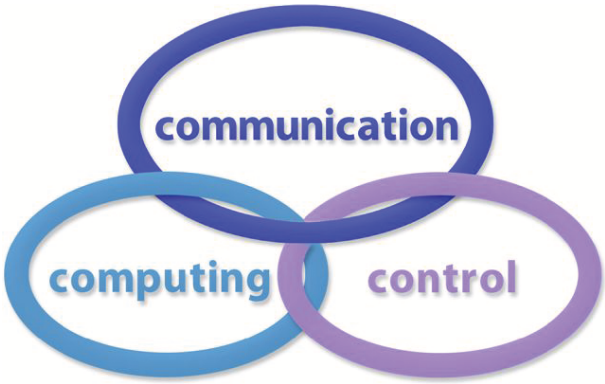
INTERNATIONAL JOURNAL

of

COMPUTERS COMMUNICATIONS & CONTROL

ISSN 1841-9836



A Bimonthly Journal
With Emphasis on the Integration of Three Technologies

Year: 2015 Volume: 10 Issue: 3 (June)

This journal is a member of, and subscribes to the principles of, the Committee on Publication Ethics (COPE).

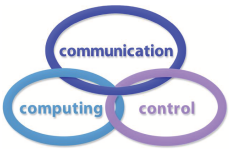
CCC Publications - Agora University Editing House

CCC Publications

http://univagora.ro/jour/index.php/ijccc/

BRIEF DESCRIPTION OF JOURNAL

Publication Name: International Journal of Computers Communications & Control.
Acronym: IJCCC; **Starting year of IJCCC:** 2006.
Abbreviated Journal Title in JCR: INT J COMPUT COMMUN.
International Standard Serial Number: ISSN 1841-9836.
Publisher: CCC Publications - Agora University of Oradea.
Publication frequency: Bimonthly: Issue 1 (February); Issue 2 (April); Issue 3 (June); Issue 4 (August); Issue 5 (October); Issue 6 (December).
Founders of IJCCC: Ioan DZITAC, Florin Gheorghe FILIP and Mişu-Jan MANOLESCU.
Logo:



Indexing/Coverage:

1. Since 2006, Vol. 1 (S), IJCCC is covered by Thomson Reuters and is indexed in ISI Web of Science/Knowledge: Science Citation Index Expanded.
2. Journal Citation Reports (JCR - Science Edition), IF = 0.694 (JCR2013).
   Subject Category:
   (a) Automation & Control Systems: Q4 (46 of 59);
   (b) Computer Science, Information Systems: Q3 (96 of 135).
3. Since 2008, 3(1), IJCCC is covered in Scopus, SJR2013 = 0.231, H index = 13.
   Subject Category:
   (a) Computational Theory and Mathematics: Q4;
   (b) Computer Networks and Communications: Q3;
   (c) Computer Science Applications: Q3.
4. Since 2007, 2(1), IJCCC is covered in EBSCO.

Focus & Scope: International Journal of Computers Communications & Control is directed to the international communities of scientific researchers in computer and control from the universities, research units and industry.

To differentiate from other similar journals, the editorial policy of IJCCC encourages the submission of original scientific papers that focus on the integration of the 3 "C" (Computing, Communication, Control).

In particular the following topics are expected to be addressed by authors:

1. Integrated solutions in computer-based control and communications;
2. Computational intelligence methods (with emphasis on fuzzy logic-based methods, ANN, evolutionary computing, collective/swarm intelligence);
3. Advanced decision support systems (with emphasis on the usage of combined solvers and/or web technologies).

Editorial Address:
Agora University/ R&D Agora Ltd. / S.C. Cercetare Dezvoltare Agora S.R.L.
Piata Tineretului 8, Oradea, jud. Bihor, Romania, Zip Code 410526
Tel./ Fax: +40 359101032
E-mail: ijccc@univagora.ro, rd.agora@univagora.ro, ccc.journal@gmail.com
Journal website: http://univagora.ro/jour/index.php/ijccc/

Gheorghe Păun
Institute of Math. of Romanian Academy
Bucharest, PO Box 1-764, Romania
gpaun@us.es

Mario de J. Pérez Jiménez
Dept. of CS and Artificial Intelligence
University of Seville, Sevilla,
Avda. Reina Mercedes s/n, 41012, Spain
marper@us.es

Dana Petcu
Computer Science Department
Western University of Timisoara
V.Parvan 4, 300223 Timisoara, Romania
petcu@info.uvt.ro

Radu Popescu-Zeletin
Fraunhofer Institute for Open
Communication Systems
Technical University Berlin, Germany
rpz@cs.tu-berlin.de

Imre J. Rudas
Óbuda University
Budapest, Hungary
rudas@bmf.hu

Yong Shi
School of Management
Chinese Academy of Sciences
Beijing 100190, China &
University of Nebraska at Omaha
Omaha, NE 68182, U.S.A.
yshi@gucas.ac.cn, yshi@unomaha.edu

Athanasios D. Styliadis
University of Kavala
Institute of Technology
65404 Kavala, Greece
styliadis@teikav.edu.gr

Gheorghe Tecuci
Learning Agents Center
George Mason University
U.S.A.
University Drive 4440, Fairfax VA
tecuci@gmu.edu

Horia-Nicolai Teodorescu
Faculty of Electronics and
Telecommunications
Technical University "Gh. Asachi" Iasi
Iasi, Bd. Carol I 11, 700506, Romania
hteodor@etc.tuiasi.ro

Dan Tufiş
Research Institute for Artificial Intelligence
of the Romanian Academy
Bucharest, "13 Septembrie" 13, 050711,
Romania
tufis@racai.ro

Lotfi A. Zadeh
Director,
Berkeley Initiative in Soft Computing (BISC)
Computer Science Division
University of California Berkeley,
Berkeley, CA 94720-1776
U.S.A.
zadeh@eecs.berkeley.edu

DATA FOR SUBSCRIBERS
Supplier: Cercetare Dezvoltare Agora Srl (Research & Development Agora Ltd.)
Fiscal code: 24747462
Headquarter: Oradea, Piata Tineretului Nr.8, Bihor, Romania, Zip code 410526
Bank: BANCA COMERCIALA FEROVIARA S.A. ORADEA
Bank address: P-ta Unirii Nr. 8, Oradea, Bihor, România
IBAN Account for EURO: RO50BFER248000014038EU01
SWIFT CODE (eq.BIC): BFER

International Journal of Computers Communications & Control
ISSN 1841-9836, Volume 10, Issue 3, June, 2015.

Contents

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):298-307, June, 2015.

Threshold Based Best Custodian Routing Protocol for Delay Tolerant Network

Q. Ayub, M. S. Mohd Zahid, S. Rashid, A. Hanan Abdullah

Qaisar Ayub*, Soperi Mohd Zahid,
Sulma Rashid, Abdul Hanan Abdullah
Faculty of Computing
Universiti Teknologi Malaysia
UTM Skudai, 81310 Johor,Malaysia
sheikhqaisar@gmail.com, soperi@utm.my,
sulmaqaiser@gmail.com, hanan@utm.my
*Corresponding author:sheikhqaisar@gmail.com

Abstract: Delay Tolerant Network (DTN) is a kind of network in which the source may not be able to establish the stable and uninterrupted path to destination due to network partitioning, dynamic topology change and frequent disconnections. In order to dealt disruption and disconnections a store, carry and forward paradigm is used in which node stores the incoming messages in its buffer, carries it while moving and forward when comes within the transmission range of other nodes. Message forwarding contributes and important role in increasing its delivery. For instance, probabilistic routing protocol forwards message to a node having high probability value to meet message destination. These protocols cannot handle a situation in which the node continually transmits messages even the probability difference is very small. In this paper, we have proposed a routing protocol known as Threshold Based best custodian Routing Protocol (TBbcRP) for delay tolerant network. We have proposed a threshold-based method to compute the quality value which is the ability of node to carry message. A self-learning mechanism has been used to remove the delivered messages from the network. Moreover, a buffer aware mechanism has been used that make sure availability of buffer space at receiver before message transmission. We have compared the performance of TBbcRP with Epidemic, PRoPHET and Delegated Forwarding. The proposed TBbcRP outperforms in terms of maximizing the delivery probability, reducing number of transmissions and message drop.
Keywords: Delay Tolerance Network , store-carry-forward, routing protocols , algorithms.

1 Introduction

With advancement in communication technologies [1-2] it is now possible to interconnect mobile nodes, stand-alone computers and provide an innovative way to join the social and business communities. Despite, other communication architectures such as LAN, WLAN, the mobile ad hoc networks have gained more popularity. In ad hoc networking routing protocols [3-6], the source and destination establishes the end-to-end path prior to the transmission of data. This prerequisite is impossible in highly disrupted wireless applications such as wildlife monitoring, deep-space communication and military networks. Such environments suffer frequent disconnections, dynamic teleology change and network partitioning due to node mobility. In addition, limited network resource, for instance, buffer space, bandwidth, energy and processing power of nodes makes routing a real challenge.

Delay Tolerance Network (DTN)[7] is a kind of network that aims to provide the communication via opportunistically connected mobile nodes. A novel method called as the store, carry and forward is used in which nodes stores the message in their buffers carries them while moving and forwards when connected to other nodes. The DTN routing protocols can be classified as

single copy and multi copy. In single copy protocols, the unique copy of the message exists in entire network [8,21]. These protocols are capable to operate under limited resource but reduce delivery ratio and raises the delivery delay. Multi copy routing protocols transmits the redundant copies of each message to all connected nodes [9-13]. Therefore, the message can reach to its destination via multiple intermediate nodes. As a result, multi copy routing protocols minimize the delivery delay and maximize the delivery [14-18].

The multi copy routing protocols are more robust, but unreliable due to consumption of high volume of network resources. The probabilistic routing protocols were proposed to reduce the resource consumption that considers node behavior such as its movement pattern, encounter history [12,14,27] before the transmission of the message. A carrier node with probabilistic protocol continues to forward message to high probable relay nodes. This issue was addressed in [19] where Vijay Erramilli et al. proposed a new message forwarding technique called as Delegated Forwarding. In this method, each node maintains a quality metric and forward the message to another node only if it has high quality metric which have been seen by the message. Later, in [20][23-24] the authors present the variations of Delegated Forwarding. The previous works has not defined a method to compute quality value of nodes.

