipolar Constraint for Multi-View Video Coding," Elsevier Signal Processing: Image Communication, vol. 27, no. 2, pp. 172–179, 2012.

[6] C. Erdem, G. Karabulut, E. Yanmaz and E. Anarim, "Motion Estimation in the Frequency Domain Using Fuzzy c-planes Clustering," IEEE Transactions on Image Processing, vol. 10, no. 2, pp. 1873-1879, 2001.

[7] J. Fabrizio, S. Dubuisson and D. Béréziat, "Motion Compensation Based on Tangent Distance Prediction for Video Compression," Elsevier Signal Processing: Image Communication, vol. 27. no. 2, pp. 153–171, 2012.

[8] M. Ghanbari, "Video Coding: An Introduction to Standard Codecs," Institution of Electrical Engineers, 1999.

[9] R. Gonzales and R. Woods, Digital Image Processing, 3rd Edition, Prentice Hall, 2008.

[10] J. R. Jain and A. K. Jain, "Displacement Measurement and Its Application in Interframe Image Coding," IEEE Transactions on Communincations, vol. 29, no. 12, pp. 1799-1808, 1981.

[11] C. Je and H-M. Park, "Optimized Hierarchical Block Matching for Fast and Accurate Image Registration," Elsevier Signal Processing: Image Communication, vol. 28, no. 7, pp. 779-791, 2013.

[12] X. Jing and L. P. Chau, "An Efficient Three-step Search Algorithm for Block Motion Estimation," IEEE Transactions on Multimedia, vol. 6, no. 3, pp. 435-438, 2004.

[13] T. Koga, K. Iinuma, A. Hirano, Y. Iijima and T. Ishiguro, "Motion Compensated Interframe Coding for Video Conferencing," in: Proc. Nat. Telecommunications Conference, pp. G5.3.1-5.3.5, 1981.

[14] N. Al-Najdawi, M. N. Al-Najdawi and S. Tedmori, "Employing a Novel Cross-diamond Search in a Modified Hierarchical Search Motion Estimation Algorithm for Video Compression," Elsevier Information Sciences, vol. 268, pp. 425-435, 2014.

[15] K. Lai, Y. Chan, C. Fu and W. Siu, "Hybrid Motion Estimation Scheme for Secondary SP-frame Coding Using Inter-Frame Correlation and FMO," Elsevier Signal Processing: Image Communications, vol. 27, no. 1, pp. 1-15, 2012.

[16] F. Di Martino, V. Loia and S. Sessa, "Fuzzy Transforms for Compression and Decompression of Color Videos," Elsevier Information Sciences, vol. 180, no. 20, pp. 3914-3931, 2010.

[17] M. Pingault and D. Pellerin, "Motion Estimation of Transparent Objects in the Frequency Domain," Journal of Signal Processing, vol. 4, no. 8, 2004.

[18] I, Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia, John Wiley & Sons, 2003.

[19] I. Richardson, Video Codec Design, John Wiley & Sons, 2002.

[20] S. Tedmori and N. Al-Najdawi, "Hierarchical Stochastic Fast Search Motion Estimation Algorithm," IET Computer Vision, vol. 6, no. 1, pp. 21-28, 2012.

[21] G. Tzimiropoulos, V. Argyriou and T. Stathaki, "Symmetry Detection Using Frequency Domain Motion Estimation Techniques," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 861-864, 2008.

[22] R. Young and N. Kingsbury, "Frequency-domain Motion Estimation Using a Complex Lapped Transform," IEEE Transactions on Image Processing, vol. 2, no. 1, pp. 2-17, 1993.

[23] S. Zhu and K. K. Ma, "A New Diamond Search Algorithm for Fast Block Matching Motion Estimation," IEEE Transactions on Image Processing, vol. 9, no. 2, pp. 287-290, 2000.

[24] H. Alzoubi and W. Pan, "Fast and Accurate Global Motion Estimation Algorithm Using Pixel Subsampling," Elsevier Information Sciences, vol. 178, no. 17, pp. 3415-3425, 2008.

[25] K. M. Nam, J. S. Kim, R. H. Park and Y. S. Shim, "A Fast Hierarchical Motion Vector Estimation Algorithm Using Mean Pyramid," IEEE Transactions on Circuits and Systems for Video Technology, vol. 5, no. 4, pp. 344–351, 1995.

[26] C. Kuo, C. Hsieh and C. Chao, "Multiresolution Video Coding Based on Kalman Filtering Motion Estimation," Elsevier Journal of Visual Communication and Image Representation, vol. 13, no. 1, pp. 348–362, 2002.

[27] B. C. Song and J. B. Ra, "A Hierarchical Block Matching Algorithm Using Partial Distortion Criterion," in: Proc. SPIE 3309, Visual Communications and Image Processing, pp. 88-95, 1998.

[28] C. K. Cheung and L. M. Po, "A Hierarchical Block Motion Estimation Algorithm Using Partial Distortion Measure," in: Proc. International Conference on Image Processing, pp. 606-609, 1997.

[29] S. Basah, A. Bab-Hadiashar and R. Hoseinnezhad, "Conditions for Motion-background Segmentation Using Fundamental Matrix," IET Computer Vision, vol. 3, no. 4, pp. 189-200, 2009.

[30] N. Al-Najdawi, S. Tedmori, O. Alzubi, O. Dorgham and J. Alzubi, "A Frequency-Based Hierarchical Fast Search Block Matching Algorithm for Fast Video Communication," International Journal of Advanced Computer Science and Applications, vol. 7, no. 4, pp. 447-455, 2016.

[31] Z. Weng, G. Chen, L. Shieh and J. Larsson, "Evolutionary Programming Kalman Filter," Elsevier Information Sciences, vol. 129, no 1-4, pp. 197-210, 2000.

[32] R. Tolimieri, M. An and C. Lu, "Algorithms for Discrete Fourier Transform and Convolution," Springer Science and Business Media, New York, pp. 55-70, 1997.

[33] MA. Diaz-Cortes, E. Cuevas and R. Rojas, "Motion Estimation Algorithm Using Block-Matching and Harmony Search Optimization," Engineering Applications of Soft Computing, Intelligent Systems Reference Library, vol. 129, 2017.

[34] S. Yu and J. Jeong, "Multidirectional Motion Estimation Algorithm for Frame Rate Up-conversion,"

21st International Conference on Circuits, Systems, Communications and Computers, 2017.

[35] A. Abdelazim, A. Hamza and D. Ait-Boudaoud, "Cross Search Frequency Domain Motion Estimation Algorithm for the High Efficiency Video Coding Standard," 9th IEEE-GCC Conference and Exhibition, 2017.

[36] S. Jia and W. Ding, "A Fast Sub-Pixel Motion Estimation Algorithm for HEVC," IEEE International Symposium on Circuits and Systems, 2016.

[37] P. Muralidhar and C. Rao, "High Performance Architecture of Motion Estimation Algorithm for Video Compression," Journal of Circuits, Systems and Computers, vol. 25, no. 8, 2016.

[38] S. Chang and R. Wang, "Novel Motion Estimation Algorithm for Image Stabilizer," Engineering Computations, vol. 34, no. 1, pp. 77-89, 2017.

[39] S. Arora, N. Rajpal and R. Purwar, "A New Fast Motion Estimation Algorithm Using Adaptive Size Diamond Pattern Search with Early Search Termination," International Journal of Computational Vision and Robotics, vol. 7, no. 6, pp. 623-643, 2017.

[40] V. Kovacevic, Z. Pantic, A. Beric and R. Jakovljevic, "Block-Matching Correlation Motion Estimation for Frame-Rate Up-Conversion," Journal of Signal Processing Systems, vol. 84, no. 2, pp. 283-292, 2016.

[41] S. Kamble, N. Thakur and P. Bajaj, "Modified Three-Step Search Block Matching Motion Estimation and Weighted Finite Automata-based Fractal Video Compression," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 4, pp. 27-39, 2017.

[42] X. Cheng, Q. Hao and M. Xie, "A Comprehensive Motion Estimation Technique for the Improvement of EIS Methods Based on the SURF Algorithm and Kalman Filter," Sensors Journal, vol. 16, no. 4, pp. 1-15, 2016.

[43] R. Purwar, "Enhanced Dynamic Pattern Search Algorithm with Weighted Search Points for Fast Motion Estimation," Signal, Image and Video Processing, vol. 11, no. 6, pp. 1001-1007, 2017.

[44] S. Arora, N. Rajpal, K. Khanna and R. Purwar, "Improved Accuracy in Initial Search Center Prediction to Fasten Motion Estimation in H.264/AVC," IETE Journal of Research, vol. 62, no. 6, pp. 842-851, 2016.

[45] S. Arora, K. Khanna and N. Rajpal, "A Novel Hybrid Approach for Fast Block-based Motion Estimation," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 6, pp. 24-30, 2017.

[46] P. Nalluri, L. Alves and A. Navarro, "Complexity Reduction Methods for Fast Motion Estimation in HEVC," Image Communication Journal, vol. 39, pp. 280-292, 2015.

[47] S. Kamble, N. Thakur and P. Bajaj,"Fractal Coding Based Video Compression Using Weighted Finite Automata," International Journal of Ambient Computing and Intelligence, vol. 9, no. 1, 2018.

ملخص البحث:

على الرغم من التطورات التي حدثت في مجال حُلول معدات الحاسوب، ما زالت مُرمِّزات الفيديو عالية الجودة الخاصة بتطبيقات الزمن الحقيقي مجالاً مفتوحاً للبحث. فقد تم اقتراح العديد من الخوارزميات التي تهدف الى التقليل من عدد العمليات الرياضية المطلوبة. وكانت النتائج تلتقي عند قيم صغرى محلية وتستمر الحاجة الى إجراء عدد كبير من العمليات الحسابية. لهذا تقترح هذه الورقة طريقة مواءمة هرمية تسهل نقل الفيديوهات عالية التفاصيل بالمقارنة مع طرق الاتصال المعيارية. والجدير بالذكر أنّ الخوارزمية المقترحة مبنية على الحقل الترددي؛ إذ تفحص الخوارزمية التشابهات بين مجموعة فرعية مختارة من الترددات، مما يقلل الى حد كبير من العدد الكلي من المقارنات والحسابات الرياضية الإجمالية المطلوبة لكل كتلة.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

ENHANCED UWB PRINTED MONOPOLE ANTENNA BASED ON GROUND PLANE MODIFICATIONS

Noor M. Awad[1], Mohammed K. Abdelazeez[2] and Ahmad Al-Sharif[3]

(Received: 16-Dec.-2017, Revised: 05-Feb.-2018, Accepted: 18-Feb.-2018)

ABSTRACT

In this paper, a UWB antenna with an enhanced bandwidth is proposed. The enhanced ultra wide band (UWB) antenna consists of a rectangular patch fed by a 50 Ω microstrip feed line and partial ground plane. The bandwidth enhancement is achieved by making three modifications on the partial ground plane; adding two rectangular sleeves, adding one rectangular groove and adding two rectangular slots. The characteristics of this antenna are investigated using high frequency structure simulator (HFSS). The proposed design achieves large bandwidth at return loss RL ≥ 10 dB of (3.4 - 22.4) GHz (147.29%). Promising peak gain with good impedance matching and omni-directional radiation pattern are obtained.

KEYWORDS

Ultra Wide Band (UWB), Sleeves, Grooves, Slots, Reflection Coefficient, Gain, Bandwidth.

1. INTRODUCTION

Printed circuit antennas are becoming more and more under consideration with a great number of research papers, reports and books to deal with this type of antennas. The rapid growth in the wireless communication systems creates demands for wideband antenna, which should have high gain and large bandwidth covering all frequency ranges used in these systems. In 2002, the Federal Communication Commission (FCC) approved the first report for UWB technology to be operating in the frequency range (3.1 - 10.6) GHz with maximum radiated power of -41.3 dBm/MHz [1]. These UWB patch antennas are designed with different geometries; i.e., triangular [2], circular disc [3] and rectangular [4]. Several methods are used to enhance the antenna bandwidths; in [5] the bandwidth enhancement is achieved by making modifications to the patch and the partial ground plane. Circular shaped slots and a ring are inserted into the patch while diagonal cuts at the top corners of the partial ground plane with two rectangular slots making the bandwidth (2.75 – 20) GHz (151.6%). In [6], modifying the ground plane is achieved with diagonal edges, rectangular slot and T-shape cut added to get -10 dB bandwidth of (2.957 – 11.89) GHz (120.27%). In [7], modifying the rectangular patch is achieved with three steps, single slot besides adding slots to the ground plane to enhance the bandwidth of (6 – 12) GHz (33.3%). In [8], three types of slotted antenna are presented to enhance the bandwidth; T- slotted for patch and feed line, couple a ring and L slots and dual symmetry L slots. In [9], the ground plane is truncated and the patch has round junction and two chamfers to get a bandwidth of (1.79 – 28.02) GHz (175.98%).

Other means are used including the utilization of feeding structure of a trident-shaped strip and a tapered impedance transformer to provide an antenna with a bandwidth of (2.75 – 16.2) GHz (141.95%) [10]. Three round ground grooves are used in [11] to get percentage bandwidth of (2.87 – 50) GHz (178.29%). One round cut in each corner of the patch antenna and one ground groove are used in [12] to get a bandwidth of (3.42 – 11.7) GHz (109.52%). Making triangular shape slots on the top of the partial ground plane in [13] increases the UWB antenna bandwidth to (2.95 – 15.45) GHz 135.87%. Adding a rectangular slot at the top edge of the partial ground plane for a circular patch UWB antenna [14] increases the bandwidth to reach (3.3 – 20) GHz 193%. One may think about increasing the substrate height, but this produces surface wave, which will reduce the antenna efficiency.

In this paper, we introduce a simple design with small size to get a large bandwidth for UWB applica-

This paper is an extended version of a short paper that was presented at the International Conference "2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)", 11-13 October 2017, Jordan.

- N. M. Awad, M. K. Abdelazeez and A. Al-Sharif are with Electrical Engineering Department, The University of Jordan, Amman, Jordan.

Emails: [1]n.awad@ju.edu.jo, [2]abdelazeez@ieee.org and [3]more.9.aw@gmail.com.

tions. The antenna consists of a microstrip feed line with rectangular patch and a modified partial ground plane as investigated in Section 2. Simulation results, using the HFSS, and discussion are presented in Section 3. Experimental verifications are outlined in Section 4 and finally, the conclusions and references are presented.

2. ANTENNA DESIGN

The antenna parameters for the proposed design are shown in Figure 1. All parameters are optimized to achieve best performance. The antenna dimensions (all lengths in mm) are as follows: the substrate is FR4-epoxy with thickness h = 1.6, tan(δ) = 0.02, ε_r = 4.4, width Ws = 30 and length Ls = 40. The patch length Lp =14 and width Wp =15. The partial ground plane length Lg = 13, width Wg = Ws = 30. The two ground sleeves length h1 = 1.15, width W1 = 1.25 and are located at y1 = 11 from the antenna edge. The ground groove length h2 = 2, width W2 = 1.5. The two ground slots are located at y2 = 7 from the edge and have width W3 = 1.5 and length h3 = 7. The microstrip feed line is designed for 50 Ω impedance with width W_f = 3.4 and length L_f = 14.

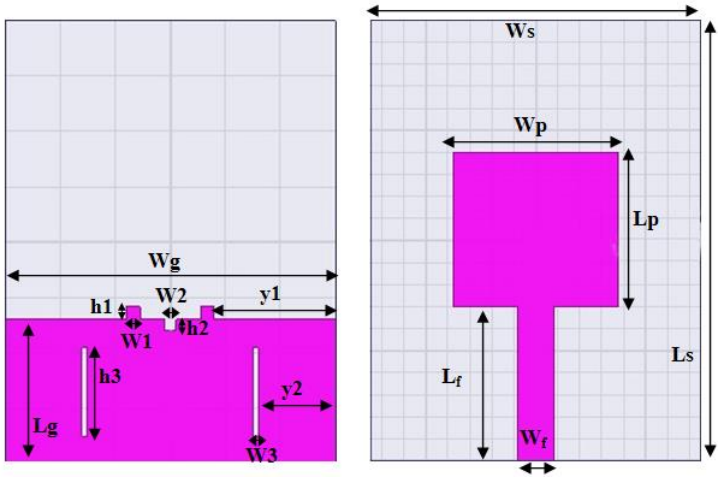


Figure 1. Antenna structure.

The antenna dimensions are obtained through parametric analysis and optimization process in order to achieve stable radiation characteristics over the frequency range of interest. Parametric studies are carried out on groove numbers and dimensions, the ground sleeves and the ground slots dimensions. In simulation, only one antenna parameter was varied each time while others were kept constant.

3. RESULTS AND DISCUSSION

The proposed design started with a simple rectangular patch, microstrip feed line and simple partial ground plane. The simulation of scattering parameter (reflection coefficient) S11 (Return Loss (RL) = - S11) versus frequency, at S11 ≤ - 10 dB, conducted using HFSS software tool, shows low covered bandwidth (3.36 - 9.64) GHz, such that the fractional (percentage) bandwidth is 96.6%. The antenna impedance matching and the covered bandwidth are enhanced by introducing three modifications in the simple antenna ground plane.

The ground modifications were added through three parts:

1) Adding sleeves as ground plane extension: these behave as an additional inductive element that generates additional resonant mode, which is used for either dual or multiband operations, or it is combined with the fundamental mode to improve the overall bandwidth [15]. Parametric analysis is carried out on the sleeves dimensions (width, length and location), number and shape. The variation of S11 versus frequency when using the ground sleeves is shown in Figure 2. Adding one or three ground sleeves degrades the overall impedance matching and subsequently the covered bandwidth compared to adding two sleeves. Sleeves width W1 is varied in the range (1.25 - 2.25) mm, as shown in Figure 3, where it can be observed that best performance is obtained when the sleeve width W1 = 1.25 mm. The sleeves length h1 variation is also studied and varied in the range (1 - 1.3) mm, as shown in Figure 4. It

is found that the resonant frequency values are highly dependent on sleeves length, choosing h1 = 1.15 mm for best impedance matching over the entire covered frequency band. The sleeves location variation from the antenna edge y1 has a high effect on S11 versus frequency as shown in Figure 5, where moving the sleeves towards the antenna edge makes the impedance matching better and yields respectively large bandwidth.

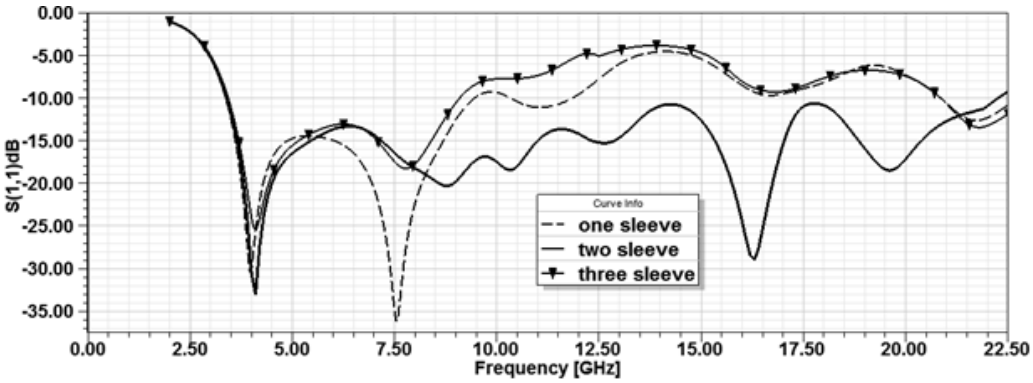


Figure 2. The scattering parameter S11 when varying the sleeves number.

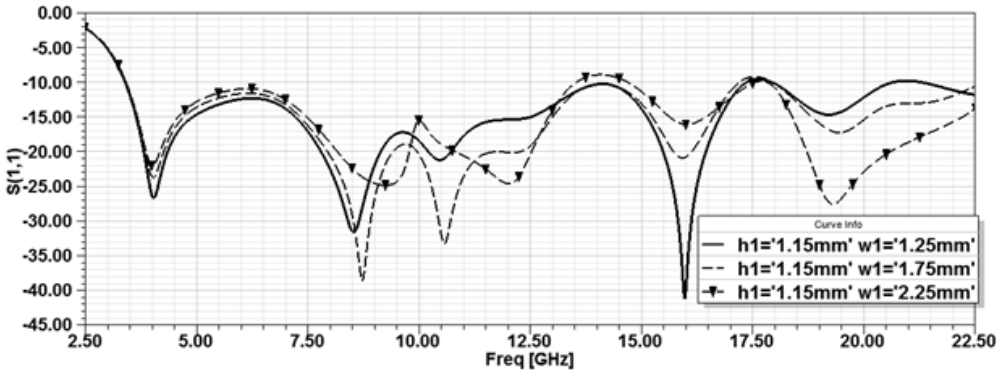


Figure 3. The scattering parameter S11 when varying the sleeves width (W1).

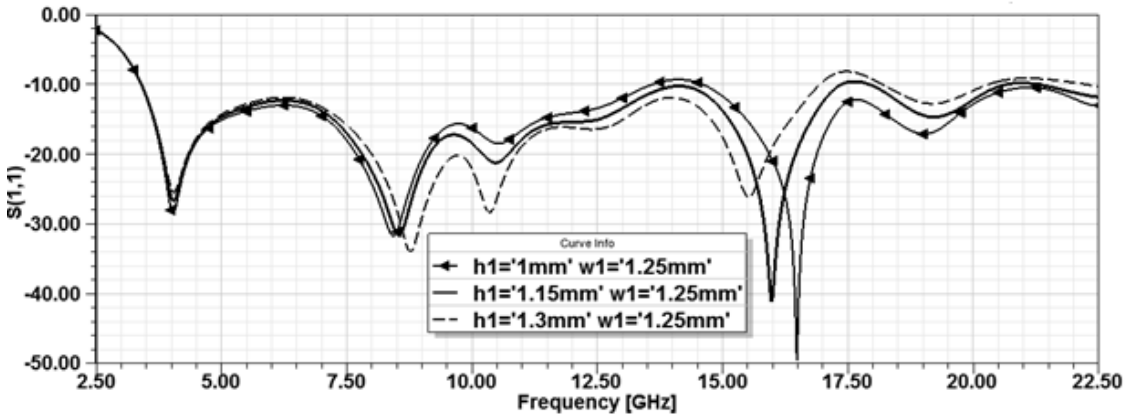


Figure 4. The scattering parameter S11 when varying the sleeves length (h1).

The sleeves shape variation is also studied using different shapes; rectangular, circular and triangular, as shown in Figure 6. Changing the sleeves shape has a high effect on the second resonance frequency values and the impedance matching over the frequency band of interest, such that the best result is obtained when using rectangular shape.

2) Inserting one rectangular groove in the middle of the ground plane. This groove is for adjusting the input impedance imaginary part to get nearly pure resistive impedance [16]. A parametric study is carried out on the groove length and width. The variation of groove length h2 and width W2 in the range

(1 – 2) mm and (0.5 – 1.5) mm, respectively, shows approximately minor effects on the impedance matching.

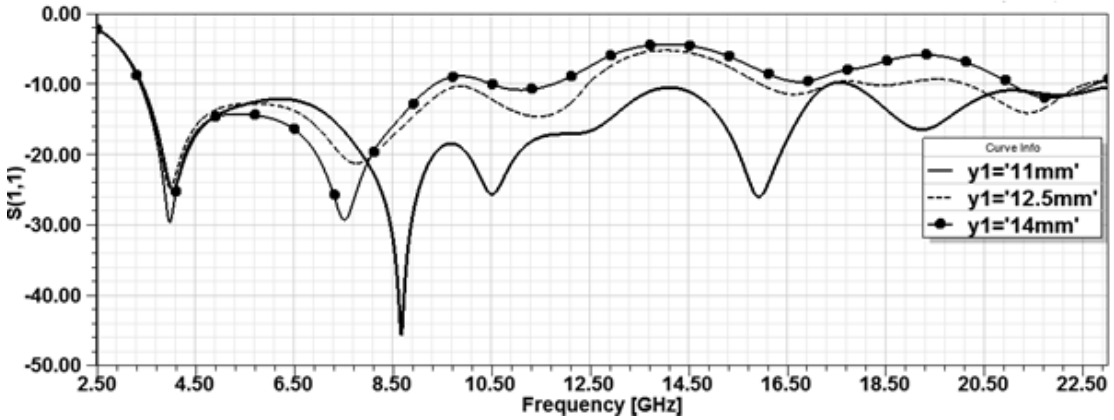


Figure 5. The scattering parameter S11 when varying the sleeves distance (y1).

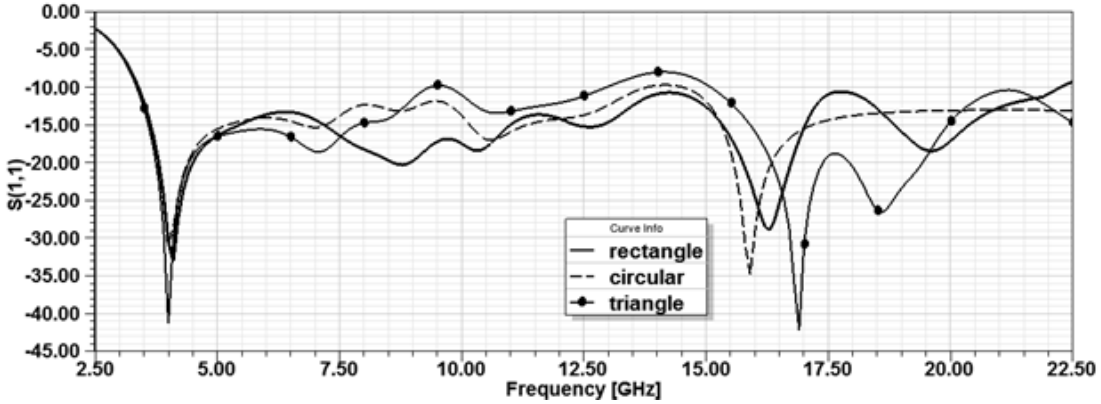


Figure 6. The scattering parameter S11 when varying the sleeves shape.

3) Adding two identical rectangular slots: this technique is used in order to get rid of or moderate the surface current reflection, thus adjusting the antenna impedance and reducing the return loss [17]. Parametric analysis is carried out on its length, width and location. The rectangular slot length h3 is varied between (5 – 7) mm as shown in Figure 7, where lengthening the slots enhances the impedance matching. The rectangular slot width W3 variation is also studied between (1.5 – 2.5) mm and no noticeable effect on the impedance matching is shown in Figure 7.

The effect of changing the rectangular slot location y2 from the antenna edge, (6 – 8) mm, is shown in Figure 8. Return loss curves show minor variations on the impedance matching for the frequency range of interest. Moving the slots either away or closer to the antenna edge makes the impedance matching worse at the lower frequency range and decreases the covered bandwidth in the high frequency range. Better results are achieved when y2 = 7 mm.

The proposed antenna is simulated using HFSS software tool. Comparison between the proposed and simple basic antennas is shown in Figure 9. The bandwidth at S11 ≤ - 10 dB starts at 3.4 GHz, while it ends at 22.4GHz, which indicates large useful bandwidth with percentage bandwidth of 147.28%.

The simulated peak gain within the operating band is shown in Figure 10, where it ranges between (2.5 – 6.5) dB within (3.4 – 22.4) GHz. The gain demonstrates moderate values in the low frequency range and increases in the high frequency range.

Radiation patterns for the E-plane and H-plane at different frequencies of 4, 6, 12 and 16 GHz are shown in Figure 11. The radiation pattern plots demonstrate that the antenna actually radiates over the frequency band of interest. The antenna exhibits a dipole-like shape in the E-plane in the low frequency range, while the number of lobes rises with the increase in frequency because of the existence of higher order

modes. The H-plane shows good omni-directional behavior in the low frequency range, while it becomes less omni-directional with an increase in frequency.

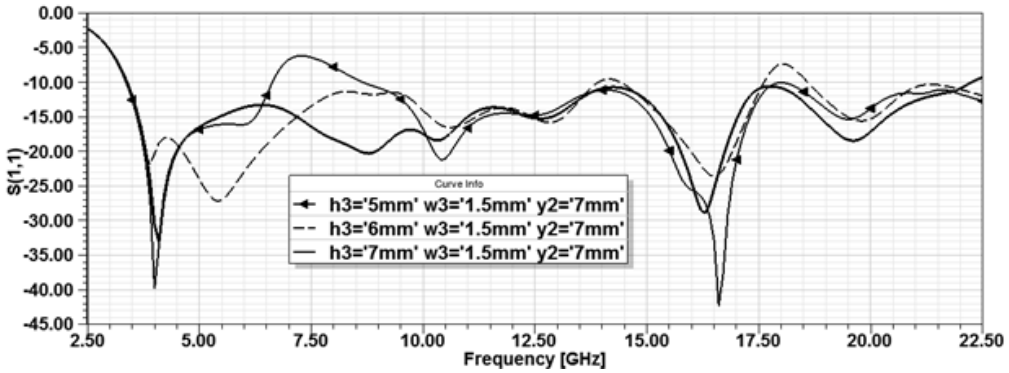


Figure 7. The scattering parameter S11 when varying the slot length (h3).

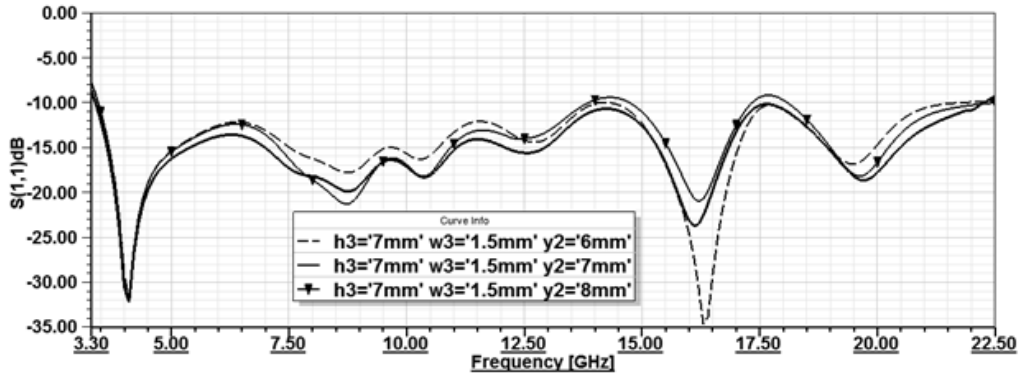


Figure 8. The scattering parameter S11 when varying the slot location (y2).

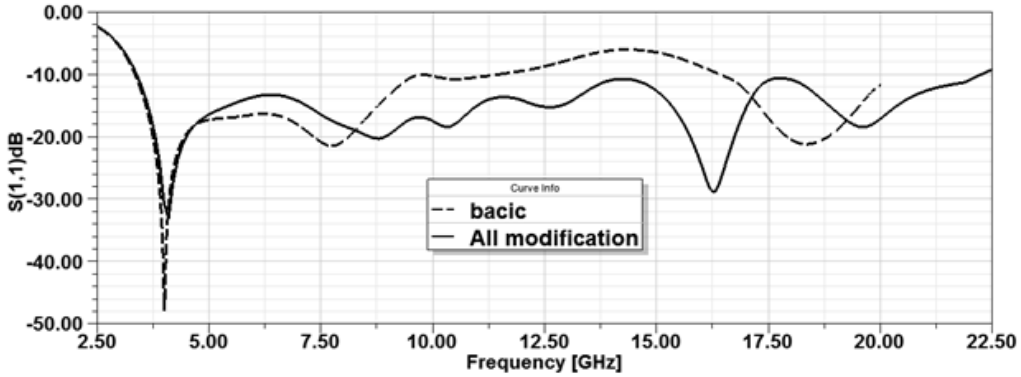


Figure 9. The scattering parameter S11 variation versus frequency.

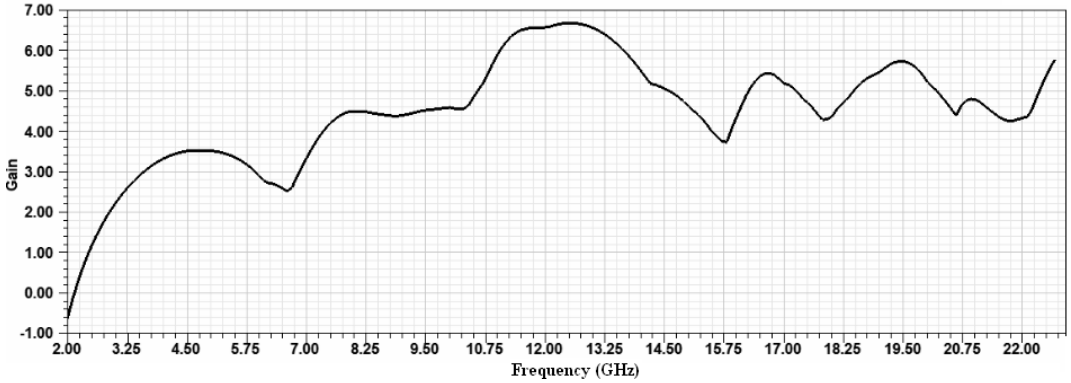


Figure 10. The simulated peak gain (dB).

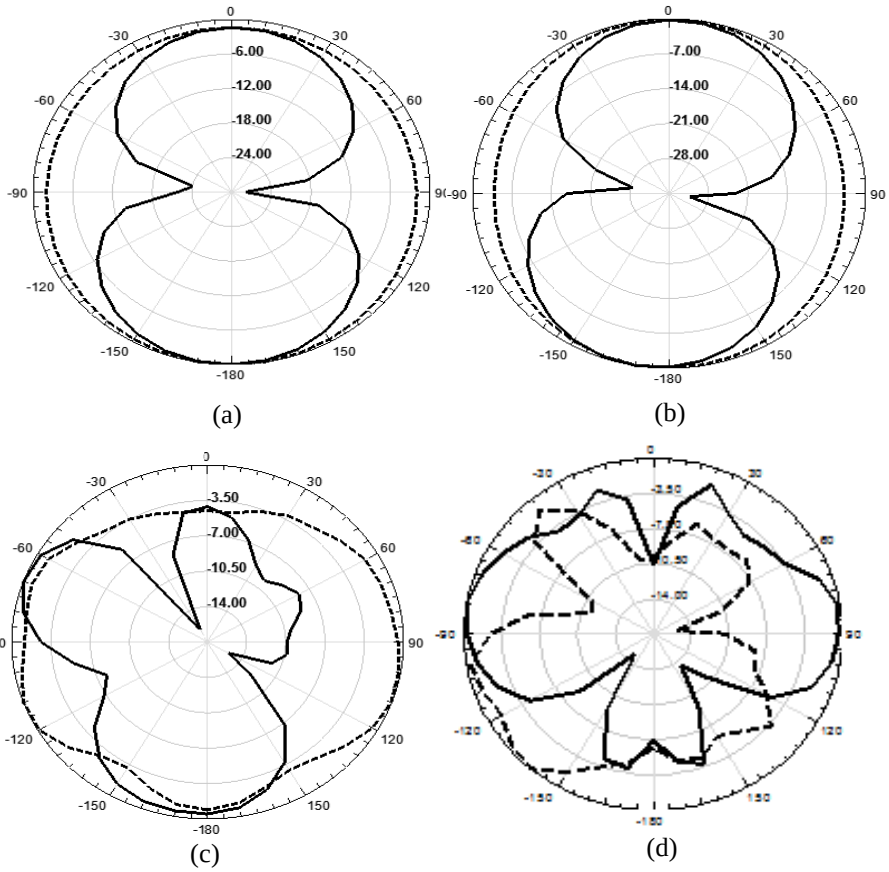


Figure 11. The radiation patterns at: (a) 4 GHz, (b) 6 GHz, (c) 12 GHz and (d) 16 GHz (_____E-plane and - - - H-plane).

The proposed antenna simulated radiation efficiency is plotted for the desired frequency range, as shown in Figure 12, which indicates good efficiency ranging between (74 - 94) %. The vector current density distributions at 3.8 GHz and 6.6 GHz is uniformly distributed and nearly flowing along the same direction for the three suggested ground modifications, as shown in Figure 13, which emphasizes the UWB characteristics.

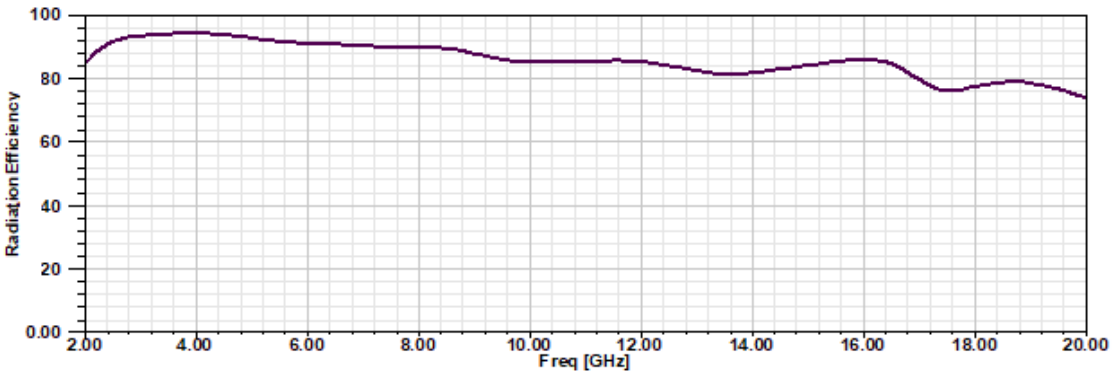
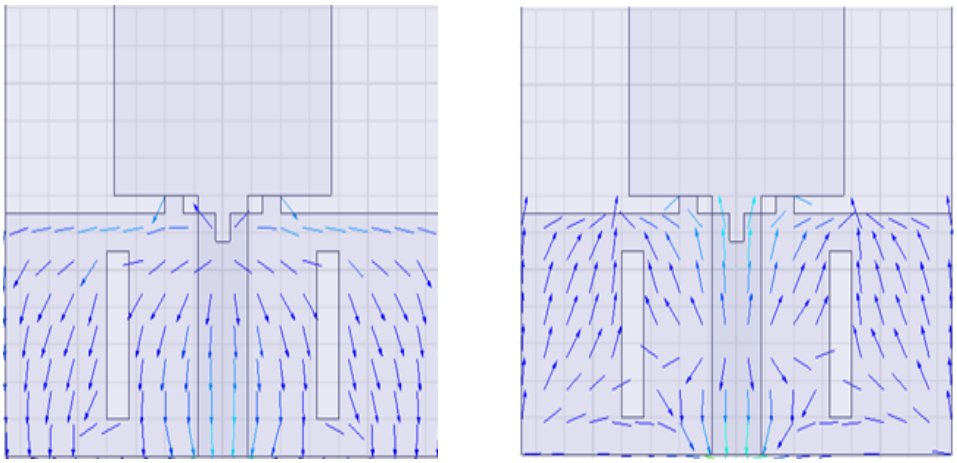


Figure 12. The radiation efficiency for the proposed antenna.

Comparison between the proposed antenna design and those presented in other research papers is shown in Table 1. Our proposed antenna is simpler in design than those antennas given in [2], [5] and [9], has larger bandwidth compared to the antennas presented in [2], [6] and [12] and is smaller in size compared to the antennas given in [5], [6] and [9].



(a) (b)

Figure 13. The vector current distribution at the ground plane at (a) 3.8 GHz and (b) 6.6 GHz.

Table 1. Comparison of the proposed antenna with those presented in other research papers.

Reference Antenna	Total Dimensions (L x W) in mm^2	Bandwidth in GHz	Percentage BW %
[2]	18 x 20	3.32 - 9.12	93.25%
[5]	30 x 50	2.75 - 20	151.6 %
[6]	30 x 51	2.96 - 11.89	120.27 %
[9]	60 x 60	1.79 - 28.02	175.98 %
[11]	30 x 40	2.87 - 50	178.29%
[12]	30 x 35	3.42 - 11.7	109.52 %
Proposed Antenna	34 x 36	3.4 - 22.5	147.49%

4. EXPERIMENTAL VERIFICATIONS

The proposed antenna is fabricated on FR4 substrate whose dielectric constant ε_r = 4.4 and thickness h = 1.6 mm as shown in Figure 14. This antenna is tested at the antenna measurement laboratory at King Abdullah Design and Development Bureau (KADDB). The scattering parameter (reflection coefficient) S11 (Return Loss (RL) = -S11) versus frequency is measured by using Agilent N5242A network analyzer with SAC-26G - 0.5 using 50 Ω cables.

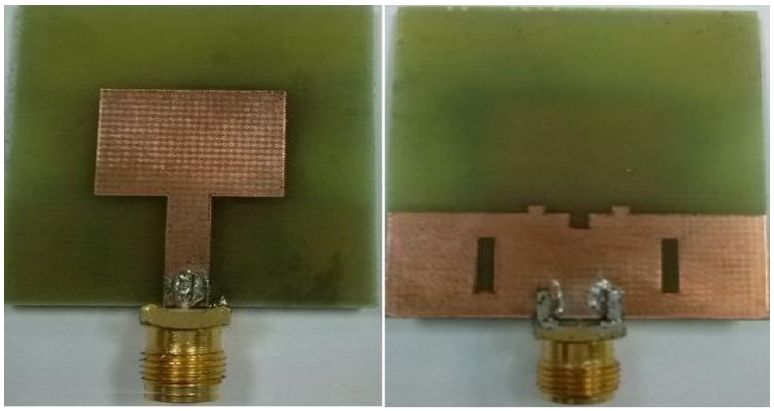


Figure 14. Photos of the fabricated antenna.

The measured S11 agrees with the simulated one in most of the desired frequency range. The measurements confirm the UWB characteristics as predicted in the simulation with a slight shift in the lower edge frequency; besides, they confirm the pass-band characteristics as shown in Figure 15.

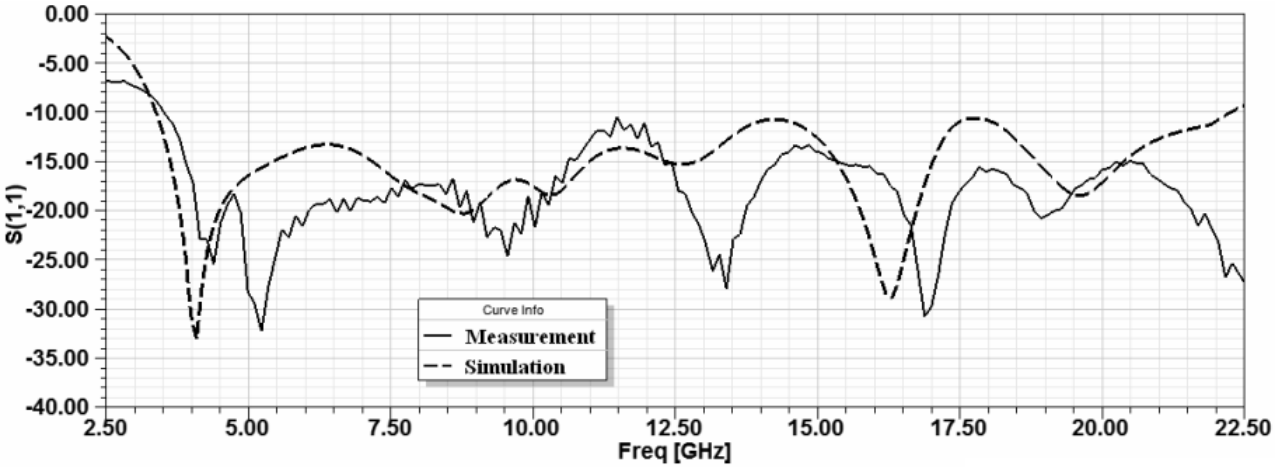


Figure 15. The simulated and measured S11 for proposed antenna.

The discrepancy between the measured and the simulated results are mostly attributed to the tolerance in fabrication and welding the SMA connector, which are not taken into account through simulation. Besides, the dielectric loss tangent of the FR4 substrate is kept constant during simulation, although it is actually a function of frequency.

5. CONCLUSION

UWB antenna is achieved by using rectangular patch antenna; 50 Ω microstrip feed line and partial ground plane with ground modifications; adding one rectangular groove, two rectangular sleeves and two rectangular slots. The simulated scattering parameter (reflection coefficient) S11 (Return Loss (RL) = - S11) versus frequency, at S11 ≤ - 10 dB using HFSS, shows high bandwidth (3.4 - 22.4) GHz, high gain and dipole-like radiation pattern in E-plane and good omni-directional radiation pattern in H-plane.

REFERENCES

[1] FCC, Revision of Part 15 of the Commission's Rules Regarding Ultra Wideband Transmission Systems, ET Docket 98 - 153, FCC 02 - 48, February 14, 2002.

[2] M. A. Ullah, F. B. Ashraf, T. Alam, M. S. Alam, S. Kibria and M. T. Islam, "A Compact Triangular-shaped Microstrip Patch Antenna with Triangular Slotted Ground for UWB Application," International Conference on Innovations on Science, Engineering and Technology (ICISET), Bangladesh, Oct. 2016.

[3] D. Pardhan, "Circular Patch with Circular Slit Patch Antenna Used for Ultra Wide Band Application," International Journal of Electrical, Electronics and Data Communication, vol. 5, pp. 84-87, 2017.

[4] N. Sameena, R. Konda and S. Mulgi, "A Novel Slot for Enhancing the Impedance Bandwidth and Gain of Rectangular Microstrip Antenna," Progress in Electromagnetics Research C, vol. 11, pp. 11–19, 2009.

[5] S. P. Kulkarni and V. G. Kasabegoudar, "Bandwidth Enhancement of Compact Circular Slot Antenna for UWB Applications," Global Journal of Researches in Engineering, vol. 17, no. 1, version 1, 2017.

[6] N. Prombutr, P. Kirawanich and P. Akkaraekthalin, "Bandwidth Enhancement of UWB Microstrip Antenna with Modified Ground Plane," International Journal of Microwave Science and Technology, vol. 2009, 2009.

[7] P. S. Ashtankar and C. G. Dethe, "Bandwidth Enhancement for UWB Antenna for HDR-WPAN Applications," International Journal of Advanced Innovations, Thoughts & Ideas, vol. 1, no. 1, 2012.

[8] Y. Rahayu, R. Ngah and T. Rahman, "Various Slotted UWB Antenna Design," 6th International Conference on Wireless and Mobile Communications, pp. 107–110, 2010.

[9] R. A. Santos and S. Jr. Cerqueira, "A Low-profile and Ultra-Wideband Printed Antenna with a 176% Bandwidth," Journal of Microwaves, Optoelectronics and Electromagnetic Applications, vol. 16, no.1, pp. 59-69, 2017.

[10] Q. Wu, R. Jin, J. Geng and M. Ding, "Printed Omni-directional UWB Monopole Antenna with Very Compact Size," IEEE Transactions on Antennas and Propagation, vol. 56, no. 3, pp. 896-899, Mar. 2008.

[11] N. Awad and M. Abdelazeez, "UWB Antenna with Super Bandwidth," APS/URSI International Symposium, 2016.

[12] N. Awad and M. Abdelazeez, "Multislotmicrostrip Antenna for Ultra-wideband Applications," Journal of King Saud University-Engineering Sciences, 2015.

[13] Rezaul Azim, Mohammad Targul Islam and Norbahiah Misran, "Design of a Planar UWB Antenna with a New Band Enhancement Technique," Applied Computational Electromagnetics Society Journal, vol. 26, no. 10, pp. 856-862, 2011.

[14] Rezaul Azim, Mohammad Targul Islam and Norbahiah Misran, "Printed Circular Planar Antenna for UWB Applications, " Telecommunication Systems, vol. 52, no. 2, pp. 1171-1177, 2013.

[15] C.–C. Lin, K.–Y. Kan and H.–R. Chauang, "A 3-8 GHz Broadband Planar Triangular Sleeves Monopole Antenna for UWB Communication," IEEE Antenna and Propagation Society International Symposium, USA, Dec. 2007.

[16] B. Ping Kasi and C. Chakrabrty, "A Compact Microstrip Antenna for Ultra Wideband Applications," European Journal of Scientific Research, vol. 67, no. 1, pp. 45–51.

[17] B. Gupta, S. Nakhate and M. Shandilya, "A Compact UWB Microstrip Antenna with Modified Ground Plane for Bandwidth Enhancement," International Journal of Computer Applications, vol. 49, no. 19, pp. 17-23, 2012.

ملخص البحث:

في هذه الورقة، يتم اقتراح هوائي ذي نطاق ترددي فائق العرض بنطاقٍ ترددي محسَّن. ويتكون الهوائي المحسَّن من رقعة مستطيلة تُغذّى بخط تغذية شريطي ميكرويّ ممانعته 50 أوم وسطح مستوٍ أرضي جزئي. ويتحقق التحسين في خصائص الهوائي المقترح عبر إجراء ثلاثة تعديلات على السطح المستوي الأرضي الجزئي، هي: إضافة كُمَّين مستطيلين، وإضافة أخدود مستطيل، وإضافة شَقّين مستطيلين.

تم استقصاء خصائص هذا الهوائي والتحقق منها عبر المحاكاة، وتبين أن الهوائي حقق نطاقاً ترددياً عريضاً جداً عند فقد رجوع يساوي أو يزيد على 10 ديسيبل؛ تراوح من 3.4 جيجاهيرتز إلى 22.4 جيجاهيرتز (147.29 %). كما تم الحصول على قيمة عظمى للكسب مع مواءمة جيدة للممانعة ونمط إشعاع في جميع الاتجاهات.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

AN EFFICIENT TWO-SERVER AUTHENTICATION AND KEY EXCHANGE PROTOCOL FOR ACCESSING SECURE CLOUD SERVICES

Durbadal Chattaraj[1], Monalisa Sarma[1] and Debasis Samanta[2]

(Received: 14-Dec.-2017, Revised: 13-Feb.-2018, Accepted: 28-Feb.-2018)

ABSTRACT

To avail cloud services; namely, Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), …etc. via insecure channel, it is necessary to establish a symmetric key between end user and remote Cloud Service Server (CSS). In such a provision, both the end parties demand proper auditing so that resources are legitimately used and privacies are maintained. To achieve this, there is a need for a robust authentication mechanism. Towards the solution, a number of single server authenticated key agreement protocols have been reported recently. However, they are vulnerable to many security threats, such as identity compromization, impersonation, man-in-the-middle, replay, byzantine, offline dictionary and privileged-insider attacks. In addition to this, most of the existing protocols adopt the single server-based authentication strategy, which are prone to single point of vulnerability and single point of failure issues. This work proposes an efficient password-based two-server authentication and key exchange protocol addressing the major limitations in the existing protocols. The formal verification of the proposed protocol using Automated Validation of Internet Security Protocols and Applications (AVISPA) proofs that it is provably secure. The informal security analysis substantiates that the proposed scheme has successfully addressed the existing issues. The performance study contemplates that the overhead of the protocol is reasonable and comparable with those of other schemes. The proposed protocol can be considered as a robust authentication protocol for a secure access to the cloud services.

KEYWORDS

Key agreement, Authentication protocol, User privacy, Cloud data security, Privacy-preserving protocol.

1. INTRODUCTION

With the exponential growth of Cloud service (e.g., SaaS, IaaS, PaaS) accessibility *via* Internet applications, it has been predicted that the annual global data traffic will reach 20.6 Zettabytes (ZB) per annum (approximately 1.7 ZB per month) by the end of 2021, which is approximately three times faster than 6.8 ZB per year (568 Exabytes) in 2016. As a result, the global data center IP traffic will grow 3-fold over the next 5 years. More precisely, data center IP traffic will grow at a Compound Annual Growth Rate (CAGR) of 25 percent from 2016 to 2021[1]. The report indicates a huge success of the cloud technology, but there is a challenge too[2]. In the cloud, a client remotely accesses his service provided by a service provider [52]. This leads to opening up a security problem as the communication takes place over insecure channel for accessing services. Towards this solution, a number of single server-based authentication and key agreement protocols are reported in the recent literature [48]-[51], [53]-[59].

[1] Cisco Global Cloud Index: Forecast and Methodology, 2016-2021 White Paper, [Online], available at: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html

[2] Parts of the paper published in ICTCS-2017 and SIN-2017.

This paper is an extended version of two conference papers. The first conference paper entitled "An Efficient Two-Server Authentication and Key Exchange Protocol", appeared in the proceedings of the International Conference "New Trends in Computing Sciences (ICTCS) 2017", 11-13 October 2017, pp. 127-132, Amman, Jordan. The second conference paper entitled "Privacy Preserving Two-Server Diffie-Hellman Key Exchange Protocol" was presented as a student paper at the 10th International Conference on "Security of Information and Networks (SIN) 2017", 13-15 October 2017, ACM SIGSEC, (DOI: 10.1145/3136825.3136871).

1. D. Chattaraj and M. Sarma are with Subir Chowdhury School of Quality and Reliability, Indian Institute of Technology Kharagpur, Kharagpur, India. Emails: dchattaraj@iitkgp.ac.in and monalisa@iitkgp.ac.in.
2. D. Samanta is with Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India. Email: dsamanta@sit.iitkgp.ernet.in.

The existing state-of-the-art authentication protocols [48]-[51], [53]-[59] provide security in the cloud environment and they follow a single server-based authentication mechanism. In order to achieve mutual authentication, all clients must interact with a particular authentication server. The single authentication server stores the credentials of all the clients in its database. Thus, this server is fully reachable for public access and is also vulnerable to a number of attacks, including dictionary, impersonation, password guessing, identity compromization and stolen-verifier attacks [52]. To mitigate these attacks, several schemes have been proposed, which are based on smart card, biometric, RFID (Radio-Frequency Identification) tag-based authentication approach. However, these approaches are economically expensive in terms of extra hardware cost [52].

In addition to this, to ensure a secure communication between the end user and the remote service server, several password-based two-server Diffie-Hellman key exchange protocols [6], [1]-[5], [7], [9], [11] are also reported in the literature. It may be noted that these protocols are not directly associated with the cloud computing domain, but these protocols can be incorporated in the same domain for accessing secure cloud services through web. Since its inception, in [6], both the intended sender and receiver establish a secret key between themselves based on previously shared common domain parameters (also called public parameters). However, this protocol does not have any mechanism to accomplish a mutual authentication between sender and receiver. As a result, an eavesdropper can easily make a man-in-the-middle attack by proxifying himself as a legitimate entity between both sender and receiver. Moreover, identity compromization attacks, impersonation attacks, replay attacks, privileged-insider attacks and offline dictionary attacks are the key security threats that are not properly addressed [10]. To avert these problems, several protocols are reported recently in the literature [24], [26]-[28], [45]-[47]. In these protocols, both parties should verify their identity before the shared secret symmetric key is settled between them. However, these existing state-of-the-art approaches do not properly address the server-side user privacy, protection from malicious insiders (i.e., byzantine attacks), single point of failure and single point of vulnerability issues [8]. In addition to this, most of the existing approaches utilize a single server-based authentication strategy, which is having the single point of failure and single point of vulnerability issues [52]. Note that in lieu of the existing extensive state-of-the-art single server-based authentication and key agreement protocol stack available for the cloud environment, in this paper, we focus only on the limitations and issues related to the password-based two-server Diffie-Hellman authenticated key agreement schemes, which could be incorporated for the same environment and finalize the following objectives.

1.1 Motivation and Research Objectives

To avert the aforesaid discussed issues and limitations of the existing password-based two-server authenticated key agreement protocols, we set the following objectives in the proposed scheme:

1. The proposed scheme should overcome several known attacks, such as password guessing, stolen-verifier, replay, man-in-the-middle, privileged-insider, impersonation, offline dictionary and identity compromization attacks.
2. The proposed scheme should provide a better server-side security and user privacy.
3. The user authentication process in the proposed scheme should be more robust and user-friendly.
4. The proposed scheme should distribute securely the session key between the entities.
5. The proposed scheme should mitigate the existing server-side single point of failure (SOF) and single point of vulnerability (SOV) issues.

The proposed approach *vis-a-vis* the above-mentioned objectives is as follows. An efficient password-based two-server authentication scheme has been proposed. This dual server model is planned in such a way that it is resilient against existing vulnerabilities; namely, man-in-the-middle, offline dictionary, byzantine, server identity compromization, client identity compromization, password guessing, stolen-verifier, privileged-insider and replay attacks. In this two-server model, the server, which is at the front-end, is responsible for interfacing with a client only, while the other server at the back-end accomplishes the authentication task. As the back-end server is hidden from public exposure, it ensures the server-side security by minimizing the risk of both SOF and SOV issues. As a solution to the impersonation

attacks, we propose a public key infrastructure (PKI)-enabled password-based two-server Diffie-Hellman authenticated key exchange scheme. To preserve user privacy and support anonymity, an Elliptic Curve Cryptography (ECC)-based digital signature generation and verification strategy has been adopted in the proposed scheme. The research contributions of the proposed approach are as follows.

1.2 Research Contributions

Following are the major research contributions as a realization of the proposed approach.

- We propose an approach to distribute the session keys without time synchronization.
- We introduce a new digital signature-based verification strategy by which each entity would be able to substantiate the other entity along with the issuer of the security credentials (i.e., session key) on which both the entities rely.
- To distribute the session key securely, we propose a pair-wise session key distribution approach using the concept of server-side in memory caching.
- The informal security analysis has been carried out to make evident that the proposed scheme can protect several known active as well as passive attacks.
- The formal security verification using the broadly-accepted AVISPA tool is carried out for the proposed scheme and the simulation results ensure that the proposed scheme is also secure.
- The suggested protocol alleviates the existing SOF and SOV issues by adopting a new dual server-based user authentication strategy.
- To ensure user friendliness of the proposed approach, a user only needs to enter his identity and password in the system for authenticated key agreement process.
- To preserve user privacy during authentication process, the user remains anonymous even if an adversary is eavesdropping the communication messages between user and remote cloud service server.

1.3 Organization of the Paper

The rest of the paper is organized as follows. Section 2 discusses the recent literature related to the work. The basic knowledge throughout the paper is presented in Section 3. Section 4 discusses the proposed protocol. The formal verification of the proposed protocol using AVISPA tool is described in Section 5. Section 6 presents the informal security analysis of the proposed protocol. The performance analyses of the proposed protocol are discussed in Section 7. Finally, Section 8 concludes the paper.

2. RELATED WORKS

In this section, we brief about the existing single server-based authenticated key exchange protocols, its issues and challenges.