The contribution to this paper is as follows

- We have proposed a routing protocol called as Threshold Based best custodian Routing Protocol for Delay Tolerant Network (TBbcRP).
- We have used a self-learning method to remove the previously delivered messages from the network.
- A threshold-based method has been used to assign the quality value to network nodes.
- We have compared the performance of TBbcRP with Epidemic, PRoPHET and Delegated Forwarding in terms of minimizing number of transmissions, number of drops, overhead while raising the deliver probability.

1.1 Review of DTN routing protocols

In dynamic surroundings like intentions topology change, node mobility and frequent network partitions, problem is to select a suitable relay node for a message. In addition, scarce network resources such as limited buffer space, bandwidth, and low power of nodes makes data routing even more challenging. Therefore, prototype of a good routing protocol must focus on minimizing the consumption of network resources and delivery of more messages to their destinations. In flooding based routing protocols such as Epidemic protocol [13] each node encounter results in the exchange of messages. The encountering nodes further diffuse the message copy and process continues.

Despite the fact that in Epidemic protocol, each message may have more than one path to reach destination, Epidemic protocol consumes high volume of network resources. The control on creation of message copies can reduce the resource requirement. In this background, qouta based routing protocols were emerged where each node was given the opportunity to transmit the n number of message copies for example Spray and Wait[10], Spray and Focus[11], Spray and Wait binary QoN Spray and Wait[26].

The Spray and Wait algorithm consist of a Spray phase, where node spread n message copies to its neighbors called as relays. If the destination is not found in the Spray phase then each node Wait until contacted to message destination. In Spray and Wait protocol, message transmission was limited only to the neighboring nodes. This problem was solved in binary Spray and Wait protocol in which a source node on encountering forwards the $n/2$ message copies to the connected

node while keeps $n/2$. In addition, the receiver node was also privileged to distribute $n/2$ message copies. This hierarchical forwarding improves the performance of Spray and Wait and increases message delivery. The spraying protocols work well when the node movement is Independent and Identically Distributed (IID) which is not possible for real world scenarios where each node exhibits its own movement pattern.

These challenges motivate the researchers and various utility functions were introduced in spraying algorithms. For example, Spray and Focus [11] protocol starts by distributing the $n/2$ message copies like Spray and Wait binary, however when the node left only one copy of message then it shifts to the Focus phase where the nodes forwards the message to neighbors by observing its suitability to meet the destination. The suitability is determined by the time since two nodes last saw each other. Quality of node Spray and Wait [26] improves the performance of Binary Spray and Wait algorithm by introducing QoN (Quality of node). The QoN is represented by an integer number which describes encountering frequency of one node to encounter with other node in a given time interval. The primary objective of Quota based routing protocols was to control the transmission of message copies.

Despite the fact that spraying algorithms has exploited the encountering history of nodes, other factor such as mobility patterns or positional coordinates can improve the routing procedure. For example In [28], the author designs a mobility based spraying strategy Most Social First, Most Mobile First (MMF), Last Seen First (LSF). The Shen Ling et [22] observe the node mobility and introduce a influence factor which is determined by the mobility of the nodes.

In [12] Lindgren et al. introduces PRoPHET protocol which reduces the number of message transmissions by introducing a new metric called as delivery predictability and transitive connectivity. The nodes are capable receive a message only if they constitute a high value of predictability and transitivity. As expected, PRoPHET protocol compared to Epidemic protocol minimizes the number of transmissions.

The variation of PRoPHET such as PROCS [29] introduces a new message forwarding method which observes movement pattern of nodes and their time sequence In [19] Vijay Erramilli et al. propose a new message forwarding technique called as Delegated Forwarding. In Delegated Forwarding, each node maintains a quality metric and forwards the message to another node only if it has high quality metric which have been seen by the message. The protocol works well to and controls the transmission of the message on the relay nodes. However, it does not provide any solution to control the transmission of the source messages. In [20] author modified the algorithm by updating source message with the probability value of high quality node. In this way, the replication was also controlled from the source messages. In [23] the author defines the Delegated Forwarding by designing cost-based drop and transmission methods. According to this, the message that is close to their destinations are assigned the high priorities by defining a replication count number called as the Delegated number. Yunsheng Wang et.al [30] proposed for single copy multicast, multi copy multicast and Delegated Forwarding multicast algorithms.

1.2 The proposed threshold based best custodian routing protocol

In proposed TBbcRP, each network node maintains the time information about encountering to other nodes in a vector called as Previous Encounter Vector (PEV). The PEV consists of node id *nid* and Encounter Time (ET). We have used ET to assign the quality value which describes the future encountering likelihood among same nodes. The high quality value indicates that node is more likely to delver message. The quality information is stored in the Quality Vector. The Quality Vector consists of node id and Quality Value (QV).

Principal-1. When nodes contact, they assigns quality values by using upstream and down-

stream time threshold:

$$TD = CurrentTime - ET(ni). \quad (1)$$

When the node encounters first time, they initializes the default quality in the Quality Vector and Current Time in the ET of PEV. If nodes have encountered previously, then the Time Difference (TD) is computed by subtracting the Current Time from the ET by using Equation 1 and Threshold Streams module is invoked. The Threshold Stream module updates the quality value for a node by mapping the TD in the pre-defined collection of threshold queue. Table 1 shows the meaning of variables used in TBbcRP.

Table 1: Meaning of variables used in TBbcRP

Symbol	Description
n_i , n_j	node i and node j
PEV,ET	Previous Encounter Vector, Encounter Time
QV	Quality Value
TD	Time Difference
QP	Quality Points
Vd	Vector Deliver

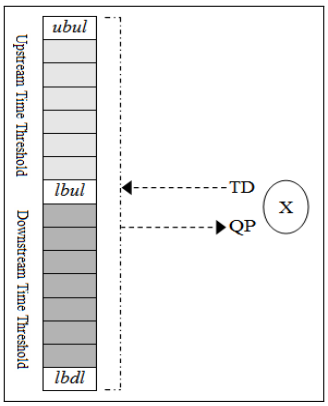


Figure 1: Structure of Threshold Queue

Figure 1 show the structure of threshold queue which is divided into Upstream Time Thresholds and Downstream Time Thresholds. The Upstream Time Thresholds further defines its Lower Bound of Upstream Limit (lbul) and Upper Bound of Upstream Limit (ubul). The Upstream Thresholds are used to decrement quality value while Downstream Thresholds are used to increase quality values. The nodes encountering after large interval of time shows high TD value and relevant Quality Points (QP) are subtracted from quality value of node. When the time difference is above the ubul then quality value of node is initialized to zero. The Downstream Threshold starts after lbul and defines its Lower Bound of Downstream Limit (lbdl). The Downstream Threshold is used to increment the quality value of nodes. For instance, nodes encountering after small interval of time are expected to encounter again. The TD is mapped and relevant Quality Points (QP) are incremented in the quality value. When time difference is lower than the lbdl then maximum Quality Points (QP) are assigned to the Quality Value (QV).

Figure 2 shows the algorithmic flow of threshold based method in which node X and node Y have established connectivity. The node X and Y has maintained Previous Encounter Vector

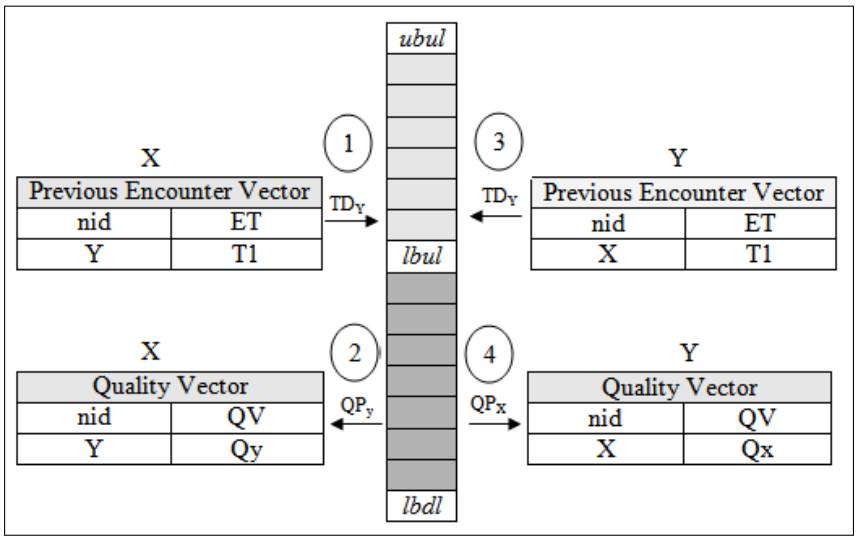


Figure 2: Threshold Based best custodian Routing Protocol Example

(PEV) and Quality Vector. In step one, X map TD_Y which represents time elapsed since X has seen Y. In step two, the relevant quality values QV_Y will be given to X. The same steps are followed by node Y.

1.3 Self learning method to remove delivered messages

Since, DTN is a highly disrupted environment where it is not possible to keep the track of the transmitted messages via central administration. Hence, most of the time the message even after finding their destinations cannot convey their delivery status to the other nodes, and message replication continues even it is delivered. When the network resources scarce then replications of delivered messages produce high overhead on the buffer space, bandwidth and energy. Despite the influence of other factors, these messages also produce the congestion that a node overcomes by dropping its stored messages. Hence, a solution is required to remove the delivered message from network. Inspired by immunity based routing protocol, we have defined a de-centralized mechanism to remove the previously delivered messages.

Principal-2. The algorithm states that when a node deliverers a message to the current connection as a final recipient then it stores the message id in Vector Delivered (Vd) and remove it from the buffer.

The TBbcRP , each network node maintains a vector called Vector Delivered (Vd). When a node forwards the message copy to a connection as final recipients it inserts the message id in Vd and removes the message from the list of its carried messages. This module is invoked after the threshold computation and before the transmission of messages.