Recently, two well-known Authenticated Key Agreement Protocol (AKAP) families are widely used in cloud industries. In the first category, several symmetric key-based single server AKAP are reported in the literature, such as, Kerberos, PKINIT, IDFusion, Sesame, …etc. [52]. Second category conveys the asymmetric key-based single server AKAP which are based on Diffie-Hellman key exchange protocol [25] and its variants reported in [24], [26]-[28]. Keeping the eye on the above fact, the literature review of our work is three-fold. First, we discuss the existing known security issues and challenges for the aforesaid two AKAP families and then elaborate the recent issues reported in the area of cloud computing platform by alleviating the state-of-the-art security threats of the existing schemes as follows:

2.1 Issues in Symmetric Key-based Single Server AKAP

In order to access cloud services (e. g. SaaS, PaaS, IaaS, …etc.) over the Internet, it is necessary for a user to enrol himself with the Cloud Service Provider (CSP). After enrolment, the end user can access cloud services remotely over the Web. Usually, according to the symmetric key based single server AKAP scheme, the CSP stores the secret information in the Key Distribution Centre (KDC), where a single point of compromization makes the whole system jeopardized and it is also vulnerable to

on/offline dictionary attacks. For example, existing approaches [29], [31]-[32], [51] enrol an end user by asking his "username" and password. This username is used as the primary credential, which is verified at the time of user authentication. In fact, selecting a "username" is not enough to be considered as a strong private entity. As a result, an adversary can easily incorporate different attacks, such as impersonation attacks and identity compromization attacks by sniffing the "username" from the insecure media. Moreover, the existing password-based enrolment strategy is vulnerable to password guessing (on/offline dictionary) attacks and stolen-verifier attacks. Additionally, the existing approaches [29]-[30] derive the client's secret key as the hash value of his password. Therefore, the key will remain the same until client changes the current password. However, changing this password needs updating in enrolled data maintained by the KDC and this, in fact, invites many key rollover problems [1]. According to the current practice, a user makes an authentication request to an authentication server (AS) by means of a plain text containing "username" [29]. In this context, an attacker can eavesdrop the "username" and later expose himself to the AS as a legitimate user. In other words, an attacker can easily determine from the transmitted message which users are currently online. In this situation, an attacker has scope to make man-in-the-middle attacks as well as replay attacks [44]. Further, an eavesdropper can make identity compromization attacks and impersonation attacks by stealing the "username" if the channel is insecure [43], [42]. Moreover, the AS issues an Authentication Ticket (AT) to an end user after verifying only "username" without verifying its password or other security credentials [43]. However, as "username" is not a confidential credential, there is an opportunity for an attacker to get multiple authentication tickets by simply sending a "username" to the AS. As a consequence, a cryptanalyst can decrypt the ciphertexts (i.e., ATs) using some knowledge about underlying user's password. Thus, this scheme is vulnerable to Ciphertext-only Attacks (COAs). In addition to this, the aforesaid discussed protocol is vulnerable to different other known security threats, including SOF and SOA issues reported in [52].

2.2 Issues in Asymmetric Key-based Single Server AKAP

Intuitively, according to the well-known Diffie-Hellman key exchange protocol [25], both the intended sender and receiver establish a secret key between themselves based on previously shared common domain parameters (also called public parameters). However, this protocol does not have any mechanism to accomplish a mutual authentication between sender and receiver. As a result, an eavesdropper can easily make a man-in-the-middle attacks by proxifying himself as a legitimate entity between both sender and receiver. In addition to this, identity compromization attacks, impersonation attacks and replay attacks are the key security threats that are not properly addressed. To avert these problems, several protocols are reported in the literature [24], [26]-[28], [45]-[47]. In these protocols, both parties should verify their identity before the shared secret symmetric key is settled between them. However, these existing state-of-the-art approaches do not properly address the server-side user privacy, protection from malicious insiders (i.e., byzantine attacks), single point of failure and single point of vulnerability issues [8].

2.3 Issues in Recent Single Server AKAP for Cloud Platform

Yang et al. [33] proposed an authentication scheme in a cloud environment setting. However, Chen et al. [34] pointed out the security pitfalls in Yang et al.'s scheme [33] that it is vulnerable to insider and impersonation attacks. To withstand these security loopholes in Yang et al.'s scheme, Chen et al. then designed a dynamic ID-based authentication scheme for cloud computing environment, which is based on the elliptic curve cryptography (ECC). Wang et al. [35] reviewed Chen et al.'s scheme [34] and proved that their scheme is vulnerable to offline password guessing as well as impersonation attacks. In addition, it was found that Chen et al.'s scheme does not provide user anonymity and also has a clock synchronization problem. Later, Hao et al. [36] presented a time-bound ticket-based mutual authentication scheme for cloud computing. The purpose of using the time bound tickets is to reduce the server's processing overhead. Unfortunately, Jaidhar [37] identified that Hao et al.'s scheme [36] is insecure against denial-of-service attack during the password change phase. Wazid et al. [38] also proposed a provably secure user authentication and key agreement scheme for cloud computing environment. Their scheme withstands the weaknesses of the existing schemes and also supports extra functionality features, such as user anonymity, efficient password and biometric update phase in multi-server environment. Recently, Gope and Das [39] proposed an anonymous mutual authentication

scheme for ubiquitous mobile cloud computing services, in which a legitimate mobile cloud user is allowed to enjoy n times all the ubiquitous services in a secure and efficient way, where the value of n may differ based on the principal he/she has paid for. In addition, Odelu et al. [40] reviewed Tsai-Lo's scheme [41] and pointed out that their scheme does not provide the session-key security and strong user credentials' privacy. To remove the security weaknesses found in Tsai-Lo's scheme, Odelu et al. designed a provably secure authentication scheme for distributed mobile cloud computing services. Since its inception, S. Kumari et al. [50] proposed a provably secure biometrics-based multi-cloud-server authentication scheme for accessing secure cloud services *via* insecure channel. In this scheme, authors have used a biometrics-based authentication scheme. M. H. Ibrahim et al. [51] proposed an attribute-based authentication protocol on the cloud for thin clients. In this scheme, authors have introduced two new authentication schemes for resource constraint client in cloud environment which support private attribute-based access to remote cloud servers. In this work, authors claimed that unlike existing attribute-based encryption and signature schemes, their scheme requires only a little amount of elliptic curve bilinear pairings and modular exponentiations. In 2015, Kalra and Sood [53] reported an authentication scheme to connect resource-constrained devices (tiny devices) to the cloud server using Elliptic Curve Cryptography (ECC) for Internet of Things (IoT). However, in 2017, S. Kumari et al. [48] have shown that Kalra and Sood's scheme [53] is vulnerable to offline password guessing and privileged insider attacks and does not achieve device anonymity, session key agreement and mutual authentication. To mitigate these vulnerabilities, S. Kumari et al. [48] proposed a secure authentication scheme based on ECC for IoT and cloud. In this scheme, authors have claimed that their scheme achieves all security requirements and is resistant to various known attacks as compared to the Kalra and Sood's scheme. In 2017, Wu et al. [49] proposed a lightweight and anonymous RFID tag authentication protocol with cloud assistance for e-healthcare application. Although these existing state-of-the-art approaches address the server-side user privacy, user anonymity, protection from malicious insiders and other known attacks, most of the schemes utilize extra hardware devices like biometric scanner, RFID-tag, smart card, …etc. and do not resolve the existing ciphertext-only attacks, single point of failure issues and single point of vulnerability issues.

Intuitively, to alleviate several known attacks and issues existing for the evolving cloud computing paradigm of the aforesaid extensive literature, in this work, we convey a new and efficient two server-based authenticated key agreement protocol. More specifically, in this paper, we address security threats, like identity compromization, server-side impersonation, offline dictionary, privileged-insider, man-in-the-middle, replay, byzantine, password guessing, stolen-verifier and COA attacks, as well as two important issues; namely, SOF and SOV issues, respectively. The basic mathematical knowledge required to understand the proposed protocol is discussed as follows.

3. PRELIMINARIES

In this section, we brief the basic security knowledge throughout the paper. In this regard, we discuss elliptic curve cryptography and its two underlying security assumptions; namely, Elliptic Curve Decisional Diffie-Hellman Problem (ECCDHP) and Elliptic Curve Discrete Logarithm Problem (ECDLP), as well as one-way cryptographic hash function as follows.

3.1 Elliptic Curve

Suppose $m, n \in Z_p$, where $Z_p = \{0,1, \cdots p-1\}$ and $p > 3$ is a prime. A non-singular elliptic curve $y^2 = x^3 + mx + n$ over the finite field Z_p is the set $E_p(m,n)$ of solutions $(x,y) \in Z_p \times Z_p$ to the congruence $y^2 \equiv x^3 + mx + n$ (mod p), where $m, n \in Z_p$ such that $4m^3 + 27n^2 \neq 0$ *(mod p)* and a point at infinity or zero point O. Note that $4m^3 + 27n^2 \neq 0$ *(mod p)* is a necessary and sufficient condition to ensure a non-singular solution for the equation $x^3 + mx + n = 0$ [18]. $4m^3 + 27n^2 = 0$ *(mod p)* implies that the elliptic curve is singular [14].

The algebraic formulae for the sum of two same or different points on $y^2 \equiv x^3 + mx + n$ *(mod p) are as follows:*

1) $R + O = O + R = R$ *for all* $R \in E_p(m,n)$.
2) *If* $R = (x, y) \in E_p(m,n)$, *then* $(x, y) + (x, -y) = O$. *(The point (x, -y) is denoted by -R).*

3) *Let $R \neq S$, where $R = (x_1, y_1) \in E_p(m,n)$ and $S = (x_2, y_2) \in E_p(m,n)$ with $x_1 \neq x2$, where $R \neq \pm S$. Then $R + S = (x_3, y_3)$, where $x_3 = \lambda^2 - x_1 - x_2 \ (mod\ p)$ and $y_3 = \lambda(x_1 - x_3) - y_1 \ (mod\ p)$. Here, $\lambda = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)$.*
4) *If $x_1 = x_2$, but $y_1 \neq y_2$, then $R_1 + R_2 = O$.*
5) *Let $R = (x_1, y_1) \in E_p(m,n)$ with $y_1 \neq 0$, where $R \neq -R$. Then $2 \cdot R = (x_3, y_3)$, where $x_3 = \lambda^2 - 2 \cdot x_1 \ (mod\ p)$ and $y_3 = \lambda \cdot (x_1 - x_3) - y_1 \ (mod\ p)$. Here, $\lambda = \left(\frac{3 \cdot x_1^2 + m}{2 \cdot y_1}\right)$.*
6) *Let $R = (x_1, y_1) \in E_p(m,n)$ with $y_1 = 0$. Then, $R_1 + R_2 = O$.*

Hasse's theorem states that the number of points on $E_p(m,n)$, denoted as $\#E$, satisfies the following inequality [19]:

$$p + 1 - 2\sqrt{p} \leq p + 1 + 2\sqrt{p} \ .$$

In other words, there are about p points on an elliptic curve $E_p(m,n)$ over Z_p. Also, $E_p(m,n)$ forms a commutative or an abelian group under addition modulo p operation.

Definition 1 (Elliptic Curve Discrete Logarithm Problem). *Given an elliptic curve $E_p(m,n)$ and two points $R, S \in E_p(m,n)$, find an integer x such that $S = x \cdot R$.*

Definition 2 (Elliptic Curve Decisional Diffie-Hellman Problem). *Given a point R on an elliptic curve $\in E_p(m,n)$ and two other points $x \cdot R, y \cdot R \in E_p(m,n)$, find $x \cdot y \cdot R$.*

3.2 One-way Hash Function

A one-way hash function $h: \{0,1\}^* \rightarrow \{0,1\}^l$ takes a binary string of variable length input, say $x \in \{0,1\}^*$ and outputs a binary string $h(x) \in \{0,1\}^l$ as an output of fixed length, say l bits. The formal definition of $h(\cdot)$ is provided as follows [16].

Definition 3 (Collision-resistant one-way hash function). *If an adversary A's advantage in finding collision in hash outputs with the execution time t is denoted by $Adv_A^{HASH}(t)$, it is defined by $Adv_A^{HASH}(t) = \Pr[(x,y) \leftarrow_R A : x \neq y \text{ and } h(x) = h(y)]$, where $\Pr[E]$ is the probability of an event E and $(x,y) \leftarrow_R A$ means that the pair (x, y) is randomly chosen by A. By an (η, t)- adversary A attacking the collision resistance of h(.), it indicates that the execution time of A is at most t and that $Adv_A^{HASH}(t) \leq \eta$.*

Examples of a one-way hash function include the Secure Hash Standard (SHA-1) hashing algorithm and the SHA-256 hashing algorithm [17].

4. PROPOSED PROTOCOL

The system architecture of our proposed protocol is shown in Figure 1. In order to demonstrate our proposed protocol, we break it into four parts: (1) System model. It tells about the system architecture, different entities involvement and individual knowledge about the initial security parameters. (2) Adversary model. It conveys the capabilities of an external entity or a malicious insider to make attacks into the security system. (3) Registration. It discusses the enrolment strategy of different entities into a trusted third-party server. (4) Authenticated key exchange. It conveys the shared session key establishment process between two distinct entities through message passing.

In our discussion, we frequently refer to some notations and symbols. All notations and symbols with their annotations are listed in Table 1.

4.1 System Model

Three types of entities are involved into our system; namely, client (C), service server (SS), back-end server (BS), where SS is the public server in two-server model and BS is the private server. The public server is reachable to everyone, whereas the private server works in the background and is controlled internally by the administrator only. C and SS both enrol themselves with BS during registration phase, but the authenticated key exchange task is carried out by both SS and BS, respectively.

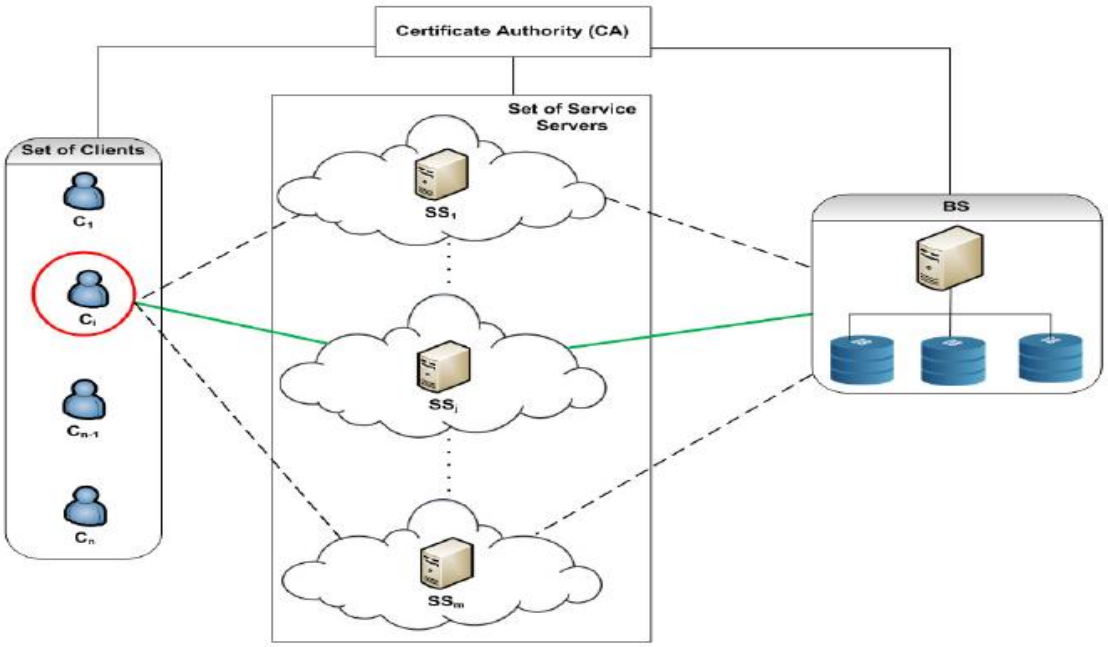


Figure 1. System architecture of the proposed protocol.

Further, there exists a certificate authority (CA), which issues a certificate C_P for each principal P. The CA selects a generator G on the elliptic curve $E_p(m, n)$ of order k and also selects two hash functions $H_1(\cdot)$ and $H_2(\cdot)$. Further, the principal P can randomly select $s_P \in \mathbb{Z}_k^*$ as secret key and computes the corresponding public key as $Q_P = s_P \cdot G$. Suppose that there exists an application B_{app} provided by BS with which the principal P can transform its secret information (i.e., password and server secret identity) offline (i.e., through a smart phone application). P registers these security credentials (after transformation) into BS *via* out-of-band channel [8]. It may be noted that the enrolments of SS are carried out by individual server administrator. At the time of authenticated key exchange, C or SS verifies the legitimacy of both SS and BS or both C and BS, respectively, without disclosing its secret credentials over the insecure channel. In addition to this, an application A_i is running into client workstation in order to access the service servers.

4.2 Adversary Model

Our adversary model is as follows: C and SS are controlled by an active adversary and BS is controlled by a passive adversary in terms of different attacks, such as offline dictionary, replay, man-in-the-middle, byzantine, identity compromization, impersonation and privileged-insider attacks. The active adversary can behave arbitrarily in order to make the above discussed attacks feasible. In contrast, we also consider the outside intruder as an active adversary. In [8], a passive adversary pursues an honest-but-curious activity; that is, it honestly executes the protocol according to the specification and does not modify any data in server's secure database. But, this adversary listens the communication channels to derive the security credentials between C and SS. Further, the passive adversary also knows all the shared secret keys and the internal state of the server.

Here, BS is trusted for both C and SS. We follow the well-accepted Dolev-Yao threat model (DY model) [20] in the proposed scheme. Under the DY model, any two parties in the network communicate over an insecure (public) channel, in which the end-point communicating parties, such as C and SS, are not considered as trustworthy entities. Therefore, under the DY model, an adversary (passive as well as active) A can then eavesdrop, modify or delete the exchanged messages during communication. Before going to the detail phases of the proposed protocol, we present a brief summary of it in Figure 2.

4.3 User and Service Server Registration Phase

Initially, both C and SS need to register themselves with BS *via* out-of-band channel [8] or postal network. Suppose that C_i wants to enrol his secret credentials with BS. Then, the detail enrolment step

is as follows:

Step 1: C_i transforms his user identity and password offline using B_{app} as $T_{C_{ID}} = H_b(C_{ID}^i)$ and $T_{C_{pwd}} = H_{b_1}(\text{PWD}||r_1)$. We may note that B_{app} has two hash functions as $H_b(.)$ and $H_{b_1}(.)$. For example, suppose that C_i executes offline B_{app} application in his smart phone and transforms his user identity using $H_{b_1}(.)$. Then, C_i chooses one random phone number from its contact information and transforms his password using $H_{b_1}(.)$. C_i needs to keep both PWD and r_1 secret.

Table 1. Notations and their descriptions.

Notations	Descriptions	Notations	Descriptions
A_i	An application running on user C_i's workstation to access service server	SS_{ID}^j	Public identity of SS_j
BS	Back End Authentication Server	SI_{ss}^j	Secret identity of SS_j
BS_{ID}	BS's public identity	SS_{sec}^{id}	Transformed identity of SS_j
B_{app}	An application computes $H_b(X_P)$ offline	s_{ss}	SS_j 's private key
C_i	i^{th} Client	s_c	C_i's private key
CA	Certificate Authority	s_b	BS's private key
C_{ID}^i	Public identity of C_i	U	A set of all points on the elliptic curve $E_p(m,n)$
C_P	Certificate for a principal P issued by CA	H_1	One-way cryptographic hash function as $\{0,1\}^* \rightarrow U$ [13]
$E_p(m,n)$	Elliptic curve $y^2 \equiv x^3 + mx + n$ (mod p) on the field Z_p	H_2	One-way cryptographic hash function as $U \rightarrow \{0,1\}^*$ [13]
ECC	Elliptic curve cryptography	H_b	One-way hash function as h(·)
ECDLP	Elliptic curve discrete logarithm problem	H_{b_1}	One-way hash function as h(J || δ)
ECDDHP	Elliptic curve decisional Diffie-Hellman problem	$T_{C_{ID}}$	C_i's transformed password
F_k	A field containing the elements 0, 1, ⋯, $(k-1)$	$T_{C_{pwd}}$	C_i's transformed user identity
G	A generator point $\in E_p(m,n)$ of some order (say, k)	X_P	Security credential X of a principal P
$h(\cdot)$	One way hash function, such as SHA-1, SHA-2 …etc.	$\mathbb{Z}_k$	A set containing the elements $0, 1, \ldots, (k-1)$
k	A 160 to 256-bit prime	(x/y)	P computes x number of exponentiations in real time and y number of exponentiations in offline mode
L	The size of a group element [1]	|q|	Bit length of q
l	The size of the hash value [1]	J	A secret information
PWD	Password of C_i	δ	Any random secret
P	A principal such as C_i, SS_j , BS	Q_{ss}	SS_j 's public key
Q_b	BS's public key	SS	Service Server
Q_c	C_i's public key	SS_j	j^{th} SS

Step 2: C_i sends C_{ID}^i, $T_{C_{ID}}$ and $T_{C_{pwd}}$ to BS via postal network.

Step 3: BS administrator creates C_i's account and updates user table with C_{ID}^i, $T_{C_{ID}}$ and $T_{C_{pwd}}$.

Step 4: BS broadcasts $T_{C_{ID}}$ to all SS's.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

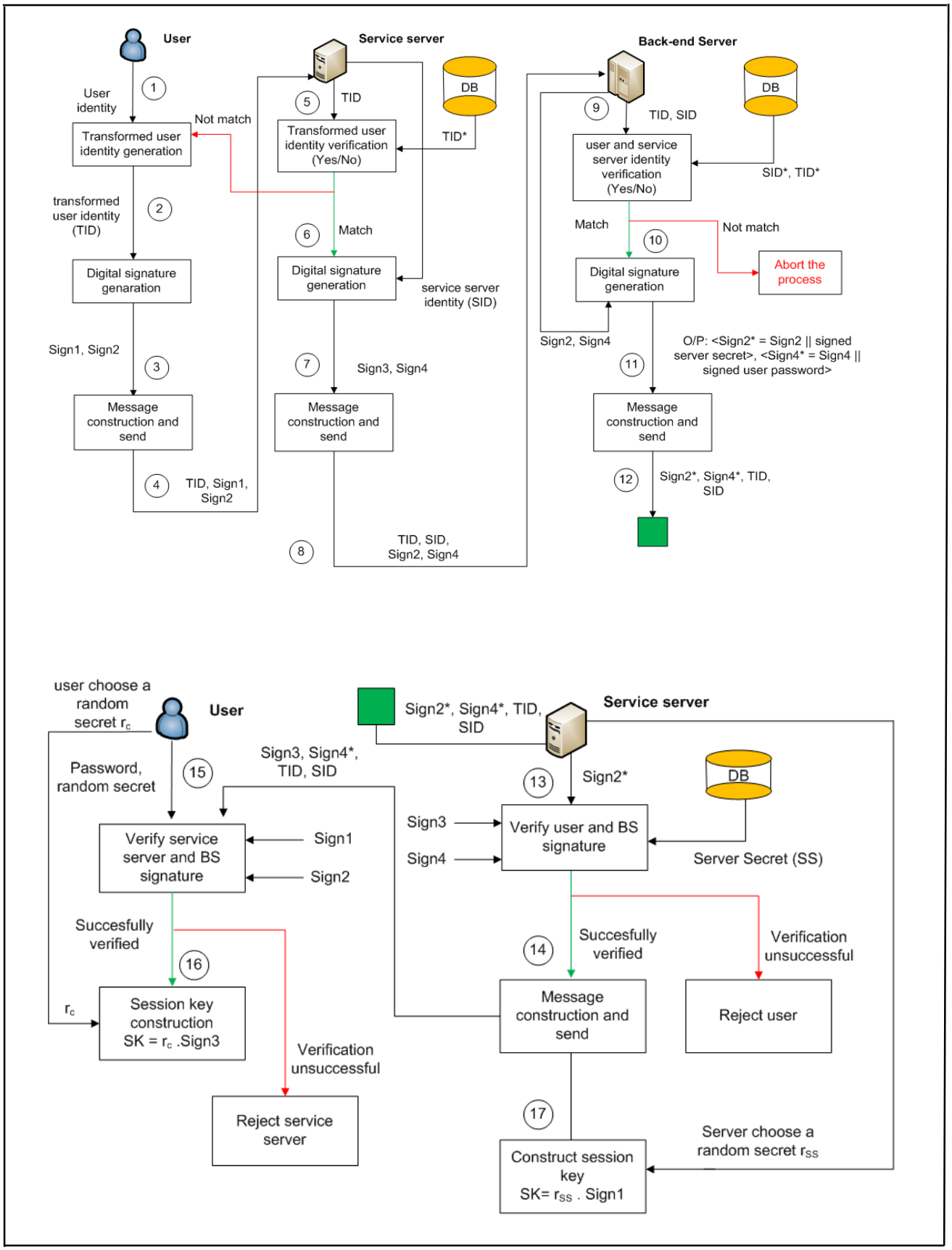


Figure 2. Summary of the proposed protocol.

Similarly, SS_j's administrator deploys server id (*i.e.*, SS_{ID}^j) and transformed secret identity (*i.e.*, $H_b(SI_{SS}^j)$) to BS. BS updates the service server table in its database. It may be noted that, the secret

identity (i.e., SI_{SS}^{j}) has been assigned to each SS by its respective administrator. This completes both C_i and SS_j enrolment process. After completion of registration, C_i can access services from SS_j followed by an authenticated key exchange task as follows:

4.4 Authenticated Key Agreement Phase

This sub-section presents the proposed password-based two-server authentication and key exchange protocol. All the message communications of the proposed protocol are shown in Fig. 3. The detail step in this process is as follows:

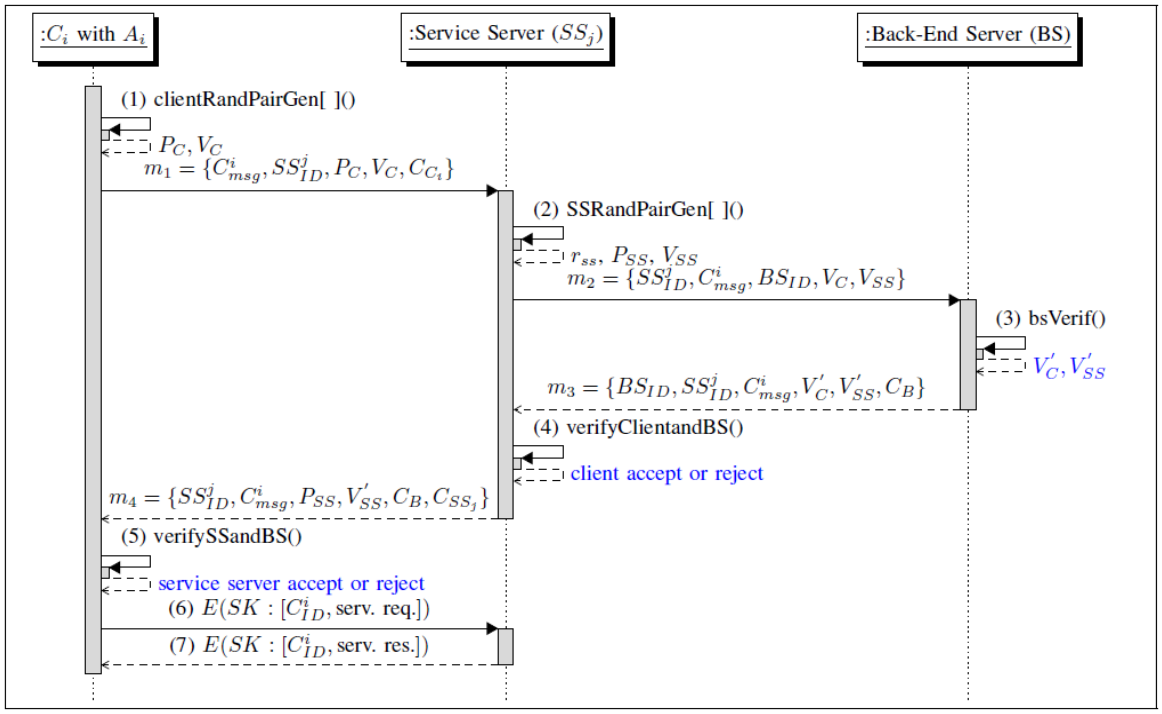


Figure 3. Summary of mutual authentication and key exchange process.

Step 1: C_i enters his identity and service server identity in application A_i.

Step 2: A_i computes C_{msg}^i and a signature pair (i.e., P_C and V_C) using clientRandPairGen[ ]() (see Fig. 4).

Step 3: A_i constructs a message m_1 = { $C_{msg}^i, SS_{ID}^j, P_C, V_C, C_{C_i}$ } and sends the same to SS_j .

Step 4: After receiving m_1, SS_j checks the transformed user identity (i.e., $C_{msg}^i = T_{C_{ID}}$) from its database.

Step 5: If user identity exists, then SS_j computes a signature pair (i.e., {P_{SS} , V_{SS}}) by taking the input of service server identity as SS_{ID}^j using SSRandPairGen[ ]() (refer to Figure 5). Step 6 is repeated, else the $C_i's$ request is rejected.

Step 6: SS_j constructs a message m_2 = { $SS_{ID}^j, C_{msg}^i, BS_{ID}, V_C, V_{SS}, C_{C_i}$ } and sends the same to BS.

Step 7: BS checks both previously stored transformed user identity and server identity from its database.

Step 8: If both the transformed identities exist, then BS modifies both random values V_C and V_{SS} as V_C' and V_{SS}', respectively, using bsVerif() (see Figure 6) and go to Step 9, else it rejects SS_j's request.