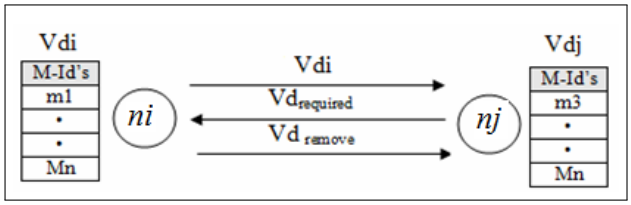


Figure 3: Exchange of previously delivered messages

Figure 3 show the technical flow of removing delivered messages. Accordingly, on encounter-

ing, ni forward Vdi [16-17] to nj. Vdi hold ids of delivered messages known by ni. nj subtracts Vdi from Vdj to get $\text{Vd}_{required}$ that holds list of delivered messages not known by nj and send it to ni by using Eq. (2).

$$Vd_{required} = (Vdi - Vdj) \tag{2}$$

ni computes Vd_{remove} by intersecting $\text{Vd}_{required}$ from Vdi and send it to nj by using Eq. (3).

$$Vd_{remove} = (Vd_{required} \cap Vdi) \tag{3}$$

Finally, nj removes the Vd_{remove} messages from buffer and update Vdj by using Eq. (4).

$$Vdj = Vd_{remove} \cup Vdj \tag{4}$$

1.4 Control on message replication by buffer space

The DTN message consists of message header and payload header. The payload header contains the actual contents of message. The message header is collection of control information such as message identification, hop count, and time to live. The DTN node utilizes control information to forward and drop messages. In TBbcRP, we have included a new data field in message header called as Recent Quality (RQ). The RQ is an integer value initialized with zero for the messages generated by the source.

Principal-3 When a transmitter forward the message copy it updates the RQ of message with the of quality value of receiver, while the receiver will update the RQ of message to its own.

The principal 3 is about the implementation of Delegated Forwarding in message header. We have used the same concept in DF++[24]. Briefly, after transmitting message copy, the sender node updates the RQ of message header to quality value of node which receives message. Similarly, receiving node update the message RQ to its own quality value. The idea is that the receiver or transmitter will not replicate the message until the encountered node has higher quality value than the RQ.

Principal-4. The transmitter will forward the message only if the QV of receiver to meet with the message destination is greater then the RQ and available buffer space at receiver is capable to store the message.

The high quality nodes are likely to encounter the message destination. However, if we forward a message to a congested high quality node then this forwarding decision may degrade the network performance. Since, the congested node will drop its previously stored messages to accommodate the new one. In our previous work, DF++ [24] and CFBARP [25] an adaptive mechanism has been defined to dealt with buffer space. Hence, the node forward the message only if the quality value of receiver is high as well as available buffer is able to accommodate the incoming message. After transmission, the transmitter will subtract the message size from available buffer.

2 Simulation and Results

This section provides the performance analysis of existing and proposed routing protocols in terms of minimizing the message transmissions, message drop and raising the delivery probability by ONE simulator [31]. ONE is event driven simulator written in java and has been designed to evaluate the DTN applications. The reality of simulation has bee increased by using a city based environment which consist on pedestrian, cars and Trains. The pedestrians were divided into two groups with 40 members at each group. The pedestrian are moving with shortest path map based movement model at the speed between 0.5km/h and 1.5 km/h. Each pedestrian has been

carrying mobiles with 2MB of buffer size. The transmission range of mobile nodes is 10 meters. The 40 cars are moving via map route movement at the speed between 10km/h-50km/h. Finally, six trains are moving via map route movement at the speed between 7km/h and10km/h. The random size message generated from the sample of 100K-300K and the inter message creation interval is 25s-35s. The bandwidth if equally distributed at 2MBPS.

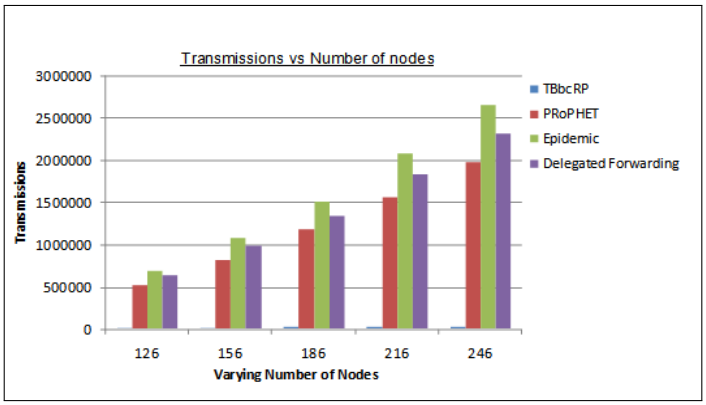


Figure 4: Transmissions by varying number of nodes

Figure 4 represents the results of exiting PRoPHET, Epidemic and Delegated Forwarding as compared to proposed TBbcRP routing protocol in terms of number of message transmissions. The flooding based Epidemic protocol has been showing high number of transmissions. The Delegated Forwarding has controlled the message diffusion as compared to Epidemic protocol but still forwards high quantity of messages compare to PRoPHET and TBbcRP protocols. The message transmissions are getting higher with increasing number of nodes. The reason is that, at high number of nodes, message exchange gets higher. It is possible to sustain such traffic under the infinite buffer space. However, in the current environment the buffer is limited resource thus a better quality node when receive a message by having no space mechanically triggers the drop event. Further, due to the multi copy of each message the same high quantity node may reputedly receive the dropped messages, thus cause high transmissions, message drop and waste of node energy. The proposed TBbcRP routing protocol has reduced the message transmissions.

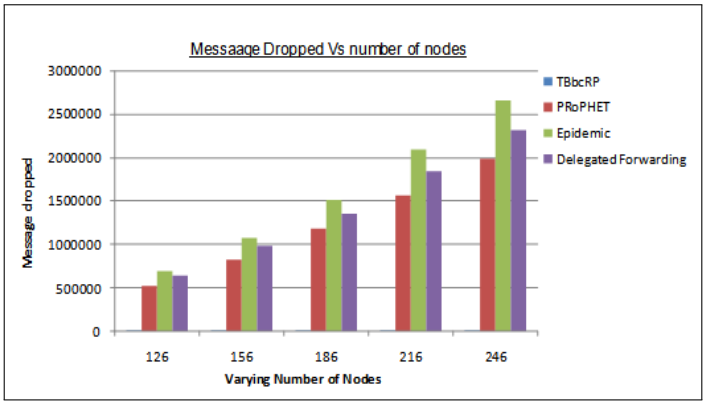


Figure 5: Message dropped by varying number of nodes

The Figure 5 depicts the results message drop by increasing the number of nodes. We can see that increasing the number of nodes has raised the message dropped. This is because the buffer space is finite, and nodes cannot accommodate all incoming messages. For instance, at increasing

number of nodes such as 186, 216, and 246, even the protocols like PRoPHET and Delegated Forwarding has dropped large number of messages. The reason is that when the encounter rate among nodes is high then multiple nodes became highly probable to receive the message. The proposed TBbcRP has shown the constant stance for all network traffic.

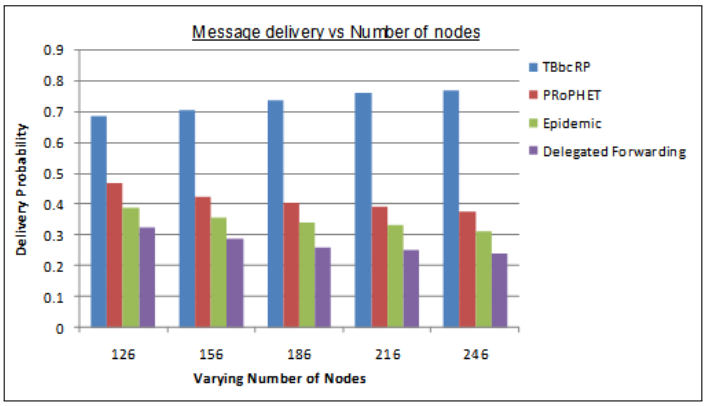


Figure 6: Delivery by varying number of nodes

The Figure 6 plot the results of existing and proposed routing protocols in terms of message delivery probability by increasing number of nodes. It can be observed that at less number of nodes such as 126, 156 the protocols such as PRoPHET, Epidemic and Delegated Forwarding has delivered more messages. Nevertheless, as the number of nodes gets higher like 186,216 and 246 less number of message find their destinations. The reason is that messages were dropped before reaching destination.

3 Conclusion

In this paper we have proposed a routing protocol called as Threshold Based best custodian Routing Protocol (TBbcRP) for delay tolerant network. A threshold based method has been proposed to compute the quality value of nodes, which is the ability of nodes to carry message. Moreover, a self learning method has been used to remove previously delivered messages from network. The proposed protocol out performs well as compared to existing strategies in terms of maximizing the delivery probability, reducing number of transmissions and message drop.

Acknowledgments

This work is financed by institution scholarship provided by UTM and Ministry of Higher Education of Malaysia.

Bibliography

[1] Ariyavisitakul, S. L. (2000); Turbo Space-Time Processing to Improve Wireless Channel Capacity. *IEEE Transactions on Communications*. 48(8): 1347-1359.

[2] Yujin Lim.,Jesung Kim.,Sang Lyul Min.,Joong Soo Ma(2001); Performance evaluation of the Bluetooth-based public Internet access point Information Networking, 2001. *Proceedings.15th International Conference on Information Networking*, 643-648.

[3] Latiff L.A.,Fisal, N.,. (2003); Routing protocols in wireless mobile ad hoc network - a review. *The 9th Asia-Pacific Conference on Communications (APCC 2003)*, 600-604.

[4] Murthy S., and J.J. Garcia-Luna-Aceves.(1996); An efficient routing protocol for wireless networks, *Mobile Networks and Applications*, 1(2):183-197.

[5]] Perkins C.E., and E.M. Royer. (1999); Ad-hoc on-demand distance vector routing, *IEEE WMCSA 99*, 90-100

[6] Johnson D.B., and D.A. Maltz.(1996); Dynamic source outing in ad hoc wireless networks, *Mobile computing*, 153-181.

[7] Fall, K. (2003); A Delay-Tolerant Network Architecture for Challenged Internets. *In SIGCOMM 03: Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. New York, NY, USA: ACM, 27-34.

[8] Spyropoulos T., K. Psounis., and C.S Raghavendra.(2004); Single-copy routing in intermittently connected mobile networks. *in Proc. IEEE Conf. Sensor and Ad Hoc Communications and Networks (SECON)*, 235-244.

[9] Ramanathan, R., Hansen, R., Basu, P., Rosales-Hain, R. and Krishnan, R. (2007); Prioritized Epidemic Routing for Opportunistic Networks. *In Proc. of the 1st International MobiSys Workshop on Mobile Opportunistic Networking*, ACM, 62-66.

[10] Spyropoulos, T., Psounis, K. and Raghavendra, C. (2005); Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks. *In Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking*, ACM, 252-259.

[11] Spyropoulos, T., Psounis, K. and Raghavendra, C. S. (2007); Spray and Focus:Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility. *In Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 07*, IEEE, 79-85.

[12] Lindgren, A., Doria, A. and Schelen, O. (2004); Probabilistic Routing in Intermittently Connected Networks. *In Service Assurance with Partial and Intermittent Resources*. Springer, 239-254.