Step 9: After verifying both C_i and SS_j , BS constructs a message as m_3= { $BS_{ID}, C_{msg}^i, SS_{ID}^j, V_C', V_{SS}', C_B$ } and sends the same to SS_j.

Step 10: After getting reply from BS, SS_j verifies the legitimacy of both C_i and BS using verifyClientandBS() (refer to Figure 8).

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

function clientRandPairGen[ ](C^i_{ID}, s_c)
{
Input: client identity and client private key
Output: two nonnegative random numbers as P_C and V_C
Data: C^i_{ID}, C_{C_i}, PWD
Compute $C^i_{msg} := H_b(C^i_{ID})$
Choose a pseudo-random number $r_c \in \mathbb{Z}^*_k$
Compute $P_C := r_c \cdot G$
Compute $V_C := r_c \cdot H_2(H_1(C^i_{msg})) + s_c \cdot H_2(P_C) \bmod k$
return P_C, V_C
}

Figure 4. Client-side signature generation process.

function SSRandPairGen[ ](SS^j_{ID}, s_{ss})
{
Input: service server identity and service server private key
Output: two nonnegative random numbers P_{SS} and V_{SS}
Data: $C^{id}_{reg} := H_b(C^i_{ID})$, $SS^{id}_{sec} := H_b(SI^j_{SS})$, SS^j_{ID}, C_{SS_j}
if($C^i_{msg} = C^{id}_{reg}$)
{
Choose a random number $r_{ss} \in \mathbb{Z}^*_k$
Compute $P_{SS} := r_{ss} \cdot G$
Compute $V_{SS} := r_{ss} \cdot H_2(H_1(SS^j_{ID})$
$+s_{ss} \cdot H_2(P_{SS}) \bmod k$
return r_{ss}, P_{SS}, V_{SS}
}
else
{
return false // reject C_i //
}
}

Figure 5. Service server-side signature generation process.

Step 11: If both of them are verified successfully, then SS_j constructs session key as $SK = r_{ss} \cdot r_c \cdot G$ and a message m_4 = { $C^i_{msg}, SS^j_{ID}, P_{SS}, V'_{SS}, C_B, C_{SS_j}$ } and goto Step 12, else it rejects C_i's request.

function bsVerif$(SS^j_{ID}, C^{id}_{reg}, V_C, V_{SS})$
{
Input: service server identity, client transform identity and a pair of random numbers
Output: two nonnegative random numbers as V'_C and V'_{SS}
Data: $C_B, C^{id'}_{reg} := H_b(C^i_{ID}), SS^j_{rgid} := SS^j_{ID}, C^{pwd}_{reg} := H_{b_1}(PWD||r_1), SS^{id}_{sec} := H_b(SI^j_{SS})$
if($C^{id'}_{reg} = C^{id}_{reg}$ & $SS^j_{ID} = SS^j_{rgid}$)
{
Compute $V'_C := V_C + s_b \cdot H_2(H_1(SS^{id}_{sec}))$
Compute $V'_{SS} := V_{SS} + s_b \cdot H_2(H_1(C^{pwd}_{reg}))$
return V'_C, V'_{SS}
}
else
{
break // unauthorized access//
}
}

Figure 6. Client and Service server verification at BS.

"An Efficient Two-Server Authentication and Key Exchange Protocol for Accessing Secure Cloud Services" , D. Chattaraj, M. Sarma and D. Samanta.

function verifySSandBS $(SS_{ID}^{j}, \mathcal{P}_{\mathcal{SS}}, \mathcal{V}'_{\mathcal{SS}}, C_B, C_{SS_j})$
{
Input: service server identity, a pair of random numbers, BS certificate, SS certificate
Output: return true or false
Data: r_c, $C_{reg}^{pwd} := H_{b_1}(PWD||r_1)$
Compute $temp_2 := V'_{SS} \cdot G$
Compute $temp_3 := P_{SS} \cdot H_2(H_1(SS_{ID}^{j})) + Q_{ss} \cdot H_2(P_{SS}) + Q_b \cdot H_2(H_1(C_{reg}^{pwd}))$
if($temp_2 = temp_3$)
{
Compute $SK := r_c \cdot P_{SS} := r_c \cdot r_{ss} \cdot G$
return true // accept SS_j and BS //
}
else
{
return false //reject SS_j and BS//
}
}

Figure 7. Client-side verification and session key generation process.

Step 12: SS_j sends the message m_4 to C_i .

Step 13: After receiving m_4 , C_i checks the legitimacy of both SS_j and BS using verifySSandBS() (see Figure 7).

Step 14: If both of them are verified successfully, then C_i constructs session key as $SK = r_c \cdot r_{ss} \cdot G$.

Finally, C_i and SS_j establish a secure channel between themselves and set up a shared secret symmetric key as $SK = r_c \cdot r_{ss} \cdot G = r_{ss} \cdot r_c \cdot G$.

function verifyClientandBS$(C_{msg}^{i}, P_C, V'_C, C_{C_i}, C_B)$
{
Input: client transformed identity, a pair of random numbers, client certificate and BS certificate
Output: return true or false
Data: $C_{reg}^{id} := H_b(C_{ID}^{i}), SS_{sec}^{id} := H_b(SI_{SS}^{j}), SS_{ID}^{j}, C_{SS_j}$
Compute $temp := V'_C \cdot G$
Compute $temp_1 := V_C \cdot G + Q_b \cdot H_2(H_1(SS_{sec}^{id}))$
if($temp = temp_1$)
{
Compute $SK := r_{ss} \cdot P_C := r_{ss} \cdot r_c \cdot G$
return true //accept C_i and BS //
}
else
{
return false // reject C_i and BS //
}
}

Figure 8. Service server-side verification and session key generation process.

4.4.1 Proof of Correctness

In order to verify the legitimacy of both sender and receiver along with back-end third party server, each sender or receiver must adhere to the following two cases:

Case 1. *For the purpose of verifying C_i and BS, SS_j needs to check*

$$V'_C \cdot G = P_C \cdot H_2(H_1(C_{msg}^{i})) + Q_c \cdot H_2(P_C) + Q_b \cdot H_2\left(H_1\left(SS_{sec}^{id}\right)\right). \quad (1)$$

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

To make the aforesaid check work, the following condition must hold:

$$V_C = r_c \cdot H_2(H_1(C_{msg}^i)) + s_c \cdot H_2(r_c \cdot G) \text{ (mod k).} \quad (2)$$

Case 2. *For the purpose of verifying SS_j and BS, C_i needs to check*

$$V'_{SS} \cdot G = P_{SS} \cdot H_2(H_1(SS_{ID}^j)) + Q_{ss} \cdot H_2(P_{SS}) + Q_b \cdot H_2\left(H_1\left(C_{reg}^{pwd}\right)\right). \quad (3)$$

To make the aforesaid check work, the following condition must be satisfied:

$$V_{SS} = r_{ss} \cdot H_2(H_1(SS_{ID}^j)) + s_{ss} \cdot H_2(r_{ss} \cdot G) \text{ (mod k)} \quad (4)$$

5. FORMAL VERIFICATION IN AVISPA

In order to check the validity of a security protocol, several formal verification tools and techniques have emerged. In order to analyse our proposed protocol, we utilize AVISPA (Automated Validation of Internet Security Protocols and Applications) tool [21]. This tool is providing a High Level Protocol Specification Language (HLPSL) [22] [23], which is the *de facto* language recommended for AVISPA, to codify a security protocol. After codification, AVISPA internally converts the protocol into an Intermediate Format (IF), in order to assess its formal specification. In addition to this, AVISPA integrates four different model checkers which is actually perform the end-to-end analysis of the protocol. Moreover, the sending/receiving channels utilized for communication among several parties are modelled using the standard Dolev-Yao mechanism [20] to assess perfect secrecy.

According to the specification of HLPSL, each principal is mapped into a basic role, where all the local and global variables are specified. In addition to this, each role needs to declare an initial state and the state transitions. These state transitions represent different interactions; such as send or receive message exchanges among other roles. We code our authentication message exchanges in HLPSL. Moreover, we have specified three different roles: a client (c), a service server (ss) and a back-end authentication server (b). However, by taking an instance of a basic role, we have initiated a composite role called "session" to configure a single protocol run.

In order to specify such security objectives in HLPSL, AVISPA equips some standard commands, for example, "secrecy_of" (used to define the secrecy of keys), "authentication_on" (used to specify strong authentication on nonce), "weak_authentication_on" (used to represent weak authentication on timestamp) …etc. We model the security goals for the proposed protocol as follows:

1) *Origin authentication or man-in-the-middle attacks prevention:* It is done by incorporating the "authentication_on" to every security goal in every data exchange.
2) *Replay attack prevention:* It is achieved by establishing strong origin authentication.
3) *Data confidentiality:* It is ensured by interpreting "secrecy_of" to the secret keys (i.e., PWD and SS_{sec}^{id}) used in the protocol.
4) *Data integrity:* It is assured by applying "secrecy_of" to the secret one time distinct values (i.e., $P_C, V_C, P_{SS}, V_{SS}, V'_C, V'_{SS}, r_c \text{ and } r_{ss}$) exchanged by the protocol.

In order to validate our proposed protocol, we have considered seven different cases to find out the attack traces in AVISPA. All the cases and session configurations are presented in Table 2 and illustrated as follows:

- **C1:** It shows a single protocol session implementation with all the legitimate agents (i.e., c, ss and b) involved into authentication.
- **C2, C3 and C4:** They consider a situation, where an adversary (i) is trying to impersonate the client (c), the service principal (ss) and the back end authentication server.
- **C5:** It defines a situation, in which two parallel sessions are being executed and client (c) is interpreting the role of the service server (ss).
- **C6:** It represents a scenario, where two concurrent sessions are being executed simultaneously and client (c) is playing the role of back-end authentication server (b).
- **C7:** This case illustrates the effects of two parallel session executions in a single protocol

run, where an adversary (i) is trying to act same as the client (c).

Table 2. Test cases' execution in AVISPA.

Cases	**Different session configurations**
C1	session (c, ss, b, kc, ks, r1, srs, a, f, h1, h2)
C2	session (c, i, b, kc, ksi, ri, sri, a, f, h1, h2)
C3	session (c, ss, i, kci, ksi, ri, sri, a, f, h1, h2)
C4	session (i, ss,b, kci, srs, ri, sri, a, f, h1, h2)
C5	session (c, ss, b, kc, ks, r1, srs, a, f, h1, h2) session (c, c, b, kc, ksi, r1, sri, a, f, h1, h2)
C6	session (c, ss, b, kc, ks, r1, srs, a, f, h1, h2) session (c, ss, c, kc, ksi, r1, sri, a, f, h1, h2)
C7	session (c, ss, b, kc, ks, r1, srs, a, f, h1, h2) session (i, ss, b, kci, srs, ri, sri, a, f, h1, h2)

[Note: kc – client symmetric key (C_{reg}^{pwd}), ks – service server secret (SS_{sec}^{id}), kci and ksi – intruder knowledge about client and service server key, r1 and srs – shared secret information of client and service server with back-end server, a – hash function, f – hash function, h1 – hash function, h2 – hash function].

After execution of all the above cases in AVISPA, it did not reveal any attack traces. Thus, we can conclude that our proposed protocol is safe from man-in-the-middle attacks, impersonation attacks, parallel-session attacks and replay attacks.

Remark 1: *We consider four hash functions, such as 'a', 'f', 'h1' and 'h2', to codify the proposed protocol in AVISPA utilizing HLPSL. Here, 'a', 'f', 'h1' and 'h2' signify arithmetic addition, ECC point multiplication, encoding of a plaintext to an ECC point (H_1) and decoding of an ECC point to a random text (H_2), respectively. Note that, in HLPSL, there is no provision to code these aforesaid operations directly using mathematical operators. To specify these operations in HPLSL, hash functions are utilized. For example, in role specification and environment configuration, we firmly declare these operations as "a, f, h1, h2: hash_func".*

Remark 2: *In this paper, we use AVISPA tool (version 1.1) to prove that our proposed protocol is safe or not. Currently, the tool supports only four known attacks; namely, man-in-the-middle, replay, impersonation and parallel session attacks. In addition to this, as per the specification of the same tool, it is not possible to trace any arbitrary security attacks. Therefore, in AVISPA version -1.1, there is no scope to implement the recent attacks, like side channel attacks, covert channel attacks, privileged-insider attacks and identity compromization attacks existing in the cloud computing domain. We will finalize this issue in the near future utilizing a later version of the AVISPA tool.*

6. INFORMAL SECURITY ANALYSIS

An informal security analysis of the proposed protocol is carried out with respect to different attacks; namely, man-in-the-middle, replay, offline dictionary, privileged-insider, byzantine, impersonation, identity compromization attacks based on the aforesaid adversary model as follows:

Claim 1: *Resilient against man-in-the-middle attacks as an active adversary without control on servers.*

Proof: Suppose an adversary *Adv* traps the message $\langle C_{msg}^{i}, SS_{ID}^{j}, P_C, V_C, C_{C_i} \rangle$ sent by C_i to SS_j. It is noted that:

$$P_C = r_c \cdot G,\ r_c \in \mathbb{Z}_k^* \tag{5}$$

$$V_C = r_c \cdot H_2\left(H_1\left(C_{msg}^{i}\right)\right) + s_c \cdot H_2(P_C)(mod\ k). \tag{6}$$

It is hard to compute r_c in Eq. 5 (or s_c in Eq. 6) from P_C (or Q_C) due to ECDLP. Further, *Adv* cannot alter either P_C or V_C or both. Suppose that *Adv* alters V_C by V_C'' . Then, to satisfy the verification condition in Eq. 6, *Adv* has to alter from P_C to P_C' . But, due to ECDLP, it is hard to alter V_C by V_C'' and then satisfy

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

the condition in Eq. 6. Analogously, it is also hard to alter P_C by P_C' and then find V_C'' so that:

$$V_C'' \cdot G = P_C \cdot H_2\left(H_1\left(C_{msg}^i\right)\right) + Q_c \cdot H_2(P_C) + Q_b \cdot H_2(H_1(SS_{sec}^{id}))$$

holds. Similarly, it holds for the other messages, such as m_2, m_3 and m_4, respectively. Further, suppose that the adversary *Adv* wants to find the session key (i.e., SK) from either the value of $P_C = r_c \cdot \mathrm{G}$ or $P_{SS} = r_{ss} \cdot G$ or both by applying man-in-the-middle attacks and guessing the value of r_c and r_{ss}. Then, *Adv* would again end up with the computationally intractable or hard problem (i.e., ECCDHP assumption) to compute $SK = r_c \cdot P_{SS} = r_{ss} \cdot P_C = r_c \cdot r_{ss} \cdot G$. Hence, we can conclude that our scheme is resilient to man-in-the-middle attacks.

Claim 2: *Resilient against server-side identity compromization attacks as a passive adversary controlling SS_j.*

Proof: In the proposed scheme, SS_j is allowed to store only hashed client identities instead of original identities. Therefore, if a malicious insider gets this information, then he cannot trace out the original identity of the client. Suppose an adversary eavesdrops message $m_1 = \{ C_{msg}^i, SS_{ID}^j, P_C, V_C, C_{C_i} \}$ and tries to get the actual identity $\mathrm{C_{ID}^i}$ of $\mathrm{C_i}$ from $\mathrm{C_{msg}^i}$. But, it is computationally hard to extract $\mathrm{C_{ID}^i}$ from $\mathrm{C_{msg}^i}$, as $\mathrm{C_{ID}^i}$ has been transformed using one-way cryptographic hash function at the time of client enrolment phase.

Claim 3: *Resilient against replay attacks.*

Proof: For a particular session, C_i chooses a pseudo-random number r_c acting as a nonce or fresh value. We can easily infer from the four communication messages (i.e., m_1, m_2, m_3 and m_4) that the V_C value is exchanged among three parties and derived from r_c. As r_c is fresh for a single protocol run, thus we can say that V_C is also fresh *vis-a-vis* the messages are also fresh. This ensures our protocol to be free from replay attacks. More precisely, supposed that there is a protection against replay attack during the authentication phase, SS_j can store the values P_C and V_C in its cache memory temporarily. SS_j first checks if $P_C = P_C'$ and $V_C = V_C'$. If this is valid, the message is treated as replay message. Otherwise, SS_j updates P_C and V_C with P_C' and V_C' in its cache. In a similar way, it can address the replay attack issue during the mutual authentication phase between SS_j and BS.

Claim 4: *Resilient against single point of vulnerability and single point of failure.*

Proof: In the proposed dual server model, BS is not directly reachable to the client or an external adversary and it manages the crucial security credentials for both C_i and SS_j. Although SSs' are the front-end interface for malicious external adversary, it is not possible to gain any knowledge about the secret parameters (i.e., PWD, SI_{ss}^j, …etc.) except the hashed identity parameters of the client. This mitigates the single point of vulnerability issue. In addition to this, distribution of secret databases, periodical back-up and auditing of the security credentials from back-end server offline represent an easy task rather doing all these stuffs along with client or external intruder interfacing at front-end. Thus, our solution avoids the single point of failure issue without affecting the service availability of the system.

Claim 5: *Resilient against byzantine attacks.*

Proof: Intuitively, in order to design a full proof system to address byzantine attack, the proposed scheme is being planned in such a way that each entity would be able to substantiate the other entity along with the issuer of the security credentials. Suppose that SS_j wants to check the legitimacy of C_i. For this, SS_j needs to verify

$$V_C' \cdot G = r_C \cdot H_2\left(H_1\left(C_{msg}^i\right)\right) \cdot G + s_c \cdot H_2(P_C) \cdot G + s_b \cdot H_2\left(H_1\left(SS_{sec}^{id}\right)\right) \cdot G.$$

From this verification condition, we can easily infer that it associates with both C_i's and BS's private key, as well as C_i 's public identity. That means, implicitly, SS_j verifies both C_i and BS legitimacy using the public keys of themselves (see Eq. 1, Eq. 2, Eq. 3 and Eq. 4, respectively) before establishing the session key. Similarly, C_i checks the legitimacy of SS_j and BS before making the shared symmetric key with SS_j. Thus, it ensures trusted third party verification and is hence free from byzantine attacks.

Claim 6: *Resilient against offline dictionary attacks as a passive adversary controlling* SS_j .

Proof: Suppose that a malicious insider or a passive adversary *A* controls SS_j. As SS_j is not storing any dictionary of passwords for C_i in its secret database, where $i = 1, 2, \cdots, n$, then it is evident that there is no chance of offline dictionary attacks on C_i's password. Further, assume that *A* eavesdrops all four messages; namely, m_1, m_2, m_3 and m_4, respectively, from the protocol transcripts and tries to guess the C_i's original password. However, it is computationally hard to trace PWD without the knowledge of r_c , s_c *and* s_b, respectively.

Claim 7: *Resilient against offline dictionary attacks as a passive adversary controlling BS.*

Proof: Suppose a malicious insider or a passive adversary *A* controls BS. Then, in order to make offline dictionary attacks on C_i 's password infeasible, we encapsulate a random number (i.e., r_1) with C_i 's original password and make a transformed password as $T_{C_{pwd}} = H_{b_1}(\text{PWD}||r_1)$. Suppose that an insider or a passive adversary *A* compromises BS's secret database and tries to compute the password offline by dictionary attack. Then, he needs the knowledge about the random number r_1 . Further, suppose *A* to eavesdrop all the protocol transcripts and try to make an offline dictionary attack on the password. It is also hard to compute C_i's PWD without the knowledge of r_{ss}, s_{ss} and s_b, respectively.

Claim 8: *Resilient against privileged-insider attacks as a passive adversary controlling BS.*

Proof: During the client enrolment phase, B_{app} asks C_i to give his password. B_{app} transforms the password as $T_{C_{pwd}} = H_{b_1}(\text{PWD}||r_1)$ and C_i needs to send $T_{C_{pwd}}$ to BS *via* postal network. Now, suppose that a privileged-insider user of the BS, being an adversary *A*, knows the information $T_{C_{pwd}}$ by stealing the BS's database. Note that the masked password $H_{b_1}(\text{PWD}||r_1)$ contains the random secret r_1 , which is only known to C_i. Even if *A* guesses a password, he cannot verify it correctly without having r_1. Therefore, it is a computationally infeasible task for *A* to derive the password PWD of C_i . This shows that the privileged-insider attacks are protected in the proposed scheme.

Claim 9: *Resilient against ciphertext-only attacks.*

Proof: Suppose that for a particular session, an adversary *A* eavesdrops all the communication messages (i.e., m_1, m_2, m_3 and m_4) during authenticated key establishment phase. *A* repeats this process for multiple sessions for C_i whose public identity is $T_{C_{ID}} = H_b(C_{ID}^i)$. In this connection, *A* is having with himself a collection of messages. Form those messages, *A* chooses one ciphertext (i.e., V'_{SS}), which is associated with the C_i's password and constructs a set as S = $\{V'_{SS}\}$. After collecting such multiple sets as $S_i = \{V'_{SS_i}\}$, $\forall\, i = 1, 2, \cdots, n$, *A* tries to incorporate ciphertext-only attacks to find out the original plaintext (i.e., PWD) of C_i. However, to extract the corresponding plaintext or client's password from the set of ciphertexts, *A* needs the knowledge about r_1, r_c and s_c parameters. Without having these parameters, extraction of the original plaintext (i.e., PWD of C_i) is computationally intractable. Thus, the proof is completed and cipertext-only attacks are prevented

Claim 10: *Resilient against impersonation attacks.*

Proof: In order to mitigate impersonation attacks, the proposed scheme adopts a strong entity authentication principle. Here, SS_j verifies the legitimacy of client C_i using his transformed identity and his digital signature approved by BS and C_i verifies the legitimacy of SS_j utilizing its original identity and its digital signature approved by BS. This digital signature-based verification of each entity makes the protocol free from impersonation attacks by achieving strong authentication (see Cases C5 and C6 in Table 2 of Section 5).

Claim 11: *Resilient against stolen-verifier attacks.*

Proof: In order to mitigate the stolen-verifier attacks, the proposed scheme adopts a strong entity authentication principle. Here, SS_j verifies the legitimacy of client C_i using his transformed identity and his digital signature approved by BS and C_i verifies the legitimacy of service server SS_j utilizing its original identity and its digital signature approved by BS. This digital signature-based verification of each entity makes the protocol free from stolen-verifier attacks. Suppose an attacker steals the hashed verifier (C_{reg}^{pwd}) of C_i from BS and tries to login into the system. But, without knowing the PWD and r_1,

it is computationally hard to satisfy the verification condition specified in Eq. 3. Therefore, we can conclude that our scheme is resilient against stolen-verifier attacks.

Claim 12: *Supports forward secrecy.*

Proof: Forward secrecy tells about an analogy that an attacker cannot find the session keys constructed in past sessions even if he discovers the C_i's password PWD and SS_j's secret key SI_{ss}^{j} . In the proposed scheme, both C_i and SS_j compute an incomparable session key $SK = r_c \cdot P_{SS} = r_c \cdot r_{SS} \cdot G = r_{SS} \cdot P_C = r_{SS} \cdot r_c \cdot G = SK^*$ in every protocol run of the proposed scheme. The attacker cannot compute SK (or SK^*) from $P_{SS} = r_{SS} \cdot G$ and $P_C = r_c \cdot G$, even if he finds out C_i 's password PWD and SS_j's secret key SI_{ss}^{j} due to the hardness of large entropy ECC points. Thus, the proposed scheme provides forward secrecy.

7. PERFORMANCE ANALYSIS

This section deals with the performance evaluation of the proposed protocol. The performance analysis in this context is two-fold. Case 1: we compare our proposed protocol with the existing password-based two-server Diffie-Hellman authenticated key agreement protocols. Case 2: we substantiate the proposed methodology with the existing schemes available in cloud computing domain as follows:

(A) Case 1: Here, three major aspects for evaluation have been considered w.r.t. the existing password-based two-server Diffie-Hellman authenticated key agreement protocol as follows:

(1) Computational time (CT): In information theoretic sense, exponentiations govern individual entity's computational overhead [8]. In this synergy, we estimate the number of exponentiations as the evaluation of execution time and outline the performance results in Table 2. Consider C_i with A_i , for instance; it needs to compute a total of 4 exponentiations (see Eq. 1 and Eq. 2); that is, for the calculation of $P_C,\ V_C, Q_b \cdot\ H_2(H_1(SS_{sec}^{id}))$ and $V_C' \cdot\ G$ (refer to Section 3), and 2 exponentiations (i.e., $P_C,\ V_C$) out of them can be performed offline using [12]. We represent the value as "x/y" notation. More precisely, "4/2" signifies out of 4 exponentiations, client needs to perform 2 exponentiations in real time and other 2 exponentiations in offline mode.

(2) Communication overhead (CO): Since $|V_C|$ is equal to $|V_{SS}|$, we cannot distinguish between these two parameters. Thus, for ease of comparison, we fix it as $L = |V_C| = |V_{SS}|$. In addition to this, we have grossly ignored the bandwidth for both principal identities and the public certificates in this aspect of evaluation.

(3) Communication rounds (CR): A single round comprises a one-way transmission of messages.

Table 3 shows that our proposed protocol is quite efficient in terms of both communication and computation. Thus, we can apply our protocol for wireless applications also.

Moreover, the proposed protocol is having several security and functional features (SFFs). Table 4 shows a comparative analysis of the proposed protocol with other existing schemes in terms of the SFFs as follows:

(B) Case 2: In order to evaluate the performance of the proposed protocol w.r.t the existing schemes available in cloud computing domain, here, we consider three major aspects; namely, computation cost (see Table 5 and Table 6), communication overhead and storage cost (see also Figure 9).

Table 3. Comparison with other existing schemes in terms of performance.

Principles	Yang et al. [8]	JWX Protocol [15]	Yi et al. [1]	DHKEP [6]	Our Scheme
C_i	CT: (4/2)	CT: (6/0)	CT: (4/0)	CT: (2/0)	CT: (4/2)
	CO: 4L + 2l	CO: 6L +2l	CO: 3L + 4l	CO: 2L	CO: 2L
	CR: 4	CR: 3	CR: 3	CR:2	CR: 2
SS_j	CT: (4/1)	CT: (8/0)	CT: (5/0)	CT: (2/0)	CT: (4/2)
	CO: 8L +3l	CO: 11L + 3l	CO: 6L + 3l	CO: 2L	CO: 4L
	CR: 8	CR: 6	CR: 4	CR:2	CR: 4
BS	CT: (3/1)	CT: (4/0)	CT: (5/0)	--	CT: (2/0)
	CO: 4L + 1l	CO: 5L +1l	CO: 6L + 3l	--	CO: 2L
	CR: 4	CR: 3	CR: 4	--	CR: 2

In addition to this, we present several security and functional features (see Table 7) of the proposed protocol. Note that, in this study, we omit device energy consumption during the proposed protocol execution.

The calculation of computation, communication and storage cost respectively, of the proposed protocol is given as follows:

Table 4. Comparison between the proposed protocol and other existing schemes in terms of attacks and other features.

SFFs	**Yang et al. [8]**	**JWX Protocol [15]**	**Yi et al. [1]**	**DHKEP [6]**	**Our Scheme**
SFF1	No	No	No	No	Yes
SFF2	No	No	No	No	Yes
SFF3	Yes	Yes	Yes	No	Yes
SFF4	No	No	No	No	Yes
SFF5	No	No	No	No	Yes
SFF6	No	No	No	No	Yes
SFF7	No	No	No	No	Yes
SFF8	Yes	Yes	Yes	No	Yes
SFF9	No	No	No	No	Yes
SFF10	No	No	No	No	Yes
SFF11	Yes	Yes	Yes	Yes	Yes

[Note: *SFF1*: Resists man-in-the-middle attacks, *SFF2*: Resists replay attacks, *SFF3*: Resists offline dictionary attacks, *SFF4*: Resists identity compromization attacks, *SFF5*: Resists privileged-insider attacks, *SFF6*: Resists single point of failure issue, *SFF7*: Resists single point of vulnerability issue, *SFF8*: Resists impersonation attacks, *SFF9*: Resists byzantine attacks, *SFF10*: Resists ciphertext-only attacks, *SFF11*: Support mutual authentication and key establishment.**]**

Table 5. A comparative summary of existing schemes in cloud considering computational complexity.

Phases	**Principals**	**Karla and Sood [53]**	**S. Kumari et al. [48]**	**Odelu et al. [54]**	**S. Kumari et al. [50]**	**Shen et al. [55]**	**Our scheme**
URP	C_i	NCCI	1T_h	1T_h + 1T_f	2B_H	1T_h	2T_h
	BS/RA	No role	No role	3T_h	2T_h	2T_h+ 1T_m	NCCI
SRP	SS_j	6T_h+ 2T_m	5T_h+ 2T_m	NCCI	NCCI	NCCI	1T_h
	BS/RA	No role	No role	2T_h	1T_h	1T_h	NCCI
AKAP	C_i	4T_h+3T_m	3T_h + 4T_m	8T_h + 1T_f +1T_s+ 3T_m	2B_H + 5T_h + 3T_m	5T_h+3T_m	2T_m (OFL)+ 2T_m (ONL) + 3T_h
	SS_j	5T_h+4T_m	4T_h + 4T_m	6T_h + 2T_s + 2T_m	5T_h + 3T_m	4T_h+1T_m	2T_m (OFL) +2T_m (ONL) + 2T_h
	BS/RA	No role	No role	11T_h+ T_s + 1T_m	6T_h + 2T_m	8T_h+2T_m	4T_h + 2T_m

[**Note:** URP – User Registration Phase, SRP – Service server Registration Phase, AKAP – Authentication and Key Agreement Phase, T_h – Represents the CPU time to execute a one-way hash function, B_H – Represents the CPU time to execute a bio-hashing operation, T_m – Represents the CPU time to execute an elliptic curve scalar point multiplication, T_f – Represents the CPU time to execute a fuzzy extraction operation, T_s – Represents the CPU time to execute a symmetric key en(de)cryption, NCCI – No computational cost involved, OFL – Offline calculation of an elliptic curve scalar point multiplication, ONL – Online calculation of an elliptic curve scalar point multiplication.]