[13] Vahdat, A., Becker, D. et al. (2000); Epidemic Routing for Partially Connected Ad hoc Networks. Technical report. Technical Report CS-200006, Duke University.

[14] de Oliveira, E. C. and de Albuquerque, C. V. (2009); NECTAR: A DTN Routing Protocol Based on Neighborhood Contact History. *In Proceedings of the 2009 ACM symposium on Applied Computing*. ACM, 40-46.

[15] Bulut, E., Geyik, S. C. and Szymanski, B. K. (2010); Conditional Shortest Path Routing in Delay Tolerant Networks. *In IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM)*. IEEE, 1-6.

[16] Srinivasa, S. and Krishnamurthy, S. (2009); CREST: An Opportunistic Forwarding Protocol Based on Conditional Residual Time. *In 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON09*, IEEE, 1-9.

[17] Hua, D., Du, X., Qian, Y. and Yan, S. (2009); A DTN Routing Protocol Based on Hierarchy Forwarding and Cluster Control.*In International Conference on Computational Intelligence and Security, CIS09*, IEEE, 2: 397-401.

[18] Wang, G., Wang, B. and Gao, Y. (2010); Dynamic Spray and Wait Routing Algorithm with Quality of Node in Delay Tolerant Network. *In International Conference on Communications and Mobile Computing (CMC)*, IEEE, 3: 452-456.

[19] Erramilli, V., et al. (2008); Delegation forwarding. *ACM MOBIHOC08*, 251-260.

[20] Chen, X., et al.(2009); Probability delegation forwarding in delay tolerant networks. *IEEE ICCCN09*, 1-6.

[21] T. Spyropoulos, K. Psounis, and C. S. Raghavendra (2007); Utility-based Message Replication for Intermittently Connected Heterogeneous Wireless Networks, *in Proc. of IEEE WoWMoM workshop on Autonomic and Opportunistic Communications (AOC)*, (INRIA Technical Report RR-6129), June 2007.

[22] Ling, S. and Wei, W. (2009); Feedback Adaptive Routing Algorithm for DTN. *In WRI International Conference on Communications and Mobile Computing, CMC09*, IEEE, 2: 267-271.

[23] Liu, C. and J. Wu (2009); An optimal probabilistic forwarding protocolin delay tolerant networks. *ACM MobiHoc 09*, 105-114.

[24] Ayub, Q., Zahid, M. S. M., Rashid, S., & Abdullah, A. H. (2013); DF++: An adaptive buffer-aware probabilistic delegation forwarding protocol for Delay Tolerant Network. *Cluster Computing*, 1-8.

[25] Qaisar Ayub; M soperi Mohd Zahid; Abdul Hanan Abdullah; Sulma Rashid (2013); Connection frequency buffer aware routing protocol for delay tolerant network. *Journal of Electrical Engineering and Technology*, 8(3): 649-657.

[26] Wang, G., Wang, B. and Gao, Y. (2010); Dynamic Spray and Wait Routing Algorithm with Quality of Node in Delay Tolerant Network. *In International Conference on Communications and Mobile Computing (CMC)*, IEEE, 3: 452-456.

[27] Nelson, S. C., Bakht, M. and Kravets, R. (2009); Encounter-Based Routing in DTNs. *In IEEE INFOCOM 2009*, IEEE, 846-854.

[28] Elwhishi, Ahmed, Pin-Han Ho, Sagar Naik, and Basem Shihada (2011); Contention aware routing for intermittently connected mobile networks, *In AFIN 2011, the third international conference on advances in future internet*, 8-15.

[29] Jathar, R. and Gupta, A. (2010); Probabilistic Routing using Contact Sequencing in Delay Tolerant Networks. *In 2010 Second International Conference on Communication Systems and Networks (COMSNETS)*, IEEE, 1-10.

[30] Wang, Y., Li, X., & Wu, J. (2010); Multicasting in delay tolerant networks: delegation forwarding. *In Global Telecommunications Conference (GLOBECOM 2010)*, IEEE, 1-5.

[31] Keranen, A., Ott, J. and Karkkainen, T. (2009); The ONE Simulator for DTN Protocol Evaluation. *Proc. of the 2nd International Conference on Simulation Tools and Techniques. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)*, 1-10.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):308-317, June, 2015.

On Observer Synchronization of Non-identical Discrete-time Hyperchaotic Maps Using Arrow Form Matrix

R.L. Filali, M. Benrejeb, P. Borne

Rania Linda Filali*, Mohamed Benrejeb, Pierre Borne
LARA. Ecole Nationale d'Ingnieurs de Tunis (ENIT)
BP 37, Le Belvdre, 1002 Tunis, Tunisia
&
LAGIS. Ecole Centrale de Lille
Cit Scientifique BP 48, F 59651, Villeneuve d'Ascq Cedex, Lille, France
rania.filali@ec-lille.fr, mohamed.benrejeb@enit.rnu.tn, pierre.borne@ec-lille.fr
*Corresponding author:rania.filali@ec-lille.fr

Abstract: In this paper, new sufficient conditions for synchronization of non-identical discrete-time hyperchaotic maps is proposed for hyperchaotic cryptosystem communication. They use aggregation techniques for stability study associated to the Benrejeb arrow form matrix for system description. In addition, suitable choice of outputs feedback brings the problem of synchronization of two non-identical hyperchaotic maps to two identical hyperchaotic maps one. The considered case of synchronization of third order hyperchaotic Hénon-Baier Klein maps shows the efficiency of the proposed approach to recover secure transmission of an image and a text.
Keywords: synchronization, discrete-time hyperchaotic maps, aggregation technique, arrow form matrix, secure communication.

1 Introduction

During the last decades, the problem of the synchronization of chaotic and hyperchaotic maps has gained a significant attention due to its potential applications, especially in the light of its application in secure communication area. Synchronization phenomena of chaotic systems makes the trajectories of the master system and the slave ones achieving synchronism after a transition time, starting from different initial conditions. Many effective methods have already been successfully applied to the problem since Pecora and Carrolls research works [1]. Recently, with the development of nonlinear control theory, various synchronization schemes have been proposed such as: adaptive control [3], observer-based control [13, 15], backstepping control [4], active control [5] or nonlinear control [14]. However, most of the methods, mentioned above, are designed to synchronize two identical chaotic systems. In this paper, the proposed synchronization of two non-identical discrete-time chaotic systems which is well adapted with the secure communication is based on establishing new output feedback stabilizing conditions. With the use of the Borne and Gentina practical criterion for stability study [7, 8] associated to the Benrejeb arrow form matrix for system description [6, 10, 11], this approach constitutes an extension of previous results on synchronization studies of identical discrete-time chaotic processes [14, 15]. In Section 2, sufficient conditions leading to conclude to the synchronization of non-identical discrete-time hyperchaotic maps are given for a secure communication scheme combining conventional cryptographic methods and synchronization of discrete-time hyperchaotic systems. In section 3, the proposed design of a complete synchronous output feedback stabilizing controller of two non-identical Baier-Klein map-Hénon maps is applied with success for secure transmission of signals as an image and a text.

2 Design of observer synchronization for non-identical chaotic maps using single channel transmission

In this section, the discrete-time hyperchaotic secure communication, called hyperchaotic cryptosystem, figure 1, shows the efficiency of the proposed non-identical hyperchaotic approach synchronization. This cryptosystem is a combination of a classical cryptographic technique and of observer-based synchronization, allowing the receiver to recover the information transmission without considering noise such that: $y_m(k) = y_s(k)$.

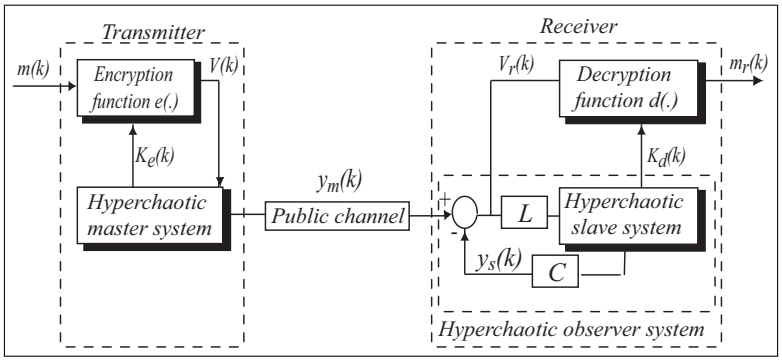


Figure 1: Block diagram of hyperchaotic communication based on cryptography

The goal is to synchronize, at the same time, a master and a slave maps by designing a suitable output control law.
Consider an n-dimensional master hyperchaotic system, described as follows:

$$\begin{aligned} x_m(k+1) &= Ax_m(k) + f(x_m(k)) + E + \delta NV(k) \\ y_m(k) &= Cx_m(k) + \delta V(k) \end{aligned} \tag{1}$$

$N = [n_1 \ \dots \ n_q]^T$ is a constant vector characterizing the way to mix the ciphertext $V(k)$ with the chaotic signal $x_m(k)$, δ a scaling factor chosen to allow the term $\delta NV(k)$ to belong to a compatible range with respect to the minimum and the maximum bounds of states variables of master and slave chaotic signals [19]
Let consider a slave system in the receiver side described as:

$$\begin{aligned} x_s(k+1) &= A_1 x_s(k) + f_1(x_s(k)) + E_1 + Bu(k) \\ y_s(k) &= Cx_s(k) \end{aligned} \tag{2}$$

with, $x_m(k) = \begin{bmatrix} x_{m1}(k) & \dots & x_{mn}(k) \end{bmatrix}^T \in R^n$, $x_s(k) = \begin{bmatrix} x_{s1}(k) & \dots & x_{sn}(k) \end{bmatrix}^T \in R^n$,
E and E_1 are constant vectors of systems (1) and (2), respectively,

$$B = I_{n\times n} \tag{3}$$

$(x_m(0), x_s(0)) = ((1, 0.1, 0), (-0.5, 0, 0.3))$ initial conditions, $A = \{a_{ij}\}$ and $A_1 = \{a_{1_{ij}}\}$ constant matrices and $f(x_m(k))$ and $f_1(x_s(k))$ nonlinear vectors.
The key $K_e(k)$ in the transmitter side is defined by:

$$K_e(k) = \text{int}\left(A\sqrt{x_{m1}^4 + \cdots + x_{mn}^4}\right) \bmod 256 \tag{4}$$

and the key $K_d(k)$ in the receiver side as following:

$$K_d(k) = \text{int}\left(A\sqrt{x_{s1}^4 + \cdots + x_{sn}^4}\right) \bmod 256 \tag{5}$$