Table 6. A comparative summary of existing schemes in cloud considering computational complexity.

Phases	Principals	Yoon and Yoo [58]	Mishra et al. [57]	Wu et al. [56]	He and Wang [59]	Our scheme
URP	C_i	1T_h	1B_H + 3T_h	1T_ h + 1T_f	1T_h	2T_h
	BS/RA	1T_h	3T_h	N/A	2T_h	NCCI
SRP	SS_j	NCCI	NCCI	1T_h	NCCI	1T_h
	BS/RA	1T_h	2T_h	N/A	1T_h	NCCI
AKAP	C_i	5T_h + 2T_m	1B_H + 9T_h	7T_h + 1T_f + 2T_s + 1T_m	3T_m + 7T_h	2T_m (OFL) +2T_m (ONL) + 3T_h
	SS_j	4T_h + 2T_m	7T_h	6T_h + 2T_ s + 1T_m	3T_m + 5T_h	2T_m (OFL)+ 2T_m (ONL) + 2T_h
	BS/RA	7T_h	No role	N/A	2T_m + 9T_h	4T_h + 2T_m

[**Note:** URP – User Registration Phase, SRP – Service server Registration Phase, AKAP – Authentication and Key Agreement Phase, T_h – Represents the CPU time to execute a one-way hash function, B_H – Represents the CPU time to execute a bio-hashing operation, T_m – Represents the CPU time to execute an elliptic curve scalar point multiplication, T_f – Represents the CPU time to execute a fuzzy extraction operation, T_s – Represents the CPU time to execute a symmetric key en(de)cryption, NCCI – No computational cost involved, OFL – Offline calculation of an elliptic curve scalar point multiplication, ONL – Online calculation of an elliptic curve scalar point multiplication.]

In [48], CPU time required to calculate T_h and T_m is 2.3 and 22.26 microseconds. Therefore, in our scheme, the total number of computations required is 9T_h + 6T_m which is equal to 13376.7 microseconds. In addition to this, to calculate the storage cost, in our scheme, we observe that SS_j needs to store the hashed identity (160 bits) of the user and BS needs to keep both hashed identity (160 bits) and hashed password (160 bits) of the user and SS_j's identity (32 bits) and SS_j's secret hashed identity (160 bits), respectively. As a result, the total storage cost required is 672 bits. Further, we calculate the communication overhead in terms of bits for the authentication message exchanges (discussed in Section 4) among different entities as follows:

m_1 = 160 bits + 32 bits + 3 * 160 bits
m_2 = 32 bits + 160 bits + 32 bits + 2 * 160 bits
m_3 = 32 bits + 32 bits + 4 * 160 bits
m_4 = 32 bits + 5 * 160 bits.

Thus, the overall communication overhead by summing up all the messages is equal to 2752 bits. We compare our protocol overhead in terms of computation, storage and communication cost with other existing schemes as follows.

From Table 5 and Table 6, we can conclude that the proposed protocol is efficient in terms of computation cost. In addition to this, from Figure 9, we can easily substantiate the aforesaid claim. Although, it is lagging in terms of communication and storage cost, but still the proposed scheme supports several SFFs mentioned in Table 7.

8. CONCLUSIONS

This paper proposes a secure and efficient two-server authentication and key agreement protocol for accessing secure cloud services. For this purpose, we have presented a new password-based two-server Diffie-Hellman authenticated key exchange protocol, in which both the client and the service server can establish a secret symmetric key between themselves after their mutual authentication. In addition to this, the proposed protocol has an efficient verification capability, where during mutual authentication

Table 7. A comparative summary: security and functional features.

Security and functional features	**Karla and Sood [53]**	**S. Kumari et al. [48]**	**Yoon and Yoo [58]**	**Mishra et al. [57]**	**Wu et al. [56]**	**Shen et al. [55]**	**Odelu et al. [54]**	**S. Kumari et al. [50]**	**Our Scheme**
SFF1	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
SFF2	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes
SFF3	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
SFF4	Yes	Yes	No	No	No	Yes	No	Yes	Yes
SFF5	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SFF6	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
SFF7	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
SFF8	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
SFF9	No	NA	No	Yes	Yes	Yes	Yes	Yes	Yes
SFF10	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SFF11	No	No	Yes	Yes	No	Yes	Yes	Yes	No
SFF12	No	No	No	No	No	No	No	No	Yes
SFF13	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
SFF14	NA	NA	NA	NA	NA	NA	NA	NA	Yes
SFF15	NA	NA	NA	NA	NA	NA	NA	NA	Yes
SFF16	Yes	Yes	No	No	No	No	No	No	Yes
SFF17	No	Yes	Yes	Yes	Yes	Yes	Yes	NA	Yes
SFF18	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SFF19	No	NA	NA	NA	NA	NA	NA	Yes	Yes
SFF20	NA	NA	NA	NA	NA	NA	NA	NA	Yes

[**Note:** SFF1 – Resists privileged-insider attack, SFF2 – Provides user anonymity, SFF3 – Resists off-line password guessing attack, SFF4 – Resists user impersonation attack, SFF5 – Resists replay attack, SFF6 – Resists cloud-server impersonation attack, SFF7 – Provides mutual authentication, SFF8 – Provides forward secrecy, SFF9 – Resists known session-specific temporary information attack, SFF10 – Provides session key agreement and verification, SFF11 – Provides freely password changing facility, SFF12 – Resists byzantine attacks, SFF13 – Resists identity compromization attacks, SFF14 – Resists single point of failure issue, SFF15 – Resists single point of vulnerability issue, SFF16 – No extra hardware cost required, SFF17 – Resists stolen-verifier attacks, SFF18 – Resists man-in-the-middle attacks, SFF19 – Resists parallel session attacks, SFF20 – Resists ciphertext-only attacks, Yes – SFF achieved, No – SFF not achieved, NA – SFF not addressed in the particular scheme.]

phase both intended parties could verify their trusted third party (i.e., the issuer of security credentials). The security part of the proposed protocol has been thoroughly executed with the help of informal security analysis. In addition, the proposed scheme is also formally verified for security analysis using the broadly-used AVISPA tool. All the security analysis results show that the proposed protocol can protect various well-known attacks against a passive as well as active adversary. Also, it successfully alleviates several limitations in the existing schemes. The performance analyses also substantiate that our protocol is efficient in terms of both computation and communication overhead.

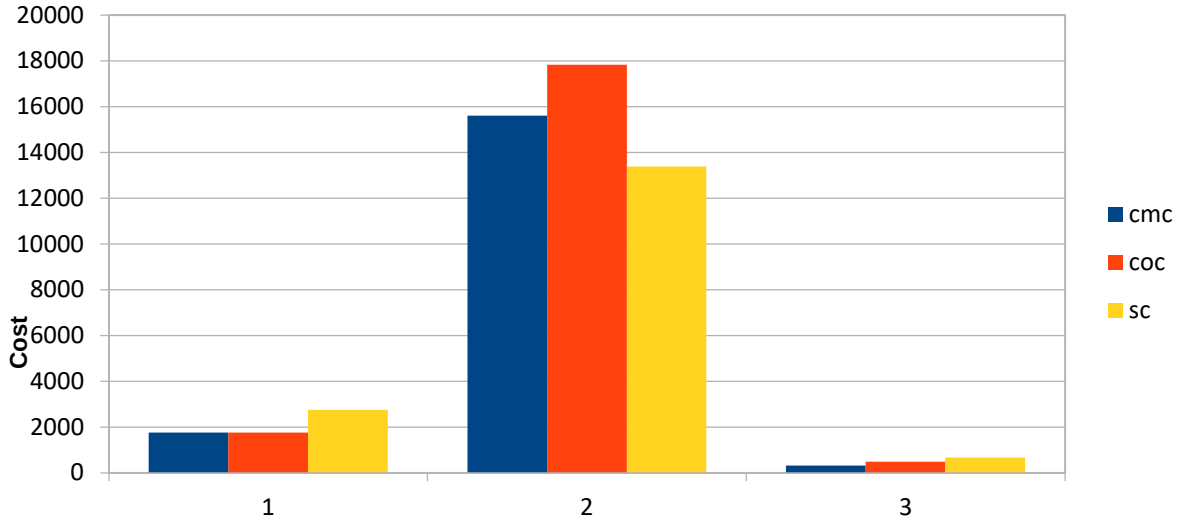


Figure 9. Performance comparison of the proposed protocol with the existing schemes.

[**Note:** Blue, orange and yellow colour plot represents Karla and Sood scheme [53], S. Kumari et al. scheme [48] and the proposed scheme, respectively. cmc – represent the communication cost in terms of bits. coc – represent the computation cost in terms of microseconds. sc – represent the storage cost in terms of bits. Here, 1, 2 and 3 represent the communication overhead, computation cost and storage cost, respectively.]

More significantly, a large-scale service server can be integrated with a single authorization centre hosting our protocol. In the future, we plan to evaluate the proposed protocol in a real-world environment setting. Note that, here, the real-world environment setting signifies an abstract network (wired/sensor/*ad-hoc* network) configuration, wherein different workstations are connected through a communication channel (wired/wireless). After implementing such an environment setting in any network simulator tool (NS2, NS3, …etc.), we can evaluate other performance metrics, such as end-to-end delay and throughput of the proposed scheme. This will allow us to fine-tune the protocol, if necessary, to offer better security and performance in a real-world deployment. The future direction also encompasses a formal treatment of our proposed scheme by utilizing Real-Or-Random model, Canetti-Krawczyk model or extended Canetti-Krawczyk model and analzing more complicated attacks, such as known key, unknown key-share, key-compromise impersonation, …etc.

ACKNOWLEDGEMENTS

The authors thank the anonymous reviewers and the Editor-in Chief for their valuable suggestions, which helped us improve significantly the presentation and quality of the paper.

REFERENCES

[1] X. Yi, S. Ling and H. Wang, "Efficient Two-server Password-only Authenticated Key Exchange," Transactions on Parallel and Distributed systems, IEEE, vol. 24, no. 9, pp. 1773-1782, 2013.

[2] X. Yi, F. Y. Rao, Z. Tari, F. Hao, E. Bertino, I. Khalil and A. Y. Zomaya, "ID2S Password-authenticated Key Exchange Protocols," Transactions on Computers, IEEE, vol. 65, no. 12, pp. 3687-3701, 2016.

[3] V. Boyko, P. MacKenzie and S. Patel, "Provably Secure Password-authenticated Key Exchange Using Diffie-Hellman," in Advances in Cryptology–Eurocrypt, Springer Berlin/Heidelberg, pp. 156-171, 2000.

[4] M. Abdalla and D. Pointcheval, "Simple Password-based Encrypted Key Exchange Protocols," in Cryptographers' Track at the RSA Conference, Springer, pp. 191-208, 2005.

[5] M. Bellare and P. Rogaway, "The AuthA Protocol for Password-based Authenticated Key Exchange," Technical Report, IEEE, vol. 1363, 2000.

[6] W. Diffie and M. Hellman, "New Directions in Cryptography," Transactions on Information Theory, IEEE, vol. 32, no. 2, pp. 644-654, 1976.

[7] X. Yi, F. Hao and E. Bertino, "ID-based Two-server Password Authenticated Key Exchange," in Proceedings of the European Symposium on Research in Computer Security, Springer, pp. 257-276, 2014.

[8] Y. Yang, R. H. Deng and F. Bao, "A Practical Password-based Two-server Authentication and Key Exchange System," Transaction on Dependable and Secure Computing, IEEE, vol. 3, no. 2, pp. 105-114, 2006.

[9] Y. Yang, R. H. Deng and F. Bao, "Fortifying Password Authentication in Integrated Healthcare Delivery Systems," in Proceedings of the ACM Symposium on Information, Computer and Communications Security, ACM, pp. 255-265, 2006.

[10] E. Bresson, O. Chevassut and D. Pointcheval, "Security Proofs for an Efficient Password-based Key Exchange," in Proceedings of the 10th ACM Conference on Computer and Communications Security, ACM, pp. 241-250, 2003.

[11] S. Bellovin and M. Merritt, "Encrypted Key Exchange: Password-based Protocol Secure against Dictionary Attack," in Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, pp. 72-84, 1992.

[12] R. P. Gallant, R. J. Lambert and S. A. Vanstone, "Faster Point Multiplication on Elliptic Curves with Efficient Endomorphisms," in Annual International Cryptology Conference, Springer, pp. 190-200, 2001.

[13] P. A. Fouque, A. Joux and M. Tibouchi, "Injective Encodings to Elliptic Curves," in Australasian Conference on Information Security and Privacy, Springer, pp. 203-218, 2013.

[14] NIST white paper, "Recommended Elliptic Curves for Federal Government Use," [Online], Available: http://csrc.nist.gov/groups/ST/toolkit/documents/dss/NISTReCur.pdf, 1999, Last accessed: 2nd April, 2017.

[15] H. Jin, D. S. Wong and Y. Xu, "An Efficient Password-only Two-server Authenticated Key Exchange System," Proceeding of 9th International Conference of Information and Communication Security, Springer, pp. 44-56, 2007.

[16] P. Sarkar, "A Simple and Generic Construction of Authenticated Encryption with Associated Data," ACM Transactions on Information and System Security, vol. 13, no. 4, pp. 33, 2010.

[17] Secure Hash Standard, FIPS PUB 180-1, National Institute of Standards and Technology (NIST), U.S. Department of Commerce, April 1995, [Online], Available: http://nvlpubs.nist.gov/nistpubs/FIPS/NIST. FIPS.180-4.pdf. Accessed in September 2017.

[18] R. W. D. Nickalls, "A New Approach to Solving the Cubic: Cardan's Solution Revealed," The Mathematical Gazette, vol. 77, no. 480, pp. 354-359, 1993.

[19] N. Koblitz, "Elliptic Curves Cryptosystems," Mathematics of Computation, vol. 48, pp. 203-209, 1987.

[20] D. Dolev and A. Yao, "On the Security of Public Key Protocols," Transactions on Information Theory, IEEE, vol. 29, no. 2, pp. 198-208, 1983.

[21] A. Armando, D. Basin, Y. Boichut et al., "The AVISPA Tool for the Automated Validation of Internet Security Protocols and Applications," International Conference on Computer-Aided Verification, Springer, pp. 281-285, 2005.

[22] L. Vigan`o, "Automated Security Protocol Analysis with the AVISPA Tool," Electronic Notes in Theoretical Computer Science, Elsevier, vol. 155, pp. 61-86, 2006.

[23] D. V. Oheimb, "The High-level Protocol Specification Language HLPSL Developed in the EU Project AVISPA," Proceedings of APPSEM Workshop, pp. 1-17, 2005.

[24] B. De Decker, "Unix Security and Kerberos," Computer Security and Industrial Cryptography, Springer, pp. 257-274, 1993.

[25] S. Blake-Wilson and A. Menezes, "Authenticated Diffie-Hellman Key Agreement Protocols, "International Workshop on Selected Areas in Cryptography, Springer, pp. 339-361, 1998.

[26] H. Krawczyk, "HMQV: A High-performance Secure Diffie-Hellman Protocol," Annual International Cryptology Conference, Springer, pp. 546-566, 2005.

[27] L. Harn, W. J. Hsin and M. Mehta, "Authenticated Diffie–Hellman key agreement protocol using a single cryptographic assumption," IEE Proceedings-Communications, vol. 152, no. 4, pp. 404-410, 2005.

[28] B. LaMacchia, K. Lauter and A. Mityagin, "Stronger Security of Authenticated Key Exchange," International Conference on Provable Security, Springer, pp. 1-16, 2007.

[29] C. Neuman, S. Hartman, T. Yu and K. Raeburn, "The Kerberos Network Authentication Service (V5)," RFC 4120, [Online], Available: https://tools.ietf.org/pdf/rfc4120.pdf, 2005.

[30] B. C. Neuman, T. Ts'o and Kerberos, "An Authentication Service for Computer Networks," IEEE Communications Magazine, vol. 32, no. 9, pp. 33-38, 1994.

[31] R. M. Needham and M. D. Schroeder, "Using Encryption for Authentication in Large Networks of Computers," Communications of the ACM, vol. 21, no. 12, pp. 993-999, 1978.

[32] L. O'gorman, A. Bagga and J. Bentley, "Query-directed passwords," Computers & Security, Elsevier, vol. 24, no. 7, pp. 546-560, 2005.

[33] J. H. Yang and P. Y. Lin, "An ID-based User Authentication Scheme for Cloud Computing," IEEE 10th Inter. Conf. on Intell. Inform. Hiding and Multimedia Signal Processing (IIH-MSP), pp. 98-101, 2014.

[34] T. H. Chen, H. L. Yeh and W. K. Shih, "An Advanced ECC Dynamic ID-based Remote Mutual Authentication Scheme for Cloud Computing," IEEE 5th International Conference on Multimedia and Ubiquitous Engineering (MUE), pp. 155-159, 2011.

[35] D. Wang, Y. Mei, C. G. Ma and Z. S. Cui, "Comments on an Advanced Dynamic ID-based Authentication Scheme for Cloud Computing," International Conference, WISM, vol. 7529, Lecture Notes in Computer Science, Springer, pp. 246-253, 2012.

[36] Z. Hao, S. Zhong and N. Yu, "A Time-bound Ticket-based Mutual Authentication Scheme for Cloud Computing," International Journal of Computers, Communications and Control, vol. 6, no. 2, pp. 227-235, 2011.

[37] C. D. Jaidhar, "Enhanced Mutual Authentication Scheme for Cloud Architecture," IEEE 3rd International Advance Computing Conference (IACC), pp. 70-75, 2013.

[38] M. Wazid, A. K. Das, S. Kumari, X. Li and F. Wu, "Provably Secure Biometric-based User Authentication and Key Agreement Scheme in Cloud Computing," Security and Communication Networks, vol. 9, no. 17, pp. 4103-4119, 2016.

[39] P. Gope and A. K. Das, "Robust Anonymous Mutual Authentication Scheme for n-times Ubiquitous Mobile Cloud Computing Services," Internet of Things Journal, IEEE, pp. 1-9, DOI: 10.1109/JIOT.2017.2723915, 2017.

[40] V. Odelu, A. K. Das, S. Kumari, X. Huang and M. Wazid, "Provably Secure Authenticated Key Agreement Scheme for Distributed Mobile Cloud Computing Services," Future Generation Computer Systems, vol. 68, pp. 74–88, 2017.

[41] J. L. Tsai and N. W. Lo, "A Privacy-Aware Authentication Scheme for Distributed Mobile Cloud Computing Services," IEEE Systems Journal, vol. 9, no. 3, pp. 805-815, 2015.

[42] I. E. Liao, C. C. Lee and M. S. Hwang, "A Password Authentication Scheme over Insecure Networks," Journal of Computer and System Sciences, vol. 72, no. 4, pp. 727-740, 2006.

[43] S. M. Bellovin and M. Merritt, "Limitations of the Kerberos Authentication System," ACM SIGCOMM Computer Communication Review, vol. 20, no. 5, pp. 119-132, 1990.

[44] G. S. Sadasivam, K. A. Kumari and S. Rubika, "A Novel Authentication Service for Hadoop in Cloud Environment," IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), pp. 1-6, 2012.

[45] R. Canetti and H. Krawczyk, "Analysis of Key-Exchange Protocols and Their Use for Building Secure Channels," International Conference on the Theory and Applications of Cryptographic Techniques, Springer, pp. 453-474, 2001.

[46] L. Law, A. Menezes, M. Qu, J. Solinas and S. Vanstone, "An Efficient Protocol for Authenticated Key Agreement," Designs, Codes and Cryptography, Springer, vol. 28, no. 2, pp. 119-134, 2003.

[47] S. K. Sood, A. K. Sarje and K. Singh, "A Secure Dynamic Identity-based Authentication Protocol for Multi-server Architecture," Journal of Network and Computer Applications, Elsevier, vol. 34, no. 2, pp. 609-618, 2011.

[48] S. Kumari, M. Karuppiah, A. K. Das, X. Li, F. Wu and N. Kumar, "A Secure Authentication Scheme Based on Elliptic Curve Cryptography for Iot and Cloud Servers," Journal of Supercomputing, Springer, DOI: 10.1007/s11227-017-2048-0, pp. 1-26, 2017.

[49] F. Wu, L. Xu, S. Kumari, X. Li, A. K. Das and J. Shen, "A Lightweight and Anonymous RFID Tag Authentication Protocol with Cloud Assistance for E-Healthcare Applications," Journal of Ambient Intelligence and Humanized Computing, Springer, DOI: 10.1007/s12652-017-0485-5, pp. 1-12, 2017.

[50] S. Kumari, X. Li, F. Wu, A. K. Das, K. R. Choo and J. Shen, "Design of a Provably Secure Biometrics-based Multi-cloud-server Authentication Scheme," Future Generation Computer Systems, Elsevier, vol. 68, pp. 320-330, 2017.

[51] M. H. Ibrahim, S. Kumari, A. K. Das and V. Odelu,"Attribute-based Authentication on the Cloud for Thin Clients," The J. of Supercomputing, Springer, DOI: 10.1007/s11227-016-1948-8, pp. 1-33, 2017.

[52] D. Chattaraj, M. Sarma and A. K. Das, "A New Two-server Authentication and Key Agreement Protocol for Accessing Secure Cloud Services," Computer Networks, Elsevier, DOI: 10.1016/j.comnet.2017.12.007, vol. 131, pp. 144-164, 2018.

[53] S. Kalra and S. K. Sood, "Secure Authentication Scheme for IoT and Cloud Servers," Pervasive and Mobile Computing, vol. 24, pp. 210-223, 2015.

[54] V. Odelu, A. K. Das and A. Goswami, "A Secure Biometrics-based Multi-Server Authentication Protocol Using Smart Cards," IEEE Trans. Inf. Forensics Secur., vol. 10, no. 9, pp. 1953-1966, 2015.

[55] H. Shen, C. Z. Gao, D. D. He and L. B. Wu, "New Biometrics-based Authentication Scheme for Multi-server Environment in Critical Systems," J. Ambient Intell. Hum. Comput., vol. 6, no. 6, pp. 825-834, 2015.

[56] F. Wu, L. Xu, S. Kumari and X. Li, "A Novel and Provably Secure Biometrics-based Three-factor Remote Authentication Scheme for Mobile Client-Server Networks," Comput. Electr. Eng., Elsevier, vol. 45, pp. 274-285, 2015.

[57] D. Mishra, A. K. Das and S. Mukhopadhyay, "A Secure User Anonymity Preserving Biometrics-based Multi-server Authenticated Key Agreement Scheme Using Smart Cards," Expert Syst. Appl., Elsevier, vol. 41, no. 18, pp. 8129-8143, 2014.

[58] E. Yoon and K. Yoo, "Robust Biometrics-based Multi-server Authentication with Key Agreement Scheme for Smart Cards on Elliptic Curve Cryptosystem," J. Supercomput, Springer, vol. 63, no. 1, pp. 235-255, 2013.

[59] D. He and D. Wang, "Robust Biometrics-based Authentication Scheme for Multi-server Environment," IEEE System Journal, vol. 9, no. 3, pp. 816-823, 2015.

ملخص البحث:

من أجل إتاحة الخدمات السحابية [البرمجيات؛ المنصّة؛ البنية التحتية]، من الضروري إنشاء مفتاح متناظر بين المستخدم النهائي والخادم البعيد الخاص بالخدمات السحابية. وعندئذٍ، يتعين على الطرفين النهائيين أن يكون بينهما تدقيق مُحكم. ولتحقيق ذلك، هناك حاجة إلى آلية تصديق متينة. حيث أن البروتوكولات الخاصة بالتصديق بواسطة خادم واحد تتسم بالهشاشة أمام العديد من التهديدات المرتبطة بالأمان.

تقترح هذه الدراسة بروتوكول تصديق وتبادل للمفاتيح يتميز بالفعالية، يرتكز على كلمة السر ويستخدم خادمين. والبروتوكول المقترح يعالج أوجُه القصور في البروتوكولات القائمة. وقد أثبت التحقق الرسمي من البروتوكول المقترح، باستخدام التحقق الآلي من بروتوكولات أمان الإنترنت وتطبيقاتها، أنّ البروتوكول المقترح في هذه الدراسة آمن. كما برهن تحليل الأمان أن البروتوكول المقترح عالج القضايا القائمة بنجاح. وتشير دراسة الأداء أن تكلفة البروتوكول معقولة مقارنة بالبروتوكولات الأخرى المماثلة. و يمكن اعتبار البروتوكول المقترح بروتوكول تصديقٍ متيناً من أجل الوصول إلى الخدمات السحابية بأمان.

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

UNMANNED GROUND VEHICLE WITH VIRTUAL REALITY VISION

Mahmood Al-Khalil, Rami Abu-Rhayyem, Ahmad Hammoudeh and Talal A. Edwan*

(Received: 21-Nov.-2017, Revised: 8-Feb.-2018, Accepted: 3-Mar.-2018)

ABSTRACT

This paper aims to describe the design and implementation of an Unmanned Ground Vehicle (UGV) and a smart phone virtual reality (VR) head mounted display (HMD) which enables visual situation awareness by giving the operator the feel of "head on rover" while sending the video feeds to separate operator computer for object detection and 3-D model creation of the UGV surrounding objects. The main contribution of this paper is of three folds: (i) the novel design of the HMD; the paper proposes an alternative design to the 3-D interface designs recently used in tele-operated search and rescue (SAR) UGVs. Unlike other designs that suggest to automatically move the whole UGV about two axes (pitch and yaw) with the movement of the head, this design suggests to let a separate unit of the UGV automatically move with the movement of the head and provide the user with VR. (ii) the distributed feature; the design allows multiple users to connect to the UGV using a wireless link in a secure way to receive video feeds from three on-board cameras. This feature facilitates cooperative team work in urban search and rescue (USAR) applications (a contemporary research issue in SAR UGV). (iii) a novel feature of the design is the simultaneous video feeds which are sent to the operator station computer for object detection using the scale-invariant feature transform (SIFT) algorithm and 3-D model construction of the UGV's surrounding objects from 2-D images of these objects. The design was realized using a smart phone-based HMD, which captures head movements in real time using its inertial measurement unit (IMU) and transmits it to three motors mounted on a rover to provide the movement about three axes (pitch, yaw and roll). The operator controls the motors via the HMD or a gamepad. Three on-board cameras provide video feeds which are transmitted to the HMD and operator computer. A software performs object detection and builds a 3-D model from the captured 2-D images. The realistic design constraints were identified, then the hardware/software functions that meet the constraints were listed. The UGV was implemented in a laboratory environment. It was tested over soft and rough terrain. Results showed that the UGV has higher visual-inspection capabilities compared to other existing SAR UGVs. Furthermore, it was found that the maximum speed of 3.3 m/s, six-wheel differential-drive chassis and spiked air-filled rubber tires of the rover gave it high manoeuvrability in open rough terrain compared to other SAR UGVs found in literature. The high visual inspection capabilities and relatively high speed of the UGV make it a good choice for planetary exploration and military reconnaissance. The three-motors and stereoscopic camera can be easily mounted as a separate unit on a chassis that uses different locomotion mechanism (e.g. leg type or tracked type) to extend the functionality of a SAR UGV. The design can be used in building disparity maps and in constructing 3-D models, or in real time face recognition, real time object detection and autonomous driving based on disparity maps.

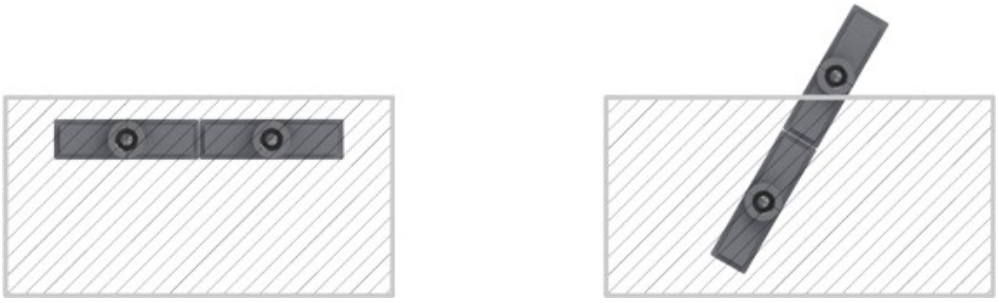
KEYWORDS

UGV, Virtual reality, Search and rescue, Robotics, Human-robot interaction.

1. INTRODUCTION

In tele-operated SAR UGV, the visual perception of the UGV environment and its presentation to the operator has a deep impact on human-robot-interaction (HRI) awareness of the UGV environment [1]-[2]. Many tele-operated SAR UGV designs rely on gathering as much data as possible from the UGV's surrounding via sensors and transferring this along with a video feed from an on-board camera to the operator's base station using wired and/or wireless communication links to enhance the cognitive ability of the UGV. [3]-[7]. The data is usually presented to the operator using a 2-D screen, hence only a proportion of the screen is used for video and the rest is used for displaying the data collected by the sensors in a user-friendly way. For example, in RAPOSA [8], only 29% of the total screen is

M. Al-Khalil, R. Abu-Rhayyem and A. Hammoudeh are students at Department of Computer Engineering, Princess Sumaya University for Technology, Amman, Jordan.
*Corresponding author: Talal A. Edwan is with Department of Computer Engineering, Princess Sumaya University for Technology, Amman, Jordan. Email: t.edwan@psut.edu.jo.



(a) Two cameras behind a barrier

(b) After rotating by 60 degrees, one camera is on top of the barrier

Figure 1. Two-camera design, where software rotation of images is not suitable.

used for video display. In addition to that, in many designs, if the operator wants to visually inspect the environment from a different viewpoint, he/she has to move the whole UGV. Similarly, Adora [9] uses a sensor box with single camera all mounted on a movable arm. More screen area was dedicated for the video feed, but some is reserved for viewing data collected from sensors. It has been found that a dichotomy in this type of research projects exists: either concentrating on conducting research on UGV mobility in rough terrain or increasing the cognitive/intelligent behaviour of the robotic assets, but rarely on both research domains [10].

Some implementations of user interfaces mitigated the problem by having a command for hiding all the clutter in the interface, leaving simply a clear cut view from the UGV visual output. The point of this is to quickly reduce sensory overload during complex situations [1]. Other implementations adopted a 3-D interface, for example, an attempt was made to increase the visual perception in RAPOSA using a 3-D interface [11]. This interface design is based on an HMD equipped with a head tracker; the HMD displays the images from a pair of video cameras located in the UGV frontal body, where the video stream of each camera is displayed to each operator eye. Because these two video feeds pertain to two slightly different viewpoints, it is possible with image rectification to endow the operator with depth perception (stereopsis). The pitch angle is used to control the UGV frontal body up/down, the yaw angle is used to rotate the UGV and the roll angle is used to rotate the images (the HMD has to counter-rotate the images to compensate for head movement in this direction). The shortcoming of this approach in our opinion is of two folds: first, the HMD highly depends on the type of the UGV in use, and second, is the inability to compensate for head movement in the roll direction by software-rotating the images in case more than one camera is used. To illustrate the latter point, we consider the situation depicted in Figure 1. a, where two cameras are initially behind a barrier, then after rotating by 60 degrees as shown in Figure 1. b, one of the cameras is positioned on top of the barrier and thus can provide a different view. We therefore conclude that the software compensation approach in [11] is suitable for a single camera, but not for multiple-camera design.

This paper proposes an alternative design to the 3-D interface designs recently used in tele-operated search and rescue (SAR) UGVs. Unlike other designs that suggest to automatically move the whole UGV about two axes (pitch and yaw) with the movement of the head, this design suggests to let a separate unit of the UGV automatically move with the movement of the head. A smart phone-based headset with VR capability captures head movements in real time using the built-in IMU in the smart phone and transmits it to three motors mounted on the same vertical axis on a six-wheel rover. The three motors are arranged to allow free movement about three axes (pitch, yaw and roll). The motors' vertical axis is equipped with a stereoscopic camera (i.e. two cameras separated by a distance) with the motors, constituting one separate unit, which captures video and transmits it in real time back to the headset. This gives the operator the flexibility to control the UGV view direction by just moving his/her head in the desired direction. We refer to this mode of operation as "automatic mode", in contrast to the "manual mode", where the operator can control the motors by a gamepad. The operator can switch between the two modes by pressing a button on the gamepad. A third camera in the midpoint between the two cameras is used to capture pictures on-demand by pressing a certain button on the gamepad. The resultant pictures are saved on an on-board SD-card and can be transmitted using a wireless connection to the operator's computer. Two special computer programs run on the

operator's computer: one is for object detection using the SIFT algorithm and the other is for 3-D model creation from the 2-D captured pictures of the UGV's surrounding objects. Further, the operator can control the six-wheel rover based on the visual feedback using a gamepad. The six-wheel rover offers relatively high climbing capabilities and performes very well over soft and rough terrain. The design allows multiple users to connect to the UGV in a secure way and receive the video feed from the cameras which facilitate cooperative team work in USAR applications.

The rest of the paper is divided as follows: the section "Design Requirements" describes the required specifications of the UGV; the section "Abridged Design" describes in a concise way the overall design; the section "Mechanical Design & Control" gives details about the hardware of the UGV, including the 3-axis rotating platform and chassis and how they are electrically controlled; the section "Stereoscopic Camera Set & Pi camera" describes the set-up of the cameras and their functions and means of communication with the operator station; the section "Data Processing Unit & Video Transmission" describes the hardware of the UGV's on-board processing unit and the software used to transmit the video to the operator station; the section "VR Headset" gives details about the HMD; whereas the section "Operator's Station" describes the operator station computer and the programs that run on it to handle the image processing; finally the section "Conclusion and Future Work" concludes the paper and gives an outline about future work.

(a) Basic design concept

(b) Main components

(c) Climbing

(d) On tough terrain

Figure 2. The UGV.

2. DESIGN REQUIREMENTS

Our work aims to achieve a set of design requirements. A brief description of each design requirement is provided below:

(i) Build an UGV able to hold a minimum weight of 3 kilograms: for a UGV to be able to carry heavy weights, its wheels and suspension systems need to be flexible enough so that they do not break if a heavy weight is applied on them. Also, the chassis needs to be rigid enough for not to break or bend due to heavy weight.

(ii) Build a UGV that can move at an approximate speed of 4 m/s: to be able to achieve this

requiremet, the UGV motors need to have enough torque to rotate the wheels that carry the UGV body and the other components that are placed on the UGV. Also, the motors are required to rotate at an approximate speed of 4 m/s while a load is applied on their shafts. To achieve such a torque, high power motors with gear reduction ratio or motors that deliver high torque and decent rotation speed are needed.

(iii) Build a 3-axis rotating platform that is able to rotate in the three Euler angles (Yaw, Pitch and Roll): the 3-axis rotating platform must have two cameras mounted on it to simulate human eyes and a third special purpose camera. The average total weight of cameras must be approximately 120 grams; high-precision motors must be used in order to rotate at a speed close to normal human-head rotating speed.

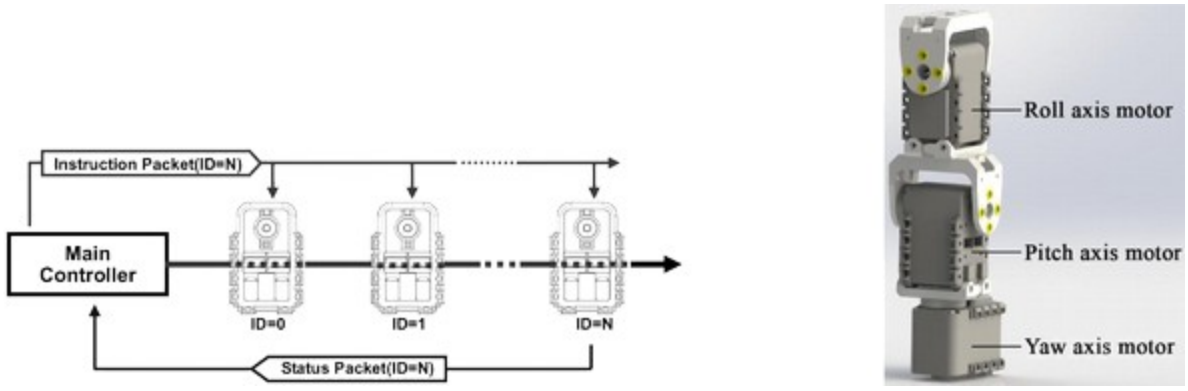
(iv) Transmit video feed from the UGV to a smartphone, with a transmission delay less than 1 second: wireless transmission is to be used to transmit video to the head set and operator's station, fast transmission to the headset is crucial to maintain synchronization between head movement and video feed. This requires the use of a wireless link with high date rates, large bandwidth, low latency and low interference.

(v) Operational radial distance from the operator's station is 50 meters: this requires a wireless link of low latency, high data rate and high signal power to operate efficiently over the required distance.

(vi) Minimum UGV trip time of 15 minutes: this requirement depends totally on the power source, the power supply must deliver enough current and voltage to drive all of the on-board circuitry and motors for the required operational time.

3. ABRIDGED DESIGN

The developed system consists of three subsystems: the operator's station, the VR headset and the UGV, all connected via an access point as depicted in Figure 2. a. The operator's station consists of a laptop and a gamepad, while the VR headset consists of a smart phone. Each of these two subsystems runs different software capable of controlling two different parts of the UGV. The laptop software reads the control data from a gamepad; this data can be used to simultaneously drive the UGV and move the stereoscopic camera by moving three control motors which constitute a 3-axis rotating platform. The smart phone (VR headset) runs different software which tracks head motion using the smart phone's IMU and transmits it to the UGV to control the 3-axis rotating platform. The operator can seamlessly switch between the manual mode (using the gamepad) and the automatic mode (using VR headset) to control the 3-axis rotating platform.



(a) Communication protocol overview (b) Assembly of the 3-axis rotating platform

Figure 3. 3-axis rotating platform.

Two video feeds are provided from the stereoscopic camera and Pi camera; the stereoscopic-camera video feed goes to the smart phone which in turn displays it to the operator using a VR software and to the operator's station for object detection, whereas the Pi camera video feed goes to the operator's station for 3-D model construction. The subsystems communicate using managed Wi-Fi (IEEE 802.11 standards). Network programming using Python and Java programming languages was used to create transmission control protocol (TCP) connections between the operator's station and the UGV on one

side and between the VR-headset and the UGV on the other side. Figure 2. b shows the main components of the UGV and Figures 2. c and 2. d show the UGV in real test over tough terrain.

4. MECHANICAL DESIGN & CONTROL

4.1 3-AXIS ROTATING PLATFORM

The 3-axis rotating platform is an essential part of the UGV; it mimics the human head movements. The platform was implemented using Dynamixel AX-12 actuator which consists of a gear reducer (to produce high torque), a high-precision DC motor and an internal controller for networking functionality. Table 1 shows the specifications of Dynamixel AX-12 actuator.

Table 1. Dynamixel AX-12 actuator specifications.

	Dynamixel AX-12
Weight (g)	55
Gear Reduction Ratio	1/254
Input Voltage	7V or 10V
Final Max Holding Torque (kgf.cm)	12 @7V and 16.5 @10V
Sec/60 degree	0.269 @7V and 0.196 @10V

The actuator has the ability to detect internal conditions, such as excessive voltage and internal temperature, using a built-in controller. Communication with the actuator is achieved by means of half duplex asynchronous serial communication channel, which uses only one wire for both transmitting and receiving the data. The size of a serial frame is 9-bits; 8-bits for data and one bit for parity. To control the Dynamixel actuator, a packet of bytes is sent through the half-duplex serial channel which contains the control command; this packet is called the "instruction packet". After the actuator's controller receives the packet, it performs the command the packet contains, afterwards, the actuator sends back a feedback packet called the "status packet", which contains either an error flag if an error occurred or data from the controllers' registers. An overview of the communication protocol is

(a)Schematic diagram of the actuators, powering source of the USB2AX device and
(b) Implementation of the 3-axis rotating platform with the stereoscopic camera mounted on it.

Figure 4. Design and implementation of the 3-axis rotating platform.

illustrated in Figure 3. a. In order to let the 3-axis rotating platform rotate in the three Euler rotation angles (Yaw, Pitch and Roll), the three Dynamixel AX-12 actuators were assembled as shown in Figure 3. b. Each actuator was given a unique ID which was written on the register of address 0x03 using the broadcast ID instruction; this was separately done for each actuator. The format of the instruction packet is as follows:

0xFF 0xFF ID LENGTH INSTRUCTION PARAMETER 1 . . . PARAMETER N CHECKSUM

- 0xFF: These two bytes indicate the beginning of an instruction packet.

- ID: This byte contains the unique ID of a Dynamixel actuator in the network; there are only 254 available ID values.

A broadcast ID is the when transmitted ID value is equal to 0xFE, which means that any instruction packet with this ID gets transmitted to all actuators connected in the network, and the actuators do not return a status packet.

- Length: This byte hold the length of packet; its value is calculated as (# of parameters + 2).

- Instruction: This byte contains the instruction that the actuator is required to perform; it can hold one of these values.

- Parameters: These bytes contain the parameters needed for the instruction (e.g.: Angle, Torque, ...).

- Checksum: A Check Sum byte is added at the end of the instruction packet to insure integrity of data being delivered. The check sum value is calculated as follows.

Checksum = ~[ID + Length + Instruction + Parameter1 + … + ParameterN]

where (~) represents the NOT logic operation.

The instruction packet sent to each actuator was of the following format:

0xFF 0xFF 0xFE 0x04 0x03 0x03 0x01 0xF6

- 0xFF: These two bytes indicate the beginning of an instruction word.

- 0xFE: This ID is the broadcast ID defined above and is used to send instruction packets to all the connected Dynamixel actuators in the network regardless of their ID values.

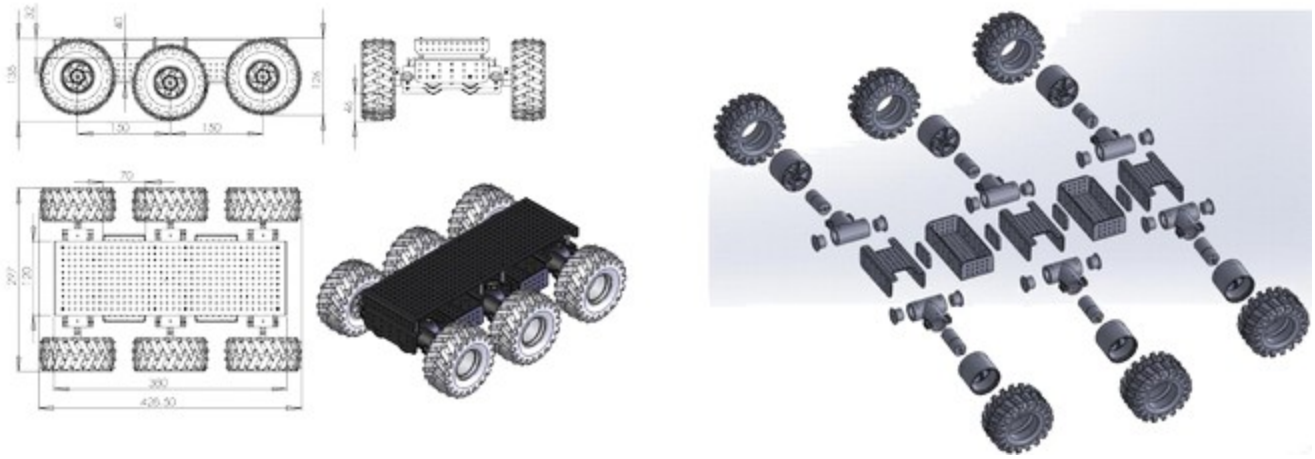
- 0x04: This byte represents the length of the instruction packet (2 parameters + 2 = 4).

- 0x03: This byte indicates a WRITE DATA instruction.

- 0x03: This byte represents the first parameter which is the address of the wanted registers for this packet. In this case it is the ID register.

- 0x01: This byte represents the second parameter which is the data to be written in the desired register (0x01, 0x02 and 0x03), respectively, for each actuator.

- 0xF6: This byte represents the checksum for the instruction packet.



(a) Chassis dimensions (b) Chassis dissemble view

Figure 5. Six-wheel rover model.

The on-board processing unit is a load-balancing cluster of two Raspberry Pi 3 model B: #1 Pi and #2 Pi. The actuators are connected to #2 Pi using a half-duplex asynchronous USB-to-TTL module (Xevelabs mini USB2AX v3.2a). The schematic diagram of the connection is illustrated in Figure 4.

(a) and the implementation in Figure 4. (b). Since the processing unit has native support for Python, the library Pyax12 was used in conjunction with Python's native socket library to continuously construct instruction packets and send them to the 3-Axis rotating platform actuators.

The processing unit has two modes for operating the 3-axis rotating platform; VR-headset mode and manual mode, (i) VR-headset mode: The processing unit uses the data extracted from the movement of the operator's head in the operator's station to rotate the actuators accordingly; this mode is the default mode. (ii) Manual mode: The processing unit uses the data received from the gamepad at the operator's station to rotate the actuators accordingly.

The processing unit initializes the serial port of the Xevelabs USB2AX device, with a baud rate of 1 Mbps. Then, it sends an instruction word to all the actuators to set the compliance slope value to 0x92, in order to avoid damage to the actuator's shafts due to continuous change in direction. The processing unit then initializes two network sockets: one for the VR mode and one for the manual mode. The sockets then listen for incoming connections. Once a connection has been established between the operator's station or VR headset and the UGV, the UGV processing unit begins receiving data (Yaw, Pitch and Roll values) via wireless link continuously as a stream of 12 bytes for each frame, then it reconstructs the Yaw, Pitch and Roll integer values using the native "struct" library, where each integer value consists of 4 bytes. Afterwards, the processing unit constructs an instruction packet for each of the Yaw, Pitch and Roll values and transmits each packet to its corresponding actuator on the 3-axis platform. If the operator's station sends a switch mode command via the wireless link by pressing a certain button on the gamepad, the processing unit stops constructing instruction packets from the Yaw, Pitch and Roll values and starts using the data coming from the gamepad of the operator's station instead. Sending another switch mode command toggles the operation mode again.

The 3-axis rotation platform is powered by the main power source of the UGV, which is a LiPo battery. The LiPo battery has a full-charge voltage of 13.4 volts. This voltage cannot be directly connected to the actuators because the operational voltage range of the actuators is 7-12 volts; therefore, a variable-voltage regulator was used to regulate the voltage of the LiPo battery to 10.5 volts so that the actuators can operate properly.

4.2 UGV Chassis

The locomotion mechanism used in the UGV is enhanced wheel type; particularly, a six-wheel driven rover with a differential-drive chassis. Vehicle-turning is accomplished by driving the motors on the two sides of the platform at different directions and speeds. For example, to rotate the rover around itself in counter-clockwise direction, the right-side tires should be moved clockwise and the left-side tires should be moved counter-clockwise. The wheels are spiked-rubber tires filled with air. The rover is "6-Wheels Wild Thumper", from Pololu [12]; the driving motors included in the chassis have a gear reduction ratio of 34:1, which provides the required speed and torque. Also, every two opposing tires are connected with a unique "super-twist" suspension system which gives the rover climbing capability. Figure 5. (a) shows the rover's dimensions and Figure 5. (b) shows a disassemble view of its 3-D model. The main specifications of the rover are listed in Table 2.

We added six DC motors; three motors on each side controlled by two H-bridge module, one for each side. The rover's controller (an Arduino UNO board) is connected to # 2 Pi by an USB-to-TTL module (CH340G). The Arduino UNO board is programmed to receive two bytes from #2 Pi each time the operator moves the joystick of the gamepad; the first byte defines the direction of the rover's movement according to Table 3, where the second byte defines the speed of the rover's movement which is a value from 0 to 9; a value of 0 means the rover is stopped and a value of 9 means that highest speed in the specified direction. Speed data is dictated to the board from # 2 Pi; accordingly, the board controls the speed of the motors using Pulse Width Modulation (PWM); that is, by adjusting the duty cycle.

The Arduino UNO board first initializes serial communication port with a baud rate of 9600 bits per second, then initializes the PWM duty cycle of all motors to zero. Once data is available on the serial communication channel, the Arduino reads two bytes. Then, the Arduino starts increasing/decreasing the PWM duty cycle factor with value of one each three milliseconds. In this way, the Arduino UNO keeps the motors safe from current spikes and instant torque on the gearbox of the motors shaft.

Table 2. UGV rover specifications.

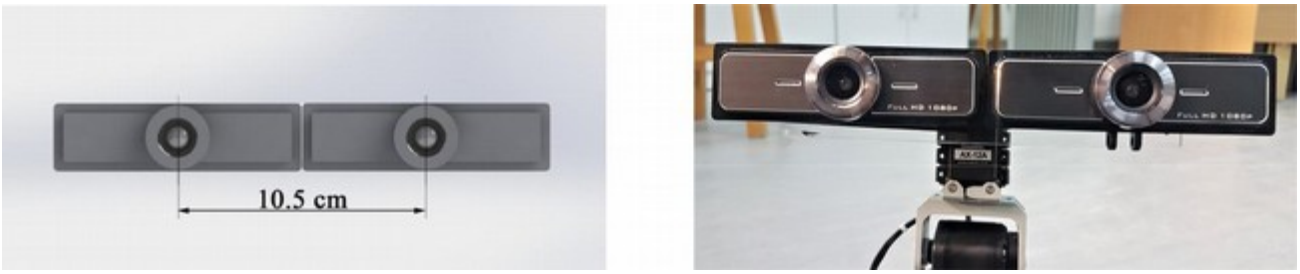
Size	420 x 300 x 130 mm (16.5" x 12" x 5").
Weight	2.7 kg (6.0 lb) (including wheels and motors only).
Max. payload	5 kg (11 lb).
Motor voltage	2 -- 7.5 V.
Stall current	6.6 A per motor.
No-load current	420 mA per motor.
No-load speed	350 RPM for the 34:1 gear reduction motors.
Stall torque at 7.2V	5 kg-cm (70 oz-in) per motor for the 34:1 gear reduction motors.

Table 3. Bytes sent to the rover's microcontroller.

First byte	Corresponding direction
0x00	Forward
0x01	Backward
0x02	Left
0x03	Right

Two H-bridge modules (two IBT-2 H-bridge modules each having two In neon BTS7960B half bridges) were used; each H-bridge controls three wheels on the same side. The module has a large heat sink and can withstand currents up to 43 amperes, with a maximum drive voltage of 27 volts. The module also has a built-in voltage regulator that provides 5 volts for the user to use in other applications. However, since the motors require 7.2 volts to operate, a DC-DC voltage step down converter was used to convert the battery's 12.4 volts (full-charge voltage) to 7.2 volts.

5. STEREOSCOPIC CAMERA SET & PI CAMERA



(a) Stereoscopic camera model, (b) Actual developed design of the stereoscopic camera set

Figure 6. Stereoscopic camera.

Two types of camera were used: a stereoscopic camera (Genius F100 widecam 12-megapixel sensor) and a Pi camera. The former is used to get video feed of the UGV point of view (POV), while the latter is used to take high-resolution pictures used to build a 3-D model of the rover's surrounding objects.

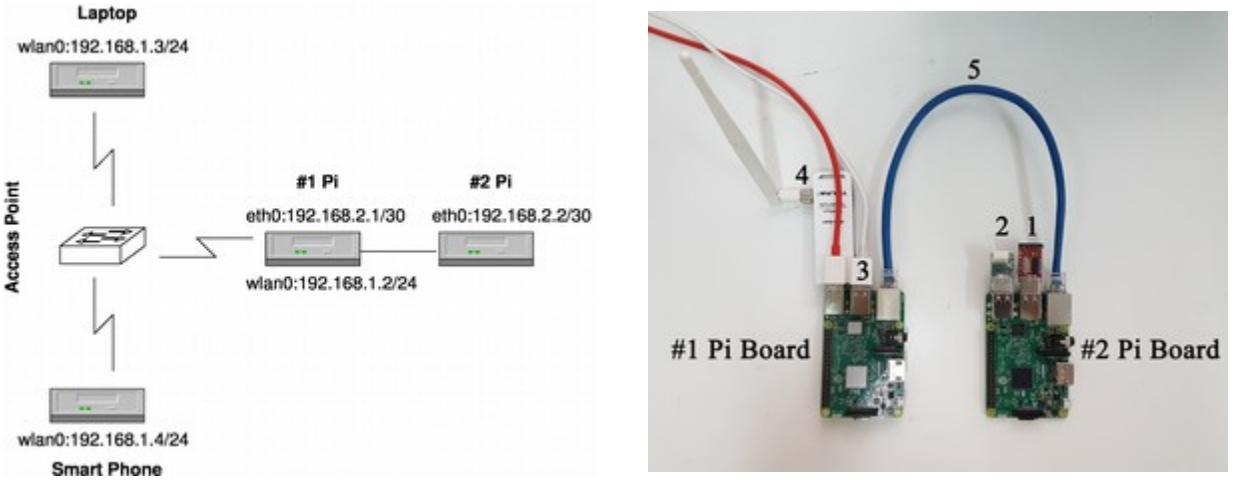
The stereoscopic camera set consists of two cameras aligned horizontally and separated by a distance of 10.5 cm measured from the centre of each lens as shown in Figure 6. (a). The distance between human eyes is usually referred to as the "Pupillary Distance"; the average value of this distance is 6.5 cm. However, larger distance results in more depth that can be felt when using the VR application. The stereoscopic camera is mounted on a "Plexiglass" plate that has a thickness of 4 mm, a length of 15.5 cm and a width of 5.5 cm. The plexiglass plate is mounted on the 3-axis rotating platform as depicted in Figure 6. (b). The plexiglass plate can be seen underneath the cameras. The two cameras were chosen mainly for their wide-angle lens, high-resolution and manual focus lens. These three specifications are crucial for any VR application. Table 4 lists the specifications of the cameras. Serial communication is achieved via a USB-2 cable connected to # 1 Pi.

The UGV is equipped with additional camera (Pi Camera Module v2 based on Sony IMX219 8-megapixel sensor); this camera is located at the mid-point between the two other cameras (see Figure

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 4, No. 1, April 2018.

Table 4. Stereoscopic camera specifications.

	Genius F100 WideCam
Image sensor	1080p full HD pixel CMOS
Focus type	Manual focus lens
Interface type	USB-2
View angle	120 degrees
Body weight (g)	82



(a) Network set-up (b) Raspberry Pi boards' cluster set-up

Figure 7. Network & cluster set-up.

2. (b)) and connected to #1 Pi using the on-board 15-pin camera serial interface (CSI-2). The camera is activated in the manual mode and is used to capture pictures for the purpose of creating a 3-D model of the UGV surrounding objects. The Python script which controls the 3-axis rotating platform is also used to capture the pictures using the Pi Camera. This is done when the operator presses the manual mode button on the gamepad then presses the capture button on the gamepad. If the operator did so, each captured picture is saved as an image le in a folder stored in an on-board storage SD card on #1 Pi board. To retrieve the captured images, a secure le transfer protocol (SFTP) server was created on #1 Pi board, where the operator can access this SFTP server by using an SFTP client software on the operator-station laptop. After the operator accesses the SFTP server, he/she can copy the images to the operator-station laptop and pass the images to dedicated software (RealityCapture) to process the images and create a 3-D model of the UGV surrounding objects. It is worth to emphasise that each camera has a separate TCP connection to the operator's laptop. that is, the stereoscopic camera has a dedicated connection to TCP port number "41222" at the laptop, while the Pi camera connects to TCP port number 22 (SFTP). Two feeds from the stereoscopic camera are sent to the HMD; one per eye. There are lenses which are placed between the operator's eyes and the screen of the smart phone. These lenses focus and reshape the image for each eye and create a stereoscopic 3D image by angling the two 2D images to mimic how each of the operator's two eyes views the UGV surrounding.

6. DATA PROCESSING UNIT & VIDEO TRANSMISSION

The data processing unit consists of a cluster of two Raspberry Pi 3 model B boards connected back-to -back using a tethered Ethernet link. As shown in Figure 7. (b), a number of devices are attached to the two boards:

1. USB-to-TTL device, which communicates with the UGV chassis microcontroller.
2. USB-to-TTL device, which communicates with the internal microcontrollers of each motor in the 3-axis rotating platform.
3. USB cable of the stereoscopic camera set.
4. USB Wi-Fi dongle, which connects #1 Pi board to the access point.
5. The Ethernet cable, which connects the Pi boards together.

For a general overview of the network set-up, a network diagram was drawn in Figure 7. (a) to aid the illustration in this section. The #1 Pi board was programmed using Java programming language and utilizing OpenCV library to transmit two video feeds; one from the stereoscopic camera to the VR headset and another from the Pi camera to the operator-station laptop. In video transmission, we can tolerate a certain amount of packet loss for low delays, that is no need for retransmission which increases delay. For this reason, UDP or RTP (unreliable data transfer) are usually used instead of TCP (reliable data transfer) when transmitting audio and/or video. Delay and jitter are the important factors when transmitting audio/video. However, in our case we used an optimised version of TCP congestion control algorithm (TCP-BBR [21]) that has low latency, yet insures reliable data transfer. TCP was used as a transmission protocol for reliable data transfer of both control commands – where reliable data transfer is necessary – and video. Inthis way, one connection and one protocol were used for data transfer. Figure 8 shows a flowchart of the main thread which runs on the #1 Pi.

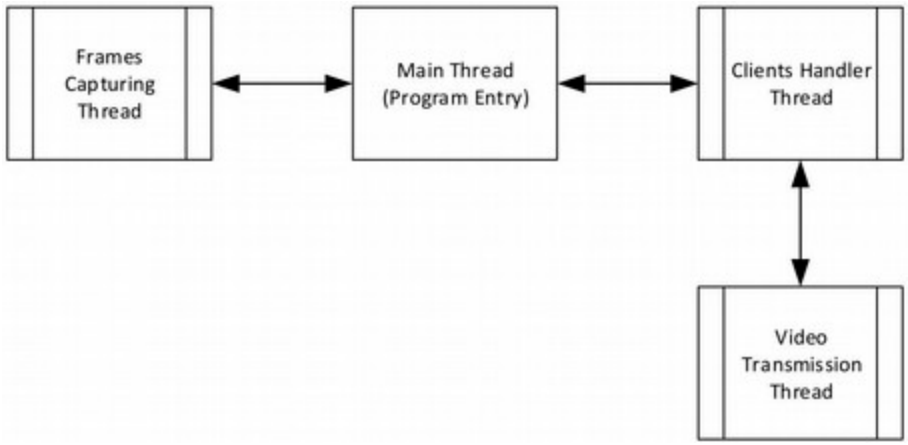


Figure 8. Flowchart of the main thread which runs on #1 Pi.