The function $int(x)$ gives the integer part of x.
The used encryption $e(.)$ is an XOR algorithm:

$$V(k) = e(m(k), K_e(k)) = m(k) \oplus K_e(k) \tag{6}$$

and the decryption function $d(.)$ is as following:

$$m_r(k) = d(V(k), K_d(k)) = V(k) \oplus K_d(k) \tag{7}$$

$m_r(k)$ is the recovered encrypted signal, $d(.)$ the corresponding decryption function. If the chaotic systems of the receiver and transmitter are synchronized, the receiver side can find the same $K_d(k)$, as in the encrypter, $K_e(k)$. The goal now is to synchronize master and slave systems and, at the same time, to design a Luenberger-like discrete observer depending on gain based on a suitable choice of the control law $u(k)$ designed to make two non-identical chaotic systems achieving the complete state synchronization.
The evolution of the error vector $e(k)$ between the master and the slave systems:

$$e(k) = x_s(k) - x_m(k) \tag{8}$$

can also be described by the following:

$$e(k+1) = A_1 x_s(k) + f_1(x_s(k)) - Ax_m(k) + f(x_m(k)) + E_1 - E + Bu(k) + N\delta V(k) \tag{9}$$

The following vector controllers $u(k)$, retained in this case, is by outputs feedback control design and by the introduction of an additional compensation term:

$$u(k) = (A - A_1) x_s(k) + f(x_s(k)) - f_1(x_s(k)) - E_1 + E + L(y_m(k) - y_s(k)) \tag{10}$$

where:

- $L(.) = \{l_i(.)\} \in R^n$ is an unknown vector gain of observer to be determined,
- $u(k)$ is chosen as (10) to bring the synchronization study of non-identical discrete-time hyperchaotic systems (1) and (2) to synchronization study of two identical hyperchaotic systems, described as (1).

The substitution of (10) into (9) yields the error system as following:

$$e(k+1) = Ae(k) + f(x_s(k)) - f(x_m(k)) + N\delta V(k) - BL(Cx_m(k) + \delta V(k) - Cx_s(k)) \tag{11}$$

$$e(k+1) = (A - BLC)e(k) + \delta V(k)(N - BL) + f(x_m(k)) - f(x_s(k)) \tag{12}$$

If the considered

$$N = BL \tag{13}$$

is satisfied, it comes:

$$e(k+1) = (A - BLC)e(k) + f(x_m(k)) - f(x_s(k)) \tag{14}$$

Once can observe that by the choice of the output control law $u(k)$ as (10), it brings the study of synchronizing of two non-identical hyperchaotic systems to the study of two identical ones.
For several chaotic systems, $f(x_s(k)) - f(x_m(k))$ can be, as shown in [18], factorized as following:

$$f(x_s(k)) - f(x_m(k)) = Q(x_m(k), x_s(k))\, e(k) \tag{15}$$

where matrix $Q\left(x_s\left(k\right), x_m\left(k\right)\right)$ is a bounded matrix whose elements depend on $x_m\left(k\right)$ and $x_s\left(k\right)$. Then, the error system can be rewritten as:

$$e\left(k+1\right) = A_a\left(x_m\left(k\right), x_s\left(k\right)\right) e\left(k\right) \tag{16}$$

with:

$$A_a\left(x_m\left(k\right), x_s\left(k\right)\right) = \left(A + Q\left(x_s\left(k\right), x_m\left(k\right)\right) - BLC\right) \tag{17}$$

The error process, described by (16), is stabilized by the a choice of suitable control law of (10), making the matrix $A_a\left(x_m\left(k\right), x_s\left(k\right)\right)$, defined by (17), in the arrow form. For this purpose, the following theorem can be established, based on the use of Borne and Gentina criterion [7, 8] associated to the canonical Benrejeb arrow form matrix $A_a\left(.\right) = \left\{a_{a_{ij}}\left(.\right)\right\} = \left(A + Q(x_s\left(k\right), x_m\left(k\right)) - BLC\right)$ [6, 10], such that:

$$A_a\left(.\right) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{n1} & & & a_{nn} \end{bmatrix} \tag{18}$$

gives sufficient conditions of synchronization of slave (1) with master (2) systems [9,12]. Theorem *Theorem* 2.1. The synchronization error, described by (16) converges towards zero, if the matrix $A_a\left(.\right)$, defined by (17) and (18), is in the arrow form such that:

1. the nonlinear elements are isolated in one row of the matrix $A_a\left(.\right)$;
2. the diagonal elements, $a_{a_{ii}}\left(.\right)$ of the matrix $A_a\left(.\right)$ are such that:
$$1 - \left|a_{a_{ii}}\left(.\right)\right| > 0,\ \forall i = 2, \ldots, n \tag{19}$$
3. there exist $\varepsilon > 0$ such that:
$$1 - \left|a_{a_{11}}\left(.\right)\right| - \sum_{i=2}^{n} \left(\begin{array}{l} \left|a_{a_{i1}}\left(.\right) a_{a_{1i}}\left(.\right)\right| \\ \times \left(1 - \left|a_{a_{ii}}\left(.\right)\right|\right)^{-1} \end{array} \right) > \varepsilon \tag{20}$$

Proof: The comparison system [8] of the error system (17), associated to the vectorial norm $p(z\left(k\right)) = [|z_1\left(k\right)| \ldots |z_n\left(k\right)|]^T$, $z\left(k\right) = [z_1\left(k\right) \ldots z_n\left(k\right)]^T$, is defined by the following equation:

$$z\left(k+1\right) = M\left(A_a\left(.\right)\right) z\left(k\right) \tag{21}$$

with $M\left(.\right) = \left\{m_{ij}\left(.\right)\right\}$ such that $m_{ij}\left(.\right) = \left|a_{a_{ij}}\left(.\right)\right| \forall i, j = 1, 2, \ldots, n$.
The error system (16) is stabilized by the output feedback law (10) if we make an appropriate choice of vectors gains L and C such as the matrix $\left(I - M\left(A_a\left(.\right)\right)\right)$ is an M matrix [16] i.e if, by application of the practical stability criterion of Borne and Gentina [7, 8], we have:

$$\begin{cases} 1 - \left|a_{a_{ii}}\left(.\right)\right)| > 0,\ \forall i = 2, \ldots, n \\ det(I - M(A_a\left(.\right))) > \varepsilon \end{cases} \tag{22}$$

The computation of the first member of the last inequality, led as following:

$$det(I - M(A_a\left(.\right))) = \left(\begin{array}{l} 1 - \left|a_{a_{11}}\left(.\right)\right| \\ -\sum_{i=2}^{n} \left(\begin{array}{l} \left|a_{a_{i1}}\left(.\right) a_{a_{1i}}\left(.\right)\right| \\ \times \left(1 - \left|a_{a_{ii}}\left(.\right)\right|\right)^{-1} \end{array} \right) \end{array} \right) \times \left(\prod_{j=2}^{n} \left(1 - \left|a_{a_{jj}}\left(.\right)\right|\right) \right) \tag{23}$$

achieves easily the proof of the theorem. □

By the design of suitable output feedback and the above theorem, synchronisation of non-identical hyperchaotic maps is satisfied.

3 Synchronization of two non-identical hyperchaotic 3D generalized Hénon map and 3D Baier-Klein map

In this section, a proposed synchronization approach for a class of two non-identical discrete-time hyperchaotic systems is applied for the case of Baier-Klein/ Hénon.
Consider the following 3D discrete-time Baier-Klein map [17] which can be described as:

$$\begin{cases} x_{m1}(k+1) = b - x_{m2}^2(k) - a.x_{m3}(k) + \delta l_1 V(k) \\ x_{m2}(k+1) = x_{m1}(k) + \delta l_2 V(k) \\ x_{m3}(k+1) = x_{m2}(k) + \delta l_3 V(k) \\ y_m(k) = c_1 x_{m1}(k) + c_2 x_{m2}(k) + c_3 x_{m3}(k) \end{cases} \tag{24}$$

For the parameters a and b such that $a = 0.1$ and $b = 1.76$, the system (24) is hyperchaotic as shown in figure 2.
In state space, (24) becomes as (1), with:

$$A = \begin{bmatrix} 0 & 0 & -0.1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{25}$$

$$f(x_m(k)) = \begin{bmatrix} -x_{m2}^2(k) & 0 & 0 \end{bmatrix}^T \tag{26}$$

and:

$$E = \begin{bmatrix} 1.76 & 0 & 0 \end{bmatrix}^T \tag{27}$$

Consider the following 3D generalized Hénon map [2] as the slave system:

$$\begin{cases} x_{s1}(k+1) = -b_1 x_{s2}(k) + u_1(k) \\ x_{s2}(k+1) = 1 + x_{s3}(k) - a_1 x_{s2}^2(k) + u_2(k) \\ x_{s3}(k+1) = b_1 x_{s2}(k) + x_{s1}(k) + u_3(k) \\ y_s(k) = c_1 x_{s1}(k) + c_2 x_{s2}(k) + c_3 x_{s3}(k) \end{cases} \tag{28}$$

The 3D generalized Hénon map (28), for the parameters $a_1 = 1.07$ and $b_1 = 0.3$, exhibits a hyperchaotic attractor as shown in figure 3.
In state space, (28) becomes as (2), with:

$$A_1(x_s(k)) = \begin{bmatrix} 0 & -0.3 & 0 \\ 0 & 0 & 1 \\ 1 & 0.3 & 0 \end{bmatrix} \tag{29}$$

$$f_1(x_s(k)) = \begin{bmatrix} 0 & -1.07x_{s2}^2(k) & 0 \end{bmatrix}^T \tag{30}$$

and:

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \tag{31}$$

The figure 4 shows the error states between systems (24) and (28), with the following initial conditions $(x_m(0), x_s(0)) = ((1, 0.1, 0), (0.1, 0.2, -0.31))$, when the control law is turned off. It is obvious that the error grows chaotically with time.