The main thread starts another two threads: the "client's handler" thread and the frames' "capturing thread". Another thread called the "video transmission thread", however, is started by the client's handler thread. When the main thread starts, it creates data containers for the compressed cameras' frames, then it starts the capturing thread and the client's handler thread. Finally, it stays in wait mode until the program is terminated.

When the capturing thread starts, it loads the OpenCV library, then it initializes the stereoscopic camera set by setting the resolution of the camera frames to 640x480 pixels. This resolution is relatively low and provides the required frame rate. To illustrate, the Raspberry Pi CPU, can process a maximum of 15 frames/second at this resolution. At higher resolutions, the frame-processing rate drops drastically. After initialization, the CPU starts receiving frames from the stereoscopic camera set. Eeach time it receives a frame, it checks the size of the frame, then it compresses the captured frame using JPEG lossy compression with a compression factor of 78%. Incorrect frame size results in the frame being dropped. Figure 9. a shows the flowchart of this thread.

The client's handler thread runs concurrently with the capturing thread, yet the CPU time is shared between the two threads. The client's handler thread is responsible for initializing a server on the #1 Pi, which accepts connections from the VR headset and the operator's station. The flow chart of the thread is shown in Figure 9. b. When the thread starts, it initializes a server socket and binds it to port number "41222", then it keeps listening on this port for incoming connections from the VR headset and/or the operator's laptop. Any new connection is passed to the video transmission thread. It is worth noting that the design can allow multiple VR headsets or operator laptops to connect and receive video feeds from the cameras, which facilitates cooperative team work in SAR operations.

The video transmission thread transmits the compressed frames by first checking whether the compressed frame's container is accessible; that is because the container is also accessible by the capturing thread and it cannot be accessed simultaneously by two threads. If the container is accessible, the thread further checks whether the container is empty or not. In case it is empty, no data

is sent. On the other hand, if the container is not empty, the thread writes the frame's bytes on the socket output stream. The #2 Pi board was programmed to deliver commands to all on-board microcontrollers through the USB-to-TTL serial interfaces.

7. VR HEADSET

The VR headset consists of two parts: the VR glasses (Arealer virtual reality glasses headset) and the smart mobile phone (Samsung S7 edge). The smart phone is used to receive video feed from the UGV stereoscopic camera and view it on a vertically split screen. The VR glasses use two lenses to correct the light angle that comes out of the smart phone screen which is placed inside the VR glasses as shown in Figure 10. a. This gives the operator a 3-D view.

The smart mobile phone runs a software which traces the user head movement based on the angles' data gathered from the built-in IMU. This data is then transmitted to the UGV. The same software is also responsible for receiving and displaying video feed from the stereoscopic camera. The two main parts of this software are discussed in the following sub-sections.

7.1 Head-Movement Tracking & Transmitting Thread

The software which runs on the smart phone runs a head-movement tracking thread which captures the Yaw, Pitch and Roll rotation angles of the smart phone using the data gathered from the Accelerometer and Geomagnetic Field Sensor of the smart phone's IMU. The data of the three angles is then transmitted to the UGV's #1 Pi board which in turn sends it to #2 Pi board. We adhere to the following definitions of the three rotation angles (we refer the reader to Figure 10.b for the orientation of the axes):

(i) Yaw (degrees of rotation about the z-axis): "This is the angle between the device's current compass direction and magnetic north. If the top edge of the device faces magnetic north, the azimuth is 0 degrees; if the top edge faces south, the azimuth is 180 degrees. Similarly, if the top edge faces east, the azimuth is 90 degrees and if the top edge faces west, the azimuth is 270 degrees".

(ii) Pitch (degrees of rotation about the x-axis): "This is the angle between a plane parallel to the device's screen and a plane parallel to the ground. If you hold the device parallel to the ground with the bottom edge closest to you and tilt the top edge of the device toward the ground, the pitch angle becomes positive. Tilting in the opposite direction and moving the top edge of the device away from the ground causes the pitch angle to become negative. The range of values is -180 degrees to 180 degrees".

(iii) Roll (degrees of rotation about the y-axis): "This is the angle between a plane perpendicular to the device's screen and a plane perpendicular to the ground. If you hold the device parallel to the ground with the bottom edge closest to you and tilt the left edge of the device toward the ground, the roll angle becomes positive. Tilting in the opposite direction and moving the right edge of the device toward the ground causes the roll angle to become negative. The range of values is -90 degrees to 90 degrees".

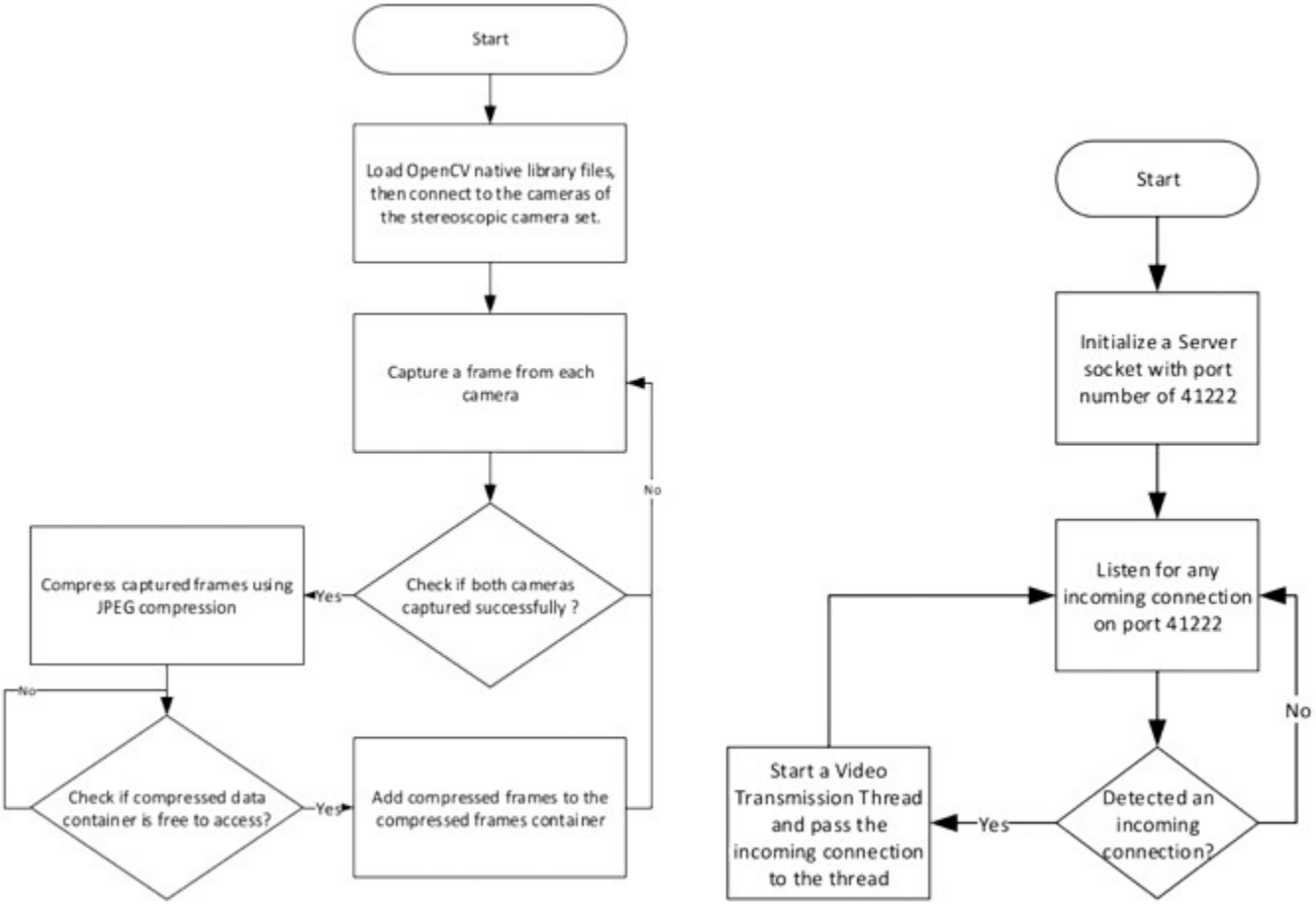
The android development environment has a native hardware support library for the IMU, which contains a number of classes: Sensor, Sensor Event, Sensor Event Listener and Sensor Manager. The thread is based on these classes:

(i) Sensor: this class is used to register the sensors in the device. It should be noted that not all android devices contain the same sensors and features.

(ii) Sensor Event: this class represents a Sensor event and holds information such as the sensor's type, accuracy and of course the sensor's data.

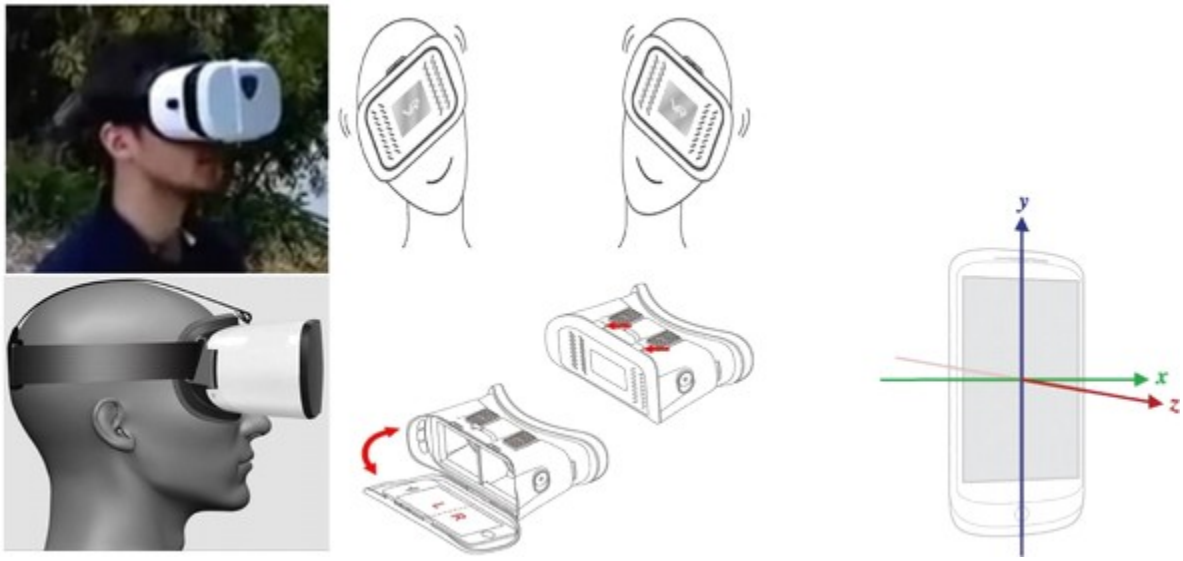
(iii) Sensor Event Listener: this class is used for receiving notifications from the Sensor Manager when there is new sensor data.

(iv) Sensor Manager: this class allows access to the device's sensors. Initializing an instance of this class is done by calling Context.getSystemService() with the argument SENSOR_SERVICE.



(a) Capturing thread (b) Client's handler thread

Figure 9. Flowcharts of threads.



(a) VR head mounted display (b) Axes orientation of the smart phone

Figure 10. VR Headset.

The flowchart of the thread is depicted in Figure 11. When the thread starts, it initializes an instance of type "Sensor Manager" in order to register the sensors which the thread wants to use. In our case, these are the Accelerometer and Geomagnetic Field Sensors. A method (onSensroChanged()) is called each time the sensors' values change to pass the new values to the getRotationMatrix() method, which

in turn computes the inclination matrix and the rotation matrix, transforming a vector from the device coordinate system to the world's coordinate system (direct orthonormal basis). The rotation matrix is then passed to another method (getOrientation()) to compute and return the three rotation angles (Yaw,Pitch and Roll angles) according to the mathematical formulae described in [13]. Once the angles are obtained, they are parsed into three integers which are then sent to the UGV #2 Pi board via #1 Pi board to rotate the 3-axis rotation platform accordingly.

7.2 Headset Video Receiver Thread

Another thread was developed to handle the video feed from the stereoscopic camera. Initially, the software asks the operator to enter the internet protocol (IP) addresses of the Pi boards (see Figure 12.a), subsequently control is moved to the video feed receiver thread to create a socket and establish a connection with the #1 Pi board. Afterwards, the headset starts receiving video feed and transmitting the head movement data. Figure 12.b illustrates the flowchart of the thread. Once the operator taps the connect button on the smart phone, the application starts a new activity that projects the received video feed in a vertically split screen, where each portion of the screen shows a different camera feed in full screen mode, then two threads are started; a thread that receives the data from the UGV and projects it on the screen, and another thread (discussed in the previous sub-section) which tracks the head movement and transmits it to the UGV.

The video feed receiver thread creates a client socket and establishes a connection to the UGV (#1 Pi) using TCP protocol. The thread then starts reading the transmitted frames from the socket's input stream: a frame of the right camera, then a second frame of the left camera, and so forth. The frames are then decompressed and displayed on the smart phone screen using Android ImageView class.

8. OPERATOR'S STATION

The operator's station consists of one processing unit and a laptop computer (Lenovo Y700, Core i7 6700HQ CPU, 16GB RAM, 1TB HDD). This unit is responsible for two operations: First, reading control data from the gamepad – which is attached to the laptop – and transmitting it to #2 Pi via #1 Pi. Second, receiving video feed from the Pi camera and performing the required image processing for: object detection using the SIFT algorithm [14] and 3-D model construction using RealityCapture software.

8.1 SIFT Algorithm

The SIFT algorithm extracts distinctive key points from images that can be used to match images of the same objects, but with different viewpoints. The algorithm is invariant to scaling and rotation and very robust for matching over affine changes, change in 3D viewpoint, noise and illumination. The extracted key points can be used for object recognition, in fact, it has been shown [14] that the algorithm is capable of robustly identifying objects among clutter and occlusion in near real-time. We give a brief overview of the SIFT algorithm steps below.

(i) Scale-space Extrema Detection

A corner may not be a corner if the image is scaled. To illustrate; in the image in Diagram – 1, a curve in the left image within a small window is linear when it is zoomed in the same window. It is therefore, obvious that we cannot use the same window to detect key points with different scale. In order to detect larger corners, larger windows are needed. SIFT algorithm uses scale-space filtering for this purpose. The scale space $L(x,y,\sigma)$ is defined as:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y),$$

where, $I(x,y)$ is the input image, the "*" is the convolution operator and the $G(x,y,\sigma)$ is the Gaussian kernel function, defined as:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma} e^{-\left(x^2+y^2\right)/2\sigma^2}$$

In principle, σ acts as a scaling parameter. For example, in Diagram 1 , a Gaussian kernel with low σ gives high value for small corner, while Gaussian kernel with high σ gives high values for larger

corner. So, local maxima across the scale and space can be found, which produces a list of (x,y,σ) values that act as a good candidates for key point at (x,y) at σ scale.

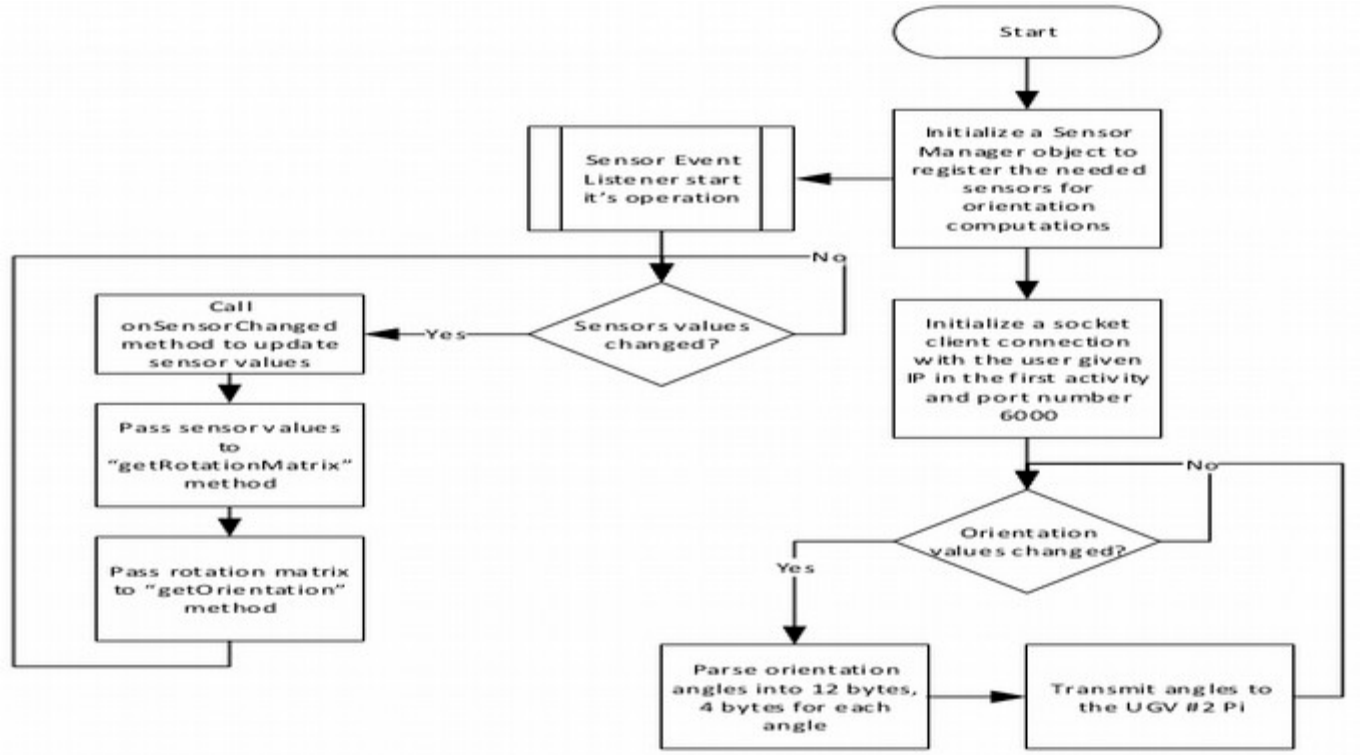


Figure 11. Flowchart of the head-tracking thread.

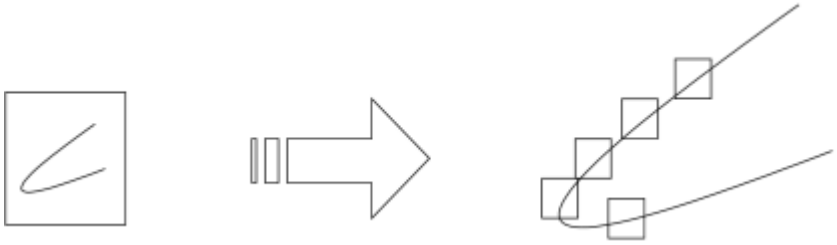


Diagram 1. Basic idea of scaling.

To improve the performance, SIFT algorithm uses Difference of Gaussians (DoG), illustrated mathematically by the following equation:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y)$$
$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma)$$

DoG is obtained as the difference of Gaussian blurring of an image with two different σ; that is σ and $k\sigma$. This process is done for different octaves of the image in Gaussian Pyramid, as shown in Diagram 2.

Once this DoG is found, images are searched for local extrema over scale and space. For example, one pixel in an image is compared with its 8 neighbours as well as 9 pixels in the next scale and 9 pixels in the previous scales – see Diagram 2. If it is a local extrema, it is a potential key point. In fact, this means that key point is best represented in that scale.

(ii) Key point Localization

The potential key points found in the first step are refined for higher accuracy. The algorithm uses Taylor series expansion of the scale space to get more accurate location of extrema. In addition to that, if the intensity at this extrema is less than a predefined threshold (0.03), it is rejected. It is worth noting that DoG has higher response for edges; so, edges also need to be removed. This is achieved by the

algorithm using a 2x2 Hessian matrix to compute the principal curvature. For edges, one eigen value is larger than the other. If the ratio of one eigen value to the other is greater than a threshold (10), the key point of the edge is discarded.

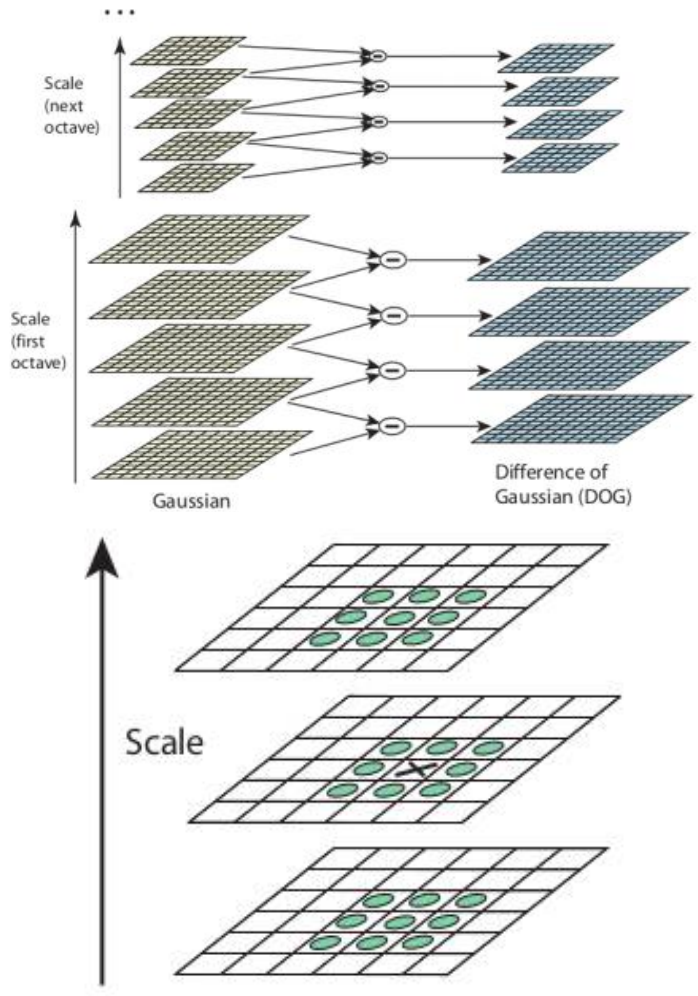


Diagram 2. Difference of Gaussians between images (Adopted from [14]).

(iii) Orientation Assignment

The orientation of each key point is then assigned to make the algorithm invariant to image rotation. Depending on the scale, a neighbourhood is taken around the key point; the gradient is then calculated in the vicinity of the key point and orientation histogram with 36 bins the width of each is 10 degrees (i.e., covering 360 degrees) is established. This is weighted by gradient magnitude and Gaussian-weighted circular window with σ equal to 1.5 times the scale of the key point. The orientation is determined using the angles that belong to the bins with peaks more than 80% of the maximum peak.

(iv) Key Point Descriptor

A key point descriptor for each key point is created by first taking 16x16 neighbourhood around the key point which is further divided into 16 sub-blocks of 4x4 size. Then, for each sub-block, 8-bin orientation histogram is created (i.e., total of 128 bin values). Finally, this is represented as a vector to form the key point descriptor.

(v) Key point Matching

Matching between key points is achieved by means of identifying their nearest neighbours. Due to noise, it is possible to have cases where the second closest-match is very near to the first. In that case, ratio of closest-distance to second-closest distance is taken and if it is greater than 0.8, they are

rejected. This eliminates approximately 90% of false matches while discarding only 5% of the correct matches [14].

8.2 The Gamepad

The gamepad used in our work is "Logitech F310". It is attached to the laptop and is used to control the 3-axis rotation platform in the manual mode. The gamepad consists of three analog joysticks; two of them move in two axes (see Figure 13.a) and the third one moves in one axis (Figure 13.b). One of the two-axes analog joysticks is used to drive the UGV and the other one with the one-axis analog joystick is used to move the 3-axis rotation platform in the manual mode.

A software was developed (see Figure 14) using Java programming language and utilizing the "Jinput" library to communicate with the gamepad using serial communication. Also, the software uses Java stream sockets to communicate with the UGV #2 Pi board using a dedicated TCP connection. Initially, the software searches for all input devices which are attached to the laptop. When the gamepad is recognized, it initializes a connection to the gamepad and starts receiving data from it through the USB port; at this stage, it also establishes a TCP connection to #2 Pi board. The gamepad sends data whenever the operator does an action on any of the buttons or joysticks of the gamepad. The data which is read from the gamepad contains information about the ID of the part that was moved and the movement (both magnitude and direction). The ID is checked by the software to determine which part was moved, e.g. if the ID matches the ID of the left joystick, this means that the operator wants to move the UGV; the software then determines the direction which the joystick was moved in (see Table 3) and maps the value of the analog joystick to a number in the range of 0-9; where number 0 means slowest speed (rover is stopped) and number 9 means the highest speed. Subsequently, the software sends the information about the magnitude and direction to the UGV #2 Pi board.

The software is also responsible for moving the 3-axis rotating platform in the manual mode. The toggling between manual mode and VR-headset mode is done by pressing the "A" button on the gamepad (see Figure 13.a). When the operator presses the button, the software sends a toggling command to the UGV #2 Pi board; subsequently, the software detects any movements of the relevant joysticks (mentioned in the first paragraph of this sub-section) and transmits the information to UGV #2 Pi board using the same TCP connection.

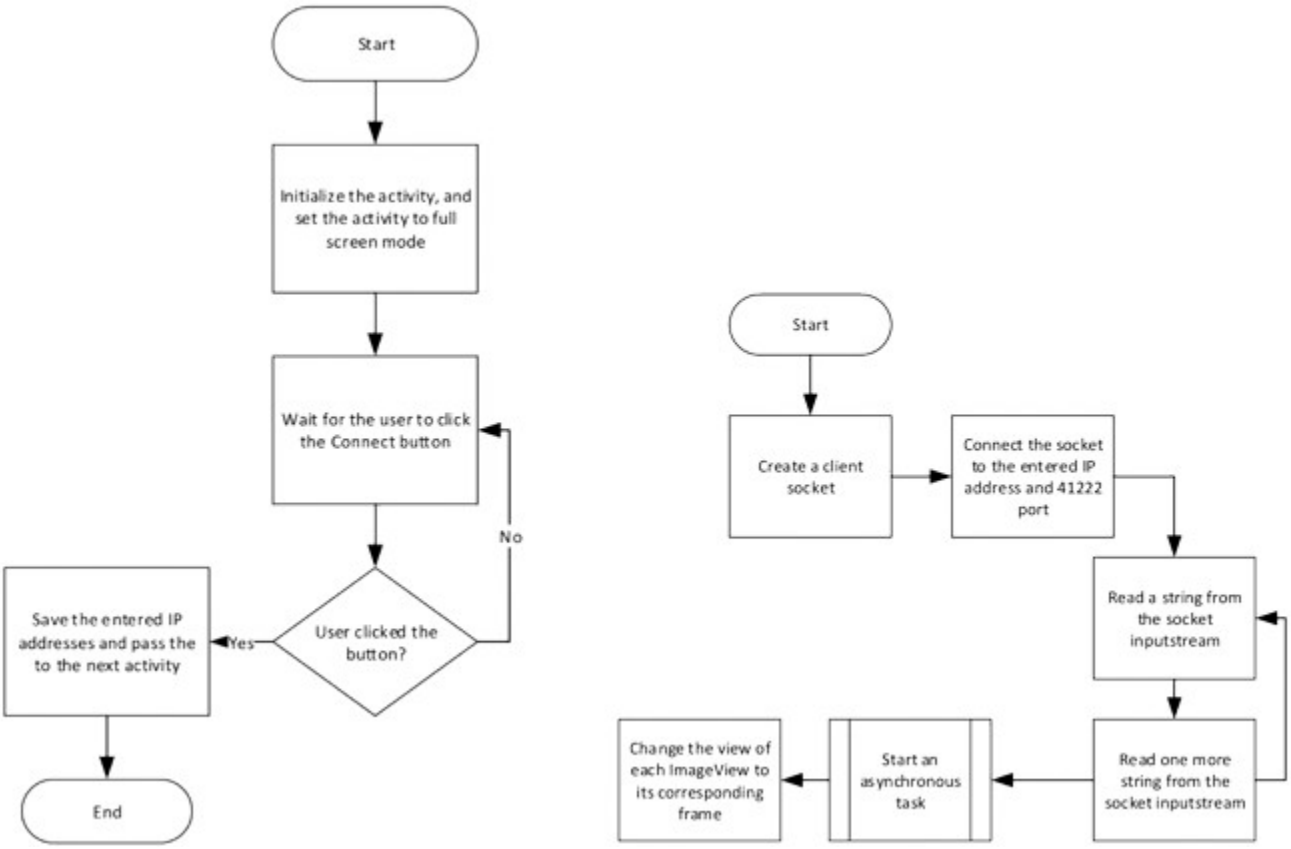
8.3 Object Detection & 3-D Model Creation

The operator-station laptop receives a video feed from both the stereoscopic camera and the Pi camera and applies image processing as follows:

1. The SIFT algorithm is used to compare the features extracted from right and left pictures received from the stereoscopic camera with the features of a reference object which the operator want to detect. To illustrate, for any object in a picture, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training reference image, can then be used to identify the object when attempting to locate the object in the pictures captured by the camera which contain many other objects. Technically speaking, this is implemented in two threads (flowcharts of which are shown in Figure 15.a and Figure 15.b). The receiving thread saves the pictures in a "List of Mat". In OpenCV library, the class "Mat" represents an n-dimensional array, which is used to store gray-scale or color images. Then, the SIFT algorithm computes key points and descriptors for both the reference image and the captured image; the distance is computed and if it is less than or equal to 25 – found experimentally to be a good match – the object is considered detected.
2. Another software (RealityCapture) is used to build a 3-D model of the UGV's surrounding environment from multiple of 2-D pictures taken by the operator using the Pi camera. The 2-D pictures are stored on an on-board SD card on #1 Pi and retrieved by the operator using a secure shell (SSH) connection. The software then automatically aligns images and detects matching keypoints of the images to create a 3-D mesh map. Finally, the operator can either choose to obtain a coloured or gray-scale high-quality texture of the object. Figure 16.a shows multiple 2-D images taken from the Pi camera in one of the rover trips, and Figure 16.b shows the 3-D model that was created from these images.