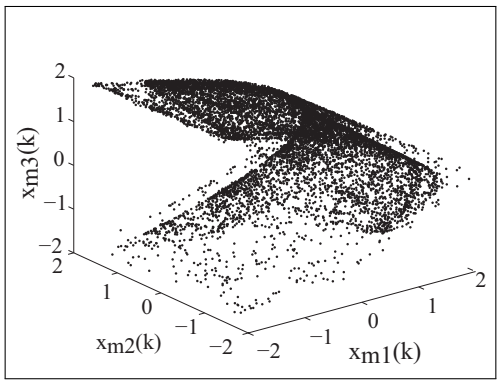


Figure 2: 3D discrete-time Baier-Klein

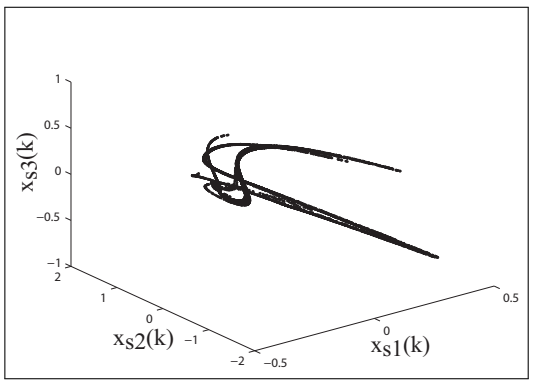


Figure 3: 3D generalized Hénon map

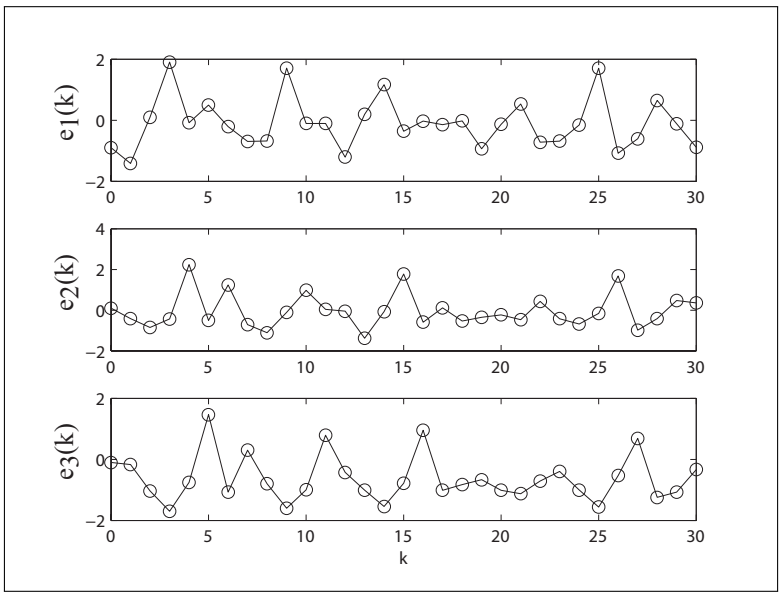


Figure 4: Error dynamics for hyperchaotic Hénon and Baier-Klein maps

Let consider the synchronization error between systems (24) and (28) described by:

$$e_i(k) = x_{si}(k) - x_{mi}(k),\ i = 1, 2, 3 \tag{32}$$

Applying (10), $u(k)$ can be written as:

$$\begin{cases} u_1(k) = 0.3x_{s2}(k) - 0.1x_{s3}(k) + 1.76 - x_{s2}^2(k) - \sum_{j=1}^{3} l_1 c_j e_j(k) \\ u_2(k) = x_{s1}(k) - x_{s3}(k) - 1 + 1.07x_{s2}^2 - \sum_{j=1}^{3} l_2 c_j e_j(k) \\ u_3(k) = -x_{s1}(k) + 0.7x_{s2}(k) - \sum_{j=1}^{3} l_3 c_j e_j(k) \end{cases} \tag{33}$$

The choice of the control law (33), transforms the synchronization of the two non-identical 3D hyperchaotic Baier-Klein and 3D generalized Hénon maps problem to two identical 3D hyperchaotic Baier-Klein hyperchaotic maps one.

In the state space, the error system is as (16) and (17), with:

$$Q(x_m(k), x_s(k)) = \begin{bmatrix} 0 & -(x_{s2}(k) + x_{m2}(k)) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{34}$$

$$B = I_{3\times 3} \tag{35}$$

$$C = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix} \tag{36}$$

and:

$$L = \begin{bmatrix} l_1 & l_2 & l_3 \end{bmatrix}^T \tag{37}$$

The matrix, $A_a(x(k))$ can be rewritten as:

$$A_a(.) = \begin{bmatrix} -l_1(.)c_1(.) & -(l_1(.)c_2(.) + x_{m2}(k) + x_{s2}(k)) & -0.1 - l_1(.)c_3(.) \\ 1 - l_2(.)c_1(.) & -l_2(.)c_2(.) & -l_2(.)c_3(.) \\ -l_3(.)c_1(.) & 1 - l_3(.)c_2(.) & -l_3(.)c_3(.) \end{bmatrix} \tag{38}$$

The choice of correction parameters c_3 and l_3 constant and satisfying the constrains:

$$\begin{cases} l_2 c_3 = 0 \\ 1 - l_3 c_2 = 0 \end{cases} \tag{39}$$

makes this matrix $A_a(x(k))$ in Benrejeb arrow form, such that:

$$A_a(.) = \begin{bmatrix} -l_1(.)c_1(.) & -(l_1(.)c_2(.) + x_{m2}(k) + x_{s2}(k)) & -0.1 \\ 1 - l_2(.)c_1(.) & -l_2(.)c_2(.) & 0 \\ -l_3(.)c_1(.) & 0 & 0 \end{bmatrix} \tag{40}$$

The system characterized by (16) is asymptotically stable, if the control gains l_i and c_j, $i, j = 1, 2, 3$, are chosen so that the following conditions are satisfied:

1. the nonlinear elements are isolated in one row of the matrix $A_a(x(k))$;
2. the diagonal element of the matrix $A_a(x(k))$ is such that:

$$1 - |l_2 c_2| > 0 \tag{41}$$

3. there exist $\varepsilon > 0$ such that:

$$1 - |l_1 c_1| - \frac{(x_{m2}(k) + x_{s2}(k) + |l_1 c_2|)(|1 - l_2 c_1|)}{1 - |l_2 c_2|} - \frac{0.1 c_1}{c_2} \geq \varepsilon \tag{42}$$

Boundedness property

The solutions $x_m(k)$ respectively $x_s(k)$ of the hyperchaotic attractor of systems (24) and (28) are forward completely and uniformly bounded.

Boundedness is a common assumption for many physical systems such as oscillators and, more particularly, chaotic oscillators. The boundedness property of (24) and (28) allows us to perform the following transformation : $|x_{mi}| < 2$ and $|x_{si}| < 2$; thus, we have: $| x_{m2} + x_{s2} + l_1 c_2 | < 4 + |$

$l_1 c_2$ |
The condition (iii) of the above theorem can be rewritten as follows:

$$1 - |l_1 c_1| - \frac{(4 + |l_1 c_2|)\,(|1 - l_2 c_1|)}{1 - |l_2 c_2|} - \left|\frac{0.1 c_1}{c_2}\right| > 0 \tag{43}$$

Then, instantaneous gains l_i and c_j, $\forall i, j = 1, 2, 3$, satisfying inequalities (41) and (43) such as:

$$C = [1.73 \quad 0.91 \quad 0] \tag{44}$$

$$L = [-0.20 \quad 0.55 \quad 1.10]^T \tag{45}$$

guaranty the synchronization of systems (24) and (28), as shown in figures 5 and 6.

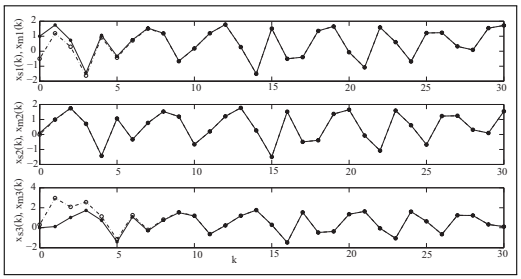


Figure 5: Time responses of master (Baier-Klein map) (—) and slave (Henon map)(−−) outputs.

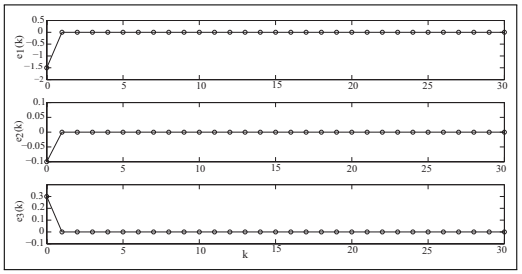


Figure 6: Error dynamics of the two diffrents hyperchaotic maps for activated controller

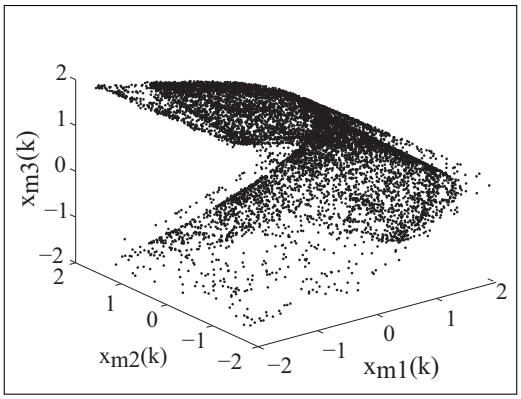


Figure 7: 3D discrete-time Baier-Klein

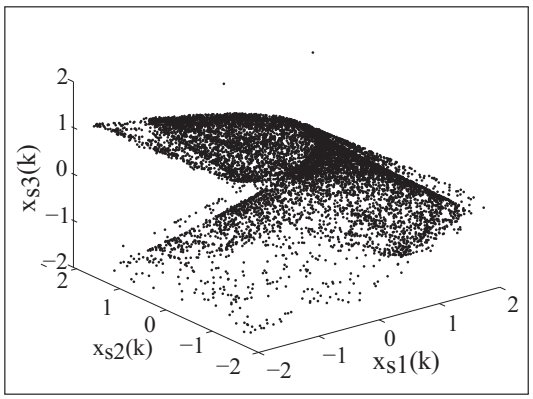


Figure 8: 3D generalized Hénon map

The transmitted messages used in this paper, to show the efficiency of the proposed output control law, are exemples of a text and of a photography of Lena.

The figures 7 and 8 show that the synchronization between two discrete-time hyperchaotic systems is achieved via an output feedback law using the proposed approach. One can see that the generalized Hénon system is controlled to become Baier-Klein system. We can conclude that the proposed approach is applied with success to recover messages, as it is shown in figures 9 and 10.

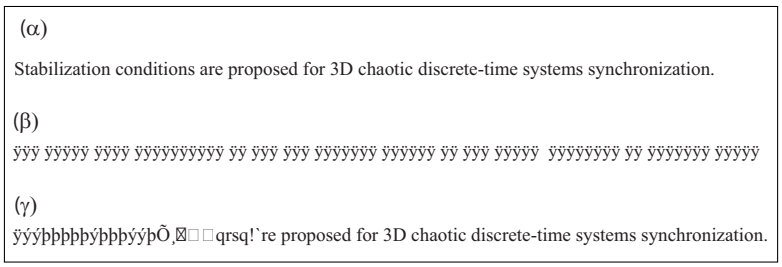


Figure 9: A single channel transmission: (α) original text $m(k)$, (β) crypted text $\delta^{-1}y_m(k)$, (γ) recovered text $m_r(k)$

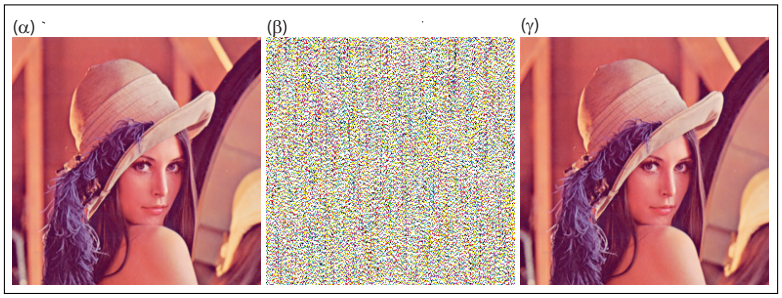


Figure 10: A single channel transmission: (α) original Lena photo $m(k)$, (β) crypted Lena photo $\delta^{-1}y_m(k)$, (γ) recovered Lena photo $m_r(k)$

4 Conclusion

In this paper, new suitable output feedback stabilizing conditions is proposed between two non-identical hyperchaotic discrete-time systems, for the case of image and text transmission. Numerical simulations illustrate the efficiency of the above stabilization conditions for the exemple of 3D generalized hyperchaotic Hénon and 3D Baier-Klein maps bringed in two identical Baier-Klein ones.

Bibliography

[1] Pecora, L.M.; Carrol, l, T.L. (1990); Synchronization in chaotic systems, *Phys. Rev. Lett.*, 64(8) : 821-824.

[2] Yan, Z. (2006); Q-S (complete or anticipated) synchronization backstepping scheme in a class of discrete-time chaotic (hyperchaotic) systems: A symbolic-numeric computation approach , *An Interdisciplinary Journal of Nonlinear Science*, ID 013119, 16(1).

[3] Fotsin, H.B.; Daafouz, J.(2004); Adaptive synchronization of uncertain chaotic Colpitts oscillators based on parameter identification, *Physics Letters, Section A: General, Atomic and Solid State Physics*, 339(3-5):304-315.

[4] Park, J.H.(2006); Synchronization of Genesio chaotic system via backstepping approach, *Chaos, Solitons and Fractals*,27(5): 967-977.

[5] Yassen, M.T.(2005); Chaos synchronization between two different chaotic systems using active control, *Chaos, Solitons and Fractals*, 23(12) : 131-140.

[6] Borne, P.; Benrejeb, M.(2008); On the representation and the stability study of large scale systems, *International Journal of Computers Communications & Control*, 3(5): 55-66.

[7] Borne, P.(1987); Non linear systems stability: vector norm approach, *Systems and Control Encyclopedia, Pergamon Press*, T.5: 3402-3406.

[8] Borne, P.; Gentina, J.C.; Laurent,F.(1976); Stability study of large scale non linear discrete systems by use of vector norms, *IFAC Symposium on Large Scale Systems Theory and Applications* Udine: 187-193.

[9] Benrejeb, M.(2010); Stability Study of Two Level Hierarchical Nonlinear Systems, *Plenary lecture, 12th IFAC Large Scale Systems Symposium: Theory and Applications, IFAC-LSS*, Lille.

[10] Benrejeb, M.; Hammami, S.(2008); New approach of stabilization of nonlinear continuous monovariable processes characterized by an arrow form matrix, *First International Conference, Systems ENgineering Design and Applications, SENDA*, Monastir.

[11] Benrejeb, M.; Soudani, D.; Sakly, A.; Borne, P.(2006); New discrete Tanaka Sugeno Kang fuzzy systems characterization and stability domain, *International Journal of Computers Communications & Control*, 1(4): 9-19.

[12] Hammami, S.; Ben Saad, K.; Benrejeb, M.(2009); On the synchronization of identical and non-identical 4-D chaotic systems using arrow form matrix, *Chaos, Solitons & Fractals*, 42(1): 101-112.

[13] Filali, R.L.; Benrejeb, M.; Borne, P.(2014); On observer-based secure communication design using discrete-time hyperchaotic systems, *Commun. Nonlinear Sci. Numer. Simulat.*, 9(5): 1424-1432.

[14] Filali, R.L.; Hammami, S.; Benrejeb, M.; Borne, P.(2012); On Synchronization, Anti-synchronization and Hybrid Synchronization of 3D Discrete Generalized Hénon Map, *Nonlinear Dynamics and Systems Theory*, 12(1): 81-95.

[15] Filali, R.L.; Hammami, S.; Benrejeb, M.; Borne, P.(2012); Synchronization of discrete-time hyperchaotic maps based on aggregation technique for encryption, *Systems, Signals and Devices (SSD), 2012 9th International Multi-Conference, Chemnitz* : 1-6.

[16] Robert, F.(1964); Normes vectorielles de vecteurs et de matrices, *RFTI Chiffres*, 17(4): 61-299.

[17] Baier, G.; Klein, M.(1990); Maximum hyperchaos in generalized Hénon circuit, *Phys. Lett. A.*, 151(67): 281-284.

[18] Jiang, G.P.; Tang, W.K.S.; Chen, G.(2003); A simple global synchronization criterion for coupled chaotic systems, *Chaos, Solitons and Fractals*, 15(5): 925-935.

[19] Millerioux, G.; Daafouz, J.(2003); An Observer-Based Approach for Input-Independent Global Chaos Synchronization of Discrete-Time Switched Systems, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 50(10) : 1270-1279.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):318-332, June, 2015.

Analysis of Reconfigurability, Control and Resource Management in Heterogeneous Wireless Networks

L. Gavrilovska, V. Atanasovski, P. Latkoski, V. Rakovic

Liljanava Gavrilovska, Vladimir Atanasovski, Pero Latkoski*, Valentin Rakovic
Faculty of Electrical Engineering and Information Technologies,
Ss. Cyril and Methodius University Skopje, Macedonia
liljana, vladimir, pero, valentin@feit.ukim.edu.mk
*Corresponding author: pero@feit.ukim.edu.mk

Abstract: Modern communications networks integrate different access technologies that require interoperability for seamless and user-transparent transfer of multimedia-reach content. Latest standardization activities in this area pinpoint the IEEE 802.21 standard as an enabler of media independent handovers in various scenarios. Additionally, the implementation of the heterogeneous network paradigm yields optimized and efficient resource management techniques emphasizing the need for reconfiguration and interoperability capabilities within future wireless networks. This paper analyzes a combination of reconfigurability, interoperability and resource management aspects in heterogeneous wireless networks based on the IEEE 802.21 standard. It introduces a novel platform for wireless heterogeneous communication systems and a prototype of a reconfigurable mobile terminal that rely on the IEEE 802.21 standard. The introduced platforms are extensively validated through simulations and laboratory experiments showcasing that the IEEE 802.21-backed interoperability is able to support uninterrupted content delivery across multiple communication technologies with high performance.
Keywords: Heterogeneous Networks; Resource Management; IEEE 802.21; Simulation; Prototyping.

1 Introduction

The exponential growth of the wireless mobile traffic, both voice and data, poses new requirements on the future networks [1]. Recent study from ABI Research [2] shows that the global volume of mobile data traffic will exceed 107 Exabytes in 2017. Furthermore, the proliferation of internet-connected mobile devices will continue to grow, forcing networks operators to dynamically increase the capacity of their networks.

Contemporary end-user mobile devices integrate a plethora of radio access technologies including 3GPP/3GPP2-based technologies as well as non-3GPP-based technologies such as Wi-Fi and Bluetooth. This terminal side heterogeneity drives R&D efforts towards efficient interworking and seamless transition among different air interfaces. The transparent transition between multiple Radio Access Technologies (RATs) is usually referred to as vertical handover (VHO). The existence of multiple different RATs triggered a development of Self Organizing Networks (SONs) that address heterogeneity in the wireless networks by introducing efficient and self-controlled methods for seamless VHOs. In addition, the heterogeneous wireless networks have a broader meaning than just a network whose radio access part comprises several air interfaces belonging to a same wireless system family (e.g. 2G, 3G and beyond). This concept needs to consider other, i.e. non-3GPP-based, wireless networks such as Wi-Fi and WiMAX, forcing new and specific solutions for transparent interworking. The introduction of the heterogeneous networking leads to resource management approach shift from technology-centric to consumer-centric. The heterogeneity is reflected in two main concepts, i.e. multiple different networks and multiple different layers in one RAT (i.e. network cell densification). Its presence in future wireless networks yields ubiquitous services through multiple communication interfaces in a

user-transparent manner.

Several international standardization bodies have taken steps towards including radio access heterogeneity in contemporary wireless systems. The enclosure of VHOs between different RATs in several existing standards paved the way for major advances and network efficiency improvement, as one of the main tools for efficient Radio Resource Management (RRM) in modern heterogeneous networks. 3GPP Release 8 (LTE) made one of the most significant contributions towards adopting and standardizing the heterogeneous networks and exploiting the benefits from their deployment [3], [4]. The previous, third generation release, provided the necessary technical specifications for wireless system interworking and introduced the VHO as an important RRM functionality that may provide uninterrupted, seamless and ubiquitous wireless connectivity and traffic offloading with more efficient radio resource utilization [3]- [5]. However, in this release, the specifications mainly focus on seamless and transparent handover between different Radio Access Networks (RANs) and air interfaces of the legacy 3GPP cellular systems such as GSM and UMTS. In Release 8, besides additional enhancements and new interfaces for interworking and VHO between the Evolved UTRAN (E-UTRAN) and legacy 3GPP RANs, 3GPP extends and utilizes the concept of VHO even further. In particular, interworking with non-3GPP RANs is one of the key design goals for the novel and evolved system architecture (SAE). The non-3GPP Interworking System Architecture in LTE includes a set of solutions in two categories. The first category contains a set of generic and loose interworking solutions that can be used with any other non-3GPP RAN (such as IEEE-based wireless systems). Mobility solutions defined in this category are referred as Handovers without Optimizations. The second category includes a specific and tighter interworking solution with one selected RAN, i.e. the cdma2000 High Rate Packet Data Air Interface. This solution category is called Handovers with Optimizations. As specified, the non-3GPP Inter-working System Architecture in LTE completely leverages on flat architecture of E-UTRAN and the flexibility and modularity of SAE. HP has proposed another state-of-the-art implementable solution that belongs to the first heterogeneity concept with multiple RATs [6]. The basic idea is to enable the communication service providers to streamline the transition to long-term evolution (LTE) and embrace a diversity of mobile access networks with a single subscriber management solution. The new HP solution bridges 2G, 3G, 4G, Wi-Fi and IP Multimedia Subsystem (IMS) networks to enable service providers to centrally manage subscribers' profiles, regardless of the networks to which they are connected.