9. RESULTS & DISCUSSION

The diversity in UGV designs and specifications (e.g. weight, payload, speed, size, ...) renders a comprehensive performance evaluation of existing UGVs a none-trivial task. We restricted our comparison to the ADORA [9] and RAPOSA [11] UGVs, since both have similar design objectives as our UGV, specifically: mobility, victim identification, mapping (i.e., ability to generate 3D maps, of the surrounding areas and present them in an intuitive way), and HRI (i.e., ability to provide the operator with visual information about the robots' immediate surroundings). In addition to that, since delay is an important factor in teleoperated UGVs [23, 24] and because our UGV relies on TCP for transmitting video over WiFi, we studied the effect of delay on received video over a range of distances.



(a) Initial activity (b) Video feed receiver

Figure 12. Flowcharts showing initial activity and video feed receiver.

(a) Gamepad buttons assignment (top view) (b) Gamepad buttons assignment (front view)

Figure 13. Gamepad top and front views.

We have found that the developed UGV outperforms ADORA UGV in terms of speed, which is depicted in Figure 17.b. The developed UGV has a speed of 3.3 m/s, while ADORA UGV has a speed of 0.61 m/s. However, ADORA UGV can handle around 1kg extra payload compared to the developed UGV, as illustrated in Figure 17.c. We look at this as a trade-off between speed and payload weight. Figure 17.a shows the weight of each UGV; a large difference in weight can be seen between the developed UGV and the other two UGVs. Subject to transmission delay constraints, we believe that the relatively light weight and high speed of the developed UGV give it high manoeuvrability.

We considered the effect of transmission delay on the quality of received video. Particularly, we studied TCP performance over a range of distances (5m – 70m) with increments of 5m. This range is the typical operational range of the developed UGV. In each experiment, we transferred a total of 5 MB from the UGV to the laptop at the operator's station and recorded the round trip time (RTT), average packet size and TCP throughput. The results are shown in Figure 17(d-f). The RTT was stable in the operation range with an average value around 50µs. The perceived quality of video over this range at this value was acceptable. The throughput and average packet size are depicted in Figure 17.e and Figure 17.f, respectively. The average packet size is around 1500 byte and the throughput is approximately 2Mbps. No packet loss or packet time-outs were reported. However, it is worth to mention that we invoked TCP-BBR [21] congestion control algorithm to manage the congestion window of TCP. TCP-BBR is optimized to achieve significant bandwidth improvements and lower latency.

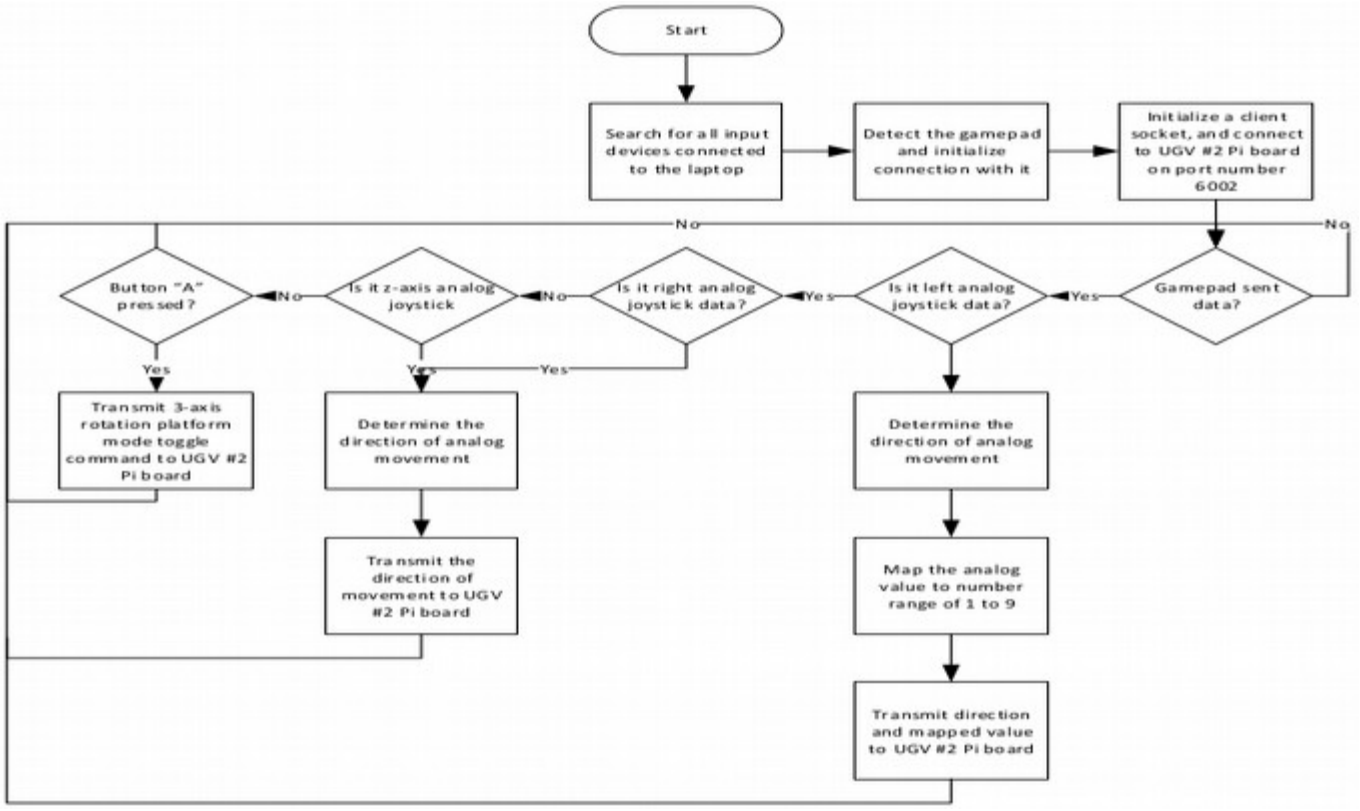
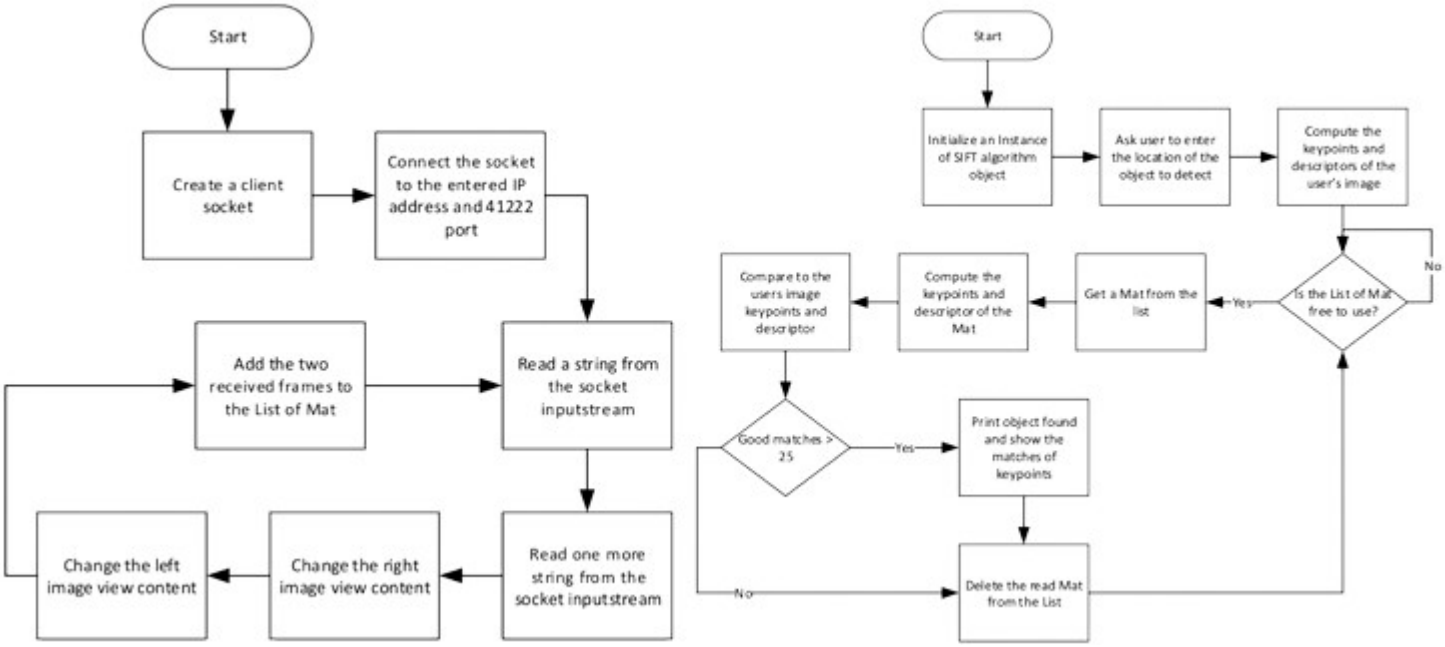


Figure 14. Flowchart of the gamepad software.

10. CONCLUSION & FUTURE WORK

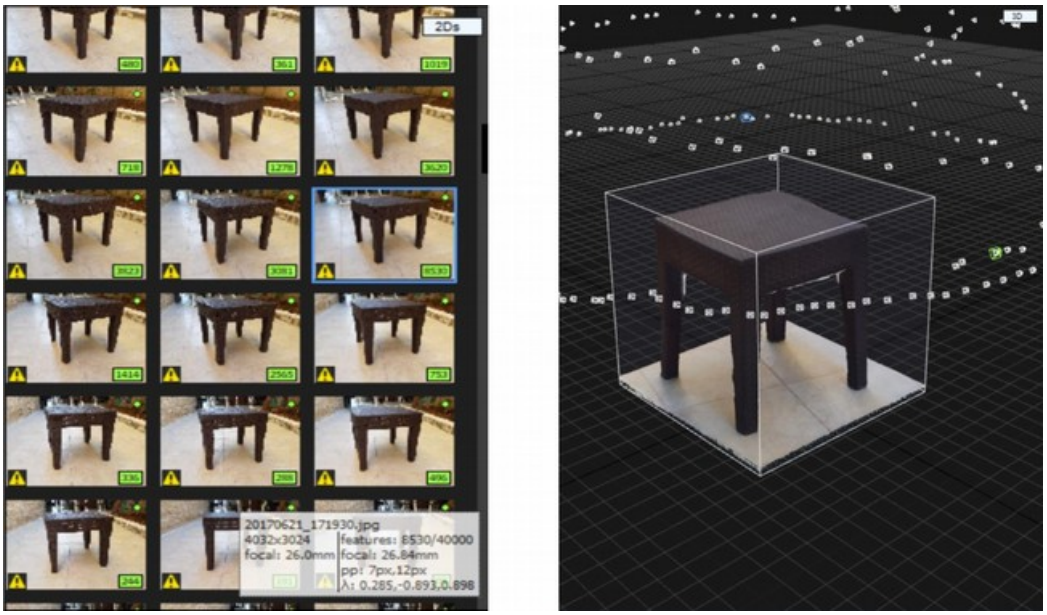
A SAR UGV with high visual inspection capabilities and relatively high speed was designed and implemented. The main contribution of the work dwells in the design of a separate user-interface unit constituting of a 3-axis rotating platform with three cameras: a stereoscopic camera and a Pi camera. The unit has two control modes, one is automatic through a smart phone-based VR-headset and the other is manual through a gamepad attached to the operator-station laptop. Video feeds from the unit are transmitted over a wireless link to both the VR-headset to provide visual feedback and to the operator-station laptop for: object detection – in the images captured by the stereoscopic camera – using the SIFT algorithm and 3-D model creation from 2-D images captured by the Pi camera.

Multiple computers can connect simultaneously to the unit to receive video feeds and since the wireless connection is realized by means of managed Wi-Fi, a team of operators can cooperate using a shared video conferencing session while viewing the video feed from the unit on their computers.



(a) Operator-station laptop video feed receiving and viewing, b) SIFT workflow thread

Figure 15. Flowcharts of Threads.

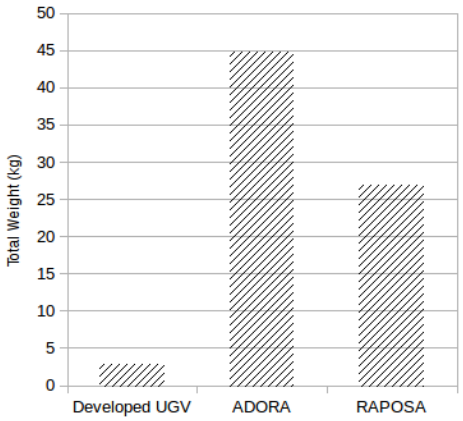


(a) 2-D pictures taken by the Pi camera (b) 3-D model created by the software

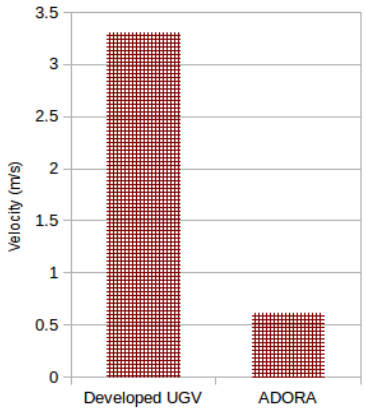
Figure 16. 3-D model creation.

A wheel-type rover was designed to hold the separate unit; the relatively high-speed of the rover (3.3 m/s) six-wheel differential-drive chassis and spiked air-filled rubber tires gave the rover high manoeuvrability in open rough terrain compared to other SAR UGVs found in literature.

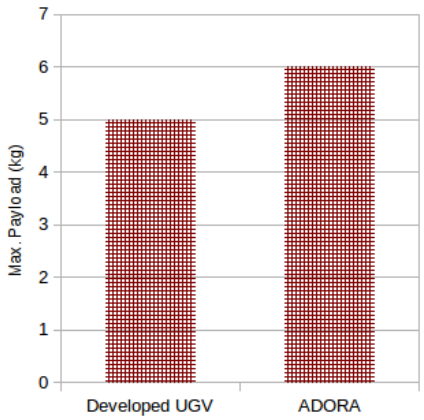
The UGV was designed and implemented in a laboratory environment. It was tested over soft and rough terrain and results showed its suitability for civil defence search and rescue applications, as well as for hazard areas (e.g. military) reconnaissance applications. Future work of this project goes in two directions: First, enhancing the night-vision capabilities of the UGV which can be done by connecting



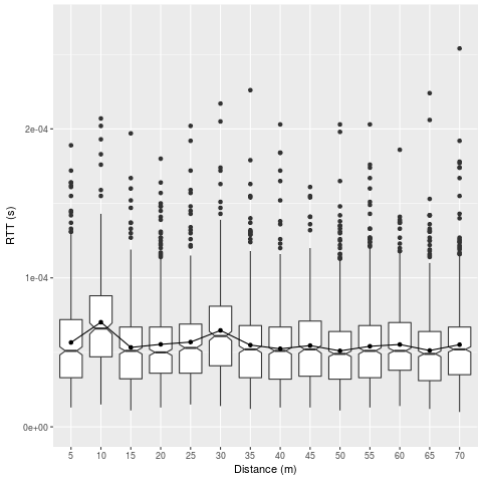
(a) Total Weight (kg)



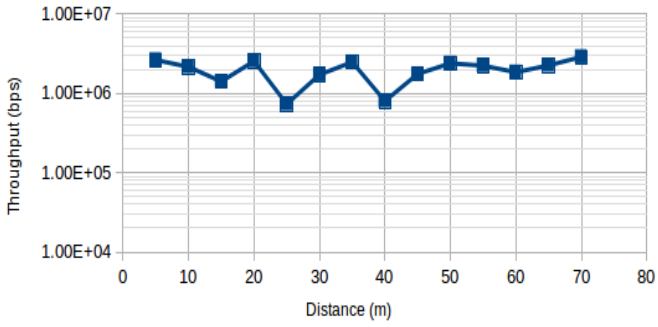
(b) Velocity (m/s)



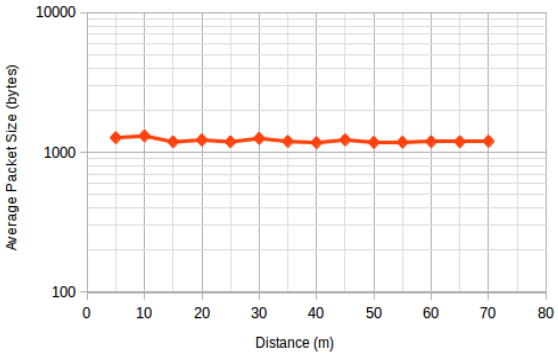
(c) Maximum Payload (kg)



(d) Round Trip Time (seconds)



(e) Throughput (bps)



(f) Average Packet Size (bytes)

Figure 17. Comparison of relevant UGV (a-c) and Performance measures of communication protocol (d-f).

an infra-red camera (Raspberry Pi Infra-red Night Vision Camera Module V2.1 IMX219 8-megapixel sensor) to #2 Pi using the on-board 15-pin CSI-2. Also, a LED torch can be mounted on top of the 3-axis platform which can be controlled by the operator through the gamepad in critical situations (e.g. in a dark pipe). Second, enhancing the exploratory capabilities of the UGV by implementing a general-purpose robotic arm. For instance, the arm can hold a special bucket to collect samples of soil, liquid, etc. Also, a separate sensor box can be added on the rover to sense temperature, humidity and gas. The data of the sensor box can be sent to a dedicated computer – other than the one used for displaying the video feed – for further analysis of the UGV's surrounding environment. Finally,

enhancing the processing power to make use of advanced software; for instance, using a dedicated graphical processing unit (GPU) such as for example NVidia Jetson development kit, which can increase the retrieved frame rate of camera, and allow to run software for real-time face recognition and autonomous driving.

REFERENCES

[1] K. Cavallin and P. Svensson, "Semi-autonomous, Teleoperated Search and Rescue Robot," 2009.

[2] J. L. Drury, J. Scholtz and H. A. Yanco, "Awareness in Human-robot Interactions," IEEE International Conference on Systems, Man and Cybernetics, vol. 1, pp. 912-918, Oct. 2003.

[3] N. Michael, S. Shen, K. Mohta, Y. Mulgaonkar, V. Kumar, K. Nagatani, Y. Okada, S. Kiribayashi, K. Otake, Yoshida et al., "Collaborative Mapping of an Earthquake-damaged Building via Ground and Aerial Robots," Journal of Field Robotics, vol. 29, no. 5, pp. 832-841, 2012.

[4] R. R. Murphy, J. Kravitz, S. L. Stover and R. Shoureshi, "Mobile Robots in Mine Rescue and Recovery," IEEE Robotics & Automation Magazine, vol. 16, no. 2, 2009.

[5] Y. Baudoin, D. Doroftei, G. De Cubber, S. A. Berrabah, C. Pinzon, F. Warlet, J. Gancet, E. Motard, M. Ilzkovitz, Nalpantidis et al., "View- finder: Robotics Assistance to Re-fighting Services and Crisis Management," IEEE International Workshop on Safety, Security & Rescue Robotics (SSRR), pp. 1-6, 2009.

[6] D. Doroftei, G. De Cubber, E. Colon and Y. Baudoin, "Behavior Based Control for an Outdoor Crisis Man-agement Robot," Proceedings of the IARP International Workshop on robotics for Risky Interventions and Environmental Surveillance, pp. 12-14, 2009.

[7] G. De Cubber, D. Doroftei, H. Sahli and Y. Baudoin, "Outdoor Terrain Traversability Analysis for Robot Navigation Using a Time-of- light Camera," Proc. RGB-D Workshop on 3D Perception in robotics, Vasteras, Sweden, 2011.

[8] C. Marques, J. Cristvo, P. Alvito, P. Lima, J. Frazo, I. Ribeiro and R. Ventura, "A Search and Rescue Robot with Teleoperated Tether Docking System," Industrial Robot: An International Journal, vol. 34, no. 4, pp. 332-338, 2007, [Online], Available: https://doi.org/10.1108/01439910710749663.

[9] N. Enayati and F. Najafi, "Design and Manufacturing of a Teleoperative Rescue UGV with a Novel Track Arrangement," Industrial Robot: An International Journal, vol. 38, no. 5, pp. 476-485, 2011, [Online], Available: https://doi.org/10.1108/01439911111154045.

[10] K. Berns, A. Nezhadfard, M. Tosa, H. Balta and G. De Cubber, "Unmanned Ground Robots for Rescue Tasks," Search and Rescue Robotics-From Theory to Practice, InTech, 2017.

[11] H. Martins and R. Ventura, "Immersive 3-D Teleoperation of a Search and Rescue Robot Using a Head-mounted Display," IEEE Conference on Emerging Technologies Factory Automation, pp. 1-8, Sept. 2009.

[12] "Pololu: Robotics & Electronics," [Online], Available: http://www.pololu.com/, 2017, accessed: 3-October-2017.

[13] M. Pedley, "Tilt Sensing Using a Three-axis Accelerometer," Freescale Semiconductor Application Note, vol. 1, pp. 2012-2013, 2013.

[14] D. G. Lowe, "Distinctive Image Features from Scale-invariant Key Points," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[15] C. M. Johnson and S.-J. Hsieh, "Board# 143: Maker: Urban Search and Rescue UGV: Visual localization and Navigation," ASEE Annual Conference & Exposition, 2017.

[16] Z. Wang and H. Gu, "A Review of Locomotion Mechanisms of Urban Search and Rescue Robot," Industrial Robot: An International Journal, vol. 34, no. 5, pp. 400-411, 2007, [Online], Available: https://doi.org/10.1108/01439910710774403.

[17] M. Neumann, T. Predki, L. Heckes and P. Labenda, "Snakelike, Tracked, Mobile Robot with Active Flippers for Urban Search and Rescue Tasks," Industrial Robot: An International Journal, vol. 40, no. 3, pp. 246-250, 2013, [Online], Available: https://doi.org/10.1108/01439911311309942.

[18] P. U. Lima, "Search and Rescue UGVs: The Civil Protection Teams of the Future," 3rd International Conference on Emerging Security Technologies, pp. 12-19, Sept. 2012.

" Unmanned Ground Vehicle with Virtual Reality Vision", M. Al-Khalil, R. Abu-Rhayyem, A. Hammoudeh and T. A. Edwan.

[19] M. W. Kadous, R. K.-M. Sheh and C. Sammut, "Caster: A Robot for Urban Search and Rescue," Proceedings of the 2005 Australasian Conference on Robotics and Automation, pp. 1-10, 2005.

[20] A. Kleiner, Mapping and Exploration for Search and Rescue with Humans and Mobile Robots, University of Freiburg, 2009.

[21] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh and V. Jacobson, "BBR: Congestion-Based Congestion Control," ACM Queue, vol. 14, pp. 20-53, 2016.

[22] Y. Wang, Z. Song, Y. Zhou and C. Zhou, "Design of A Tele-Operated Field Robot," Defense Science Research Conference and Expo (DSR), pp. 1-4, Aug. 2011.

[23] D. Sanders "Comparing Speed to Complete Progressively More Difficult Mobile Robot Paths Between Human Tele-operators and Humans," Assembly Automation, vol. 29, pp. 230-248, 2017.

[24] D. Sanders, "Analysis of the Effects of Time Delays on the Teleoperation of a Mobile Robot in Various Modes of Operation," Industrial Robot: An International Journal, vol. 36, pp. 570-584, 2009.

ملخص البحث:

تصف هذه الورقة البحثية تصميم و تنفيذ مركبة أرضية غير مأهولة ووحدة إظهار هاتف ذكي تركب على الرأس تسمح بادراك الوضع البصري، بحيث يتم إرسال صور الفيديو إلى حاسوب منفصل لكشف الأشياء و إنشاء نموذج ثلاثي الأبعاد لمحيط المركبة من الصور ثنائية الأبعاد للأشياء المحيطة بها.

وقد تم تحقيق التصميم المطلوب باستخدام وحدة إظهار متصلة بهاتف ذكي تلتقط حركات الرأس في الزمن الحقيقي وترسلها من ثم إلى ثلاثة محركات مركبة على عربة لتوفير الحركة حول ثلاثة محاور، بينما يعمل المشغل على التحكم بتلك المحركات إما عبر وحدة إظهار متصلة بهاتف ذكي أو عن طريق لوحة ألعاب. و تعمل ثلاث كاميرات محمولة على متن المركبة على التقاط الصور التي يتم إرسالها إلى وحدة الإظهار المتصلة بالهاتف الذكي و إلى حاسوب المشغل. و تقوم برمجية بتحويل الصور ثنائية الأبعاد الملتقطة لمحيط المركبة إلى نموذج ثلاثي الأبعاد.

لقد تم تنفيذ المركبة في بيئة مخبرية ومن ثم فحصها. وبينت النتائج التي جرى الحصول عليها أن المركبة المقترحة تمتلك خصائص أفضل للتفتيش البصري مقارنة بتصاميم أخرى لمركبات مماثلة. و قد منحتها تلك الخصائص – و أبرزها سرعة قصوى مقدارها 3.3 م/ث، و هيكل دفع تفاضلي بست عجلات، و إطارات مطاطية مملوءة بالهواء – قدرات عالية على المناورة، مما يجعلها خيارا ملائما لاستكشاف الكواكب و الأغراض العسكرية. و يمكن تركيب المحركات و الكاميرات بسهولة على وحدة منفصلة تتضمن هيكلا يستخدم آليات قطر مختلفة لزيادة فعالية عمل المركبة. و يمكن للتصميم المقترح أن يستخدم في بناء الخرائط، وإنشاء النماذج ثلاثية الأبعاد، وكشف الأشياء في الزمن الحقيقي، والتعرف على الوجوه في الزمن الحقيقي، والقيادة الذاتية.

EDITORIAL BOARD SUPPORT TEAM

LANGUAGE EDITOR

Haydar Al-Momani

EDITORIAL BOARD SECRETARY

Eyad Al-Kouz

JJCIT ADDRESS

WEBSITE: www.jjcit.org

EMAIL: jjcit@psut.edu.jo

ADDRESS: Princess Sumaya University for Technology, Khalil Saket Street, Al-Jubaiha.

B.O. BOX: 1438 Amman 11941 Jordan.

TELEPHONE: +962-6-5359949.

FAX: +962-6-7295534.

المجلة الأردنية للحاسوب و تكنولوجيا المعلومات

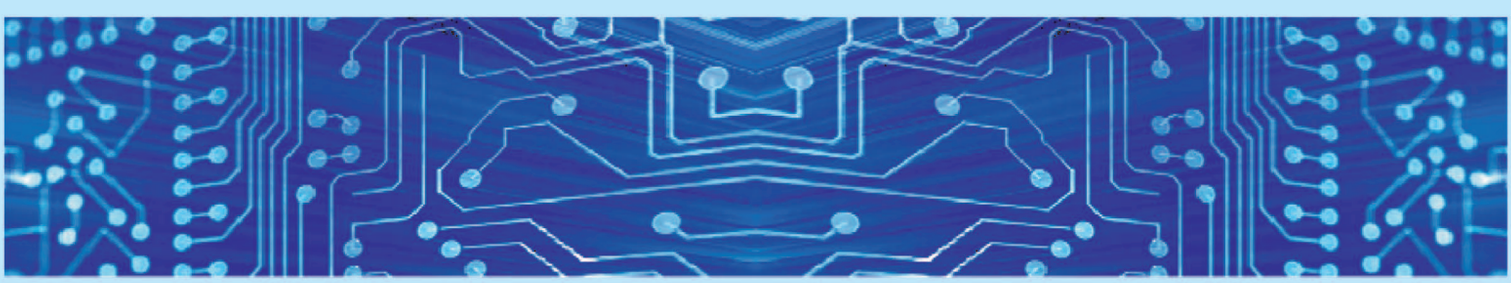
ISSN 2415 - 1076 (Online)
ISSN 2413 - 9351 (Print)

نيسان ٢٠١٨ المجلد ٤ العدد ١

JJCIT

www.jjcit.org

jjcit@psut.edu.jo



مجلة علمية عالمية متخصصة محكمة

تصدر بدعم من صندوق دعم البحث العلمي