Another example is the solution proposed by Nokia-Siemens, named Smart Wi-Fi [7]. It is a carrier-grade, end-to-end solution for building, optimizing and controlling Wi-Fi networks that integrate seamlessly with cellular networks. The need for general system interworking framework led to the initiation (in 2004) of the IEEE 802.21 standard, which was published in 2008 [8]. The standard supports various technical solutions that enable seamless VHOs introducing the so called Media Independent Handover (MIH). In particular the standard allows transparent interoperability among different underlying technologies such as 802.3, 802.11, 802.15, 802.16, 3GPP and 3GPP2 families. The forthcoming subsections present and elaborate on IEEE 802.21 standard as an important general system interworking framework for heterogeneous networks.

All previous examples belong to heterogeneity cases with multiple RATs. According to the other concept, multi-layering in a single radio access technology can also provide heterogeneity in the network. The growing demand for affordable mobile broadband connectivity is driving the interest towards implementation of a multitude of small cells, such as micro, pico and femto cells. Small cells were used mainly for fill-in purposes to improve the network coverage in the early days of GSM and until recently with HSPA. Today, there is a constant outdoor small cell (micro and pico) densification, which is cost-effective and relatively simple option for adding capacity to the wireless network. For indoor traffic offload, femtocell deployment (with the so-called Home eNodeB) represents a promising solution [9], [10], where the end users are making the deploy-

ment without any interaction of the network operators. Since more than 80 per cent of global wireless data traffic will be generated indoor [11], the femtocell concept will play important role in the future cellular networks. Recently, several international standardization committees have expressed their intents to provide standardized approach towards wireless network densification. 3GPP Release 10 and LTE-Advanced (LTE-A) [12] introduce a number of technical specifications towards small cell deployments and network densification with strong emphasis on femtocell deployments. The LTE-A evolved framework for Inter-Cell Interference Coordination (eICIC) and the support for Self-Organization, provides two types of multilayer coupling. The first type comprises closely coupled cells (such as macro cells), which are subjected to frequency planning and RRM. The second type consists of loosely coupled cells (e.g. picocells or femtocells) that auto-configure and auto-optimize and provide frequency reuse with factor one. Both types of cell coupling exploit the X2 interface for coordination. It is expected that the network densification process and the aggressive frequency reuse will enable dramatic increase in wireless capacity, data rates and quality of user experience [13].

Among many previously mentioned examples and approaches to cope with the problems related to the heterogeneous networks, this paper focuses on the IEEE 802.21 standard as the enabler of VHO in heterogeneous networks. It provides extensive simulation-based validation of the IEEE 802.21 heterogeneous network performance in various scenarios. The potentials of the standard for seamless interoperability are extensively exploited to develop a complete solution for reconfigurable and interoperable network resource management (RM) mechanisms for heterogeneous environments. The proposed platform is then tested and validated in a state-of-the-art laboratory simulator and is completely ported to a laboratory prototype of a reconfigurable multi-interface mobile terminal. Up to the best of authors' knowledge, this prototype is the first in the literature that relies on the IEEE 802.21 standard and incorporates an advanced RM.

2 Reconfigurability and Interoperability in Heterogeneous Networks

The modern network devices now include a plethora of built-in communication interfaces, of both wired and wireless type. The most popular consumer electronic devices such as tablets, laptops, personal digital assistants or even smart-phones are supporting wideband connectivity through both cellular technologies and Wireless Local Area Networks (WLANs). At the same time, communication devices are capable of exchanging, processing and displaying multimedia-reach contents in real time [1]. These new multi-access devices push the boundaries of the modern communications towards a network environment that is referred to as the fourth-generation of mobile communications (4G). The major benefits of the previous third-generation (3G) communications involve increased data rates (compared to the second generation of mobile devices), seamless mobility within large geographical areas, along with a global reachability. The biggest challenge for the 4G networks includes beyond transparent mobile communication within only one access network, and a global reachability in form of anytime, anyplace using heterogeneous communication technologies. For this purpose, different communication technologies need to integrate into a single heterogeneous platform [15]. This platform should be capable of supporting user transparent roaming and efficient delivery of multimedia traffic. On the other hand, the terminal devices participating in a heterogeneous network need to be able for autonomous operation, requiring the capability of self-reconfiguration. As a result, the notion of VHO in heterogeneous scenarios becomes a necessity. The VHO concept requires the need to research and develop novel solutions that support interoperability among different communications technologies. This gives rise to the concept of reconfigurable interoperability [16].

The reconfigurable interoperability is a cornerstone of modern communications and is essential driving force towards the multimedia content delivery in heterogeneous environment. The reconfigurable interoperability can be obtained on network side, on user side, or on both. This brings benefits for both, network providers and users. At the same time it contributes to the robustness of the provided services, allowing seamless and transparent network management.

When implemented on network level, the reconfigurable interoperability enables the network providers with a possibility to choose between a variety of wireless access networks. In this case, the access technology selection could be based on several criteria, such as:

- Comparison of access resources availability and specific service requirements (e.g. channel state, outage probability, VHO probability, user QoS requirements, context awareness etc.);
- Traffic load sharing and distribution between different coexisting networks;
- Efficient spectrum sharing;
- Network discovery and preferred gateway selection;
- Network congestion;

When implemented on user side, the reconfigurable interoperability will lead to more efficient end-to-end connectivity and service delivery in heterogeneous environments, easier worldwide roaming and dynamic adaptation to regional contexts, enhanced personalization and richer services. The users' terminal devices can reconfigure based on:

- Available resource capabilities;
- Minimization of the service cost when multiple underlying technologies are available;
- Anticipation of communication quality, as well as user contexts and preferences;
- User's mobility [17].

Autonomic decision making and self-healing capabilities directly provided by the reconfigurable interoperability can greatly improve the communication reliability. Furthermore, network providers can use the reconfigurability to introduce the value-added services more easily. They can exploit these features at the application level, since they have the possibility to introduce new services of various types. This will lead to more vibrant market movement and increased consumers' choices.

The reconfigurable interoperability in future convergent and multimedia reach communications must not be provided by isolated proprietary and vendor related solutions (such as Nokia's multimode Unlicensed Mobile Access - UMA concept), but rather should be supported by internationally recognized standard, like IP Multimedia Subsystem (IMS) and LTE's Evolved Packed Core (EPC) - based solutions. These approaches address the aspect of interoperability and VHOs by incorporating variety of advanced mechanisms (e.g. PCC [15], SRVCC [19], [20] PMIP [21], etc.) that support the VHO processes. As discussed in [22], the known solutions show handover performance inconsistencies that lead to degradation and in worst case complete failure of the handover procedures. This is where the new IEEE 802.21 standard positions.

The following two subsections will give an overview of the IEEE 802.21 standard, along with the necessary validation of its benefits in reconfiguration and inter-operation of heterogeneous networks.

2.1 IEEE 802.21 standard and Network Reconfiguration

The IEEE 802.21 standard [16] is a result of the work performed within the Media Independent Handover Services group of IEEE. The working group has been initiated in 2004 and the latest draft version of the standard was accepted in November 2008. The standard itself has been published in January 2009. The actual deployment of the standard is taking place at the moment and is predicted to intensify in the near future.

In the heart of the IEEE 802.21 framework lays the Media Independent Handover Function - MIHF. All IEEE 802.21 compatible devices should contain a MIHF in order to provide communication with different terminals, networks and remote MIHFs. MIHF provides abstract services to the higher layers (MIH User) using a unified interface (located on layer 2.5 according to the Open System Interconnection-OSI reference layer model). MIHF defines three different services: *Media Independent Event Service (MIES)*, *Media Independent Command Service (MICS)* and *Media Independent Information Service (MIIS)*.

Changes and condition of the Link Layer trigger appropriate MIES events. MICS provides the upper layers necessary commands to manage the link behavior. MIIS provides information about the neighboring networks and their current status and capabilities.

The IEEE 802.21 standard aims to provide transparent communication in heterogeneous environments by enabling seamless HOs between available access technologies. The standard defines mechanisms for network-in-range discovery and execution of intelligent VHOs, based on established link conditions and mobile devises' capabilities and preferences. For instance, WLAN can be preferred when available, instead of expensive cellular network communication, especially for heavy data transfer. As another example, the device can choose the strongest signal network in order to obtain itself with best QoS, or the network can balance traffic load in order to obtain stable communications. Present devices do not possess capabilities for intelligent self-reconfiguration. Users can only manually select the communication technology interface based on their knowledge and review of the network state. IEEE 802.21-capable devices are always aware of the available access networks and changes in link conditions. By combining this information with some intelligent RM implemented within mobile devices or on the network side, the communication is getting closer to the envisioned concept of 4G.

The main contribution of the IEEE 802.21 relies on a technology-independent abstraction layer which provides a generic interface to the processes operating on upper protocol layers. In this way the upper layers do not need to be specialized in processing the technology specific primitives, resulting in much simpler and complexity-free upper layer processes. The mobile device maintains a relevant set of networks and link status in a generic manner and the resource manager utilizes this set of information, containing static and dynamic aspects of the links, in order to fulfill its comprehensive decision-making activities. The resource manager can further involve Mobile IP aspects and functions, especially in the handover decision making process.

2.2 Simulation and Validation of IEEE 802.21 based Interoperability

This subsection validates the benefits of the IEEE 802.21 standard by presenting a simulation platform, which combines a commercially available network simulator (QualNet) and a Specification and Description Language (SDL)-based protocol developer [16]. This combination facilitates novel protocols and architectures development and their subsequent validation in a state-of-the-art simulator. The targeted simulation scenario for IEEE 802.21 performance analysis Fig.1 comprises a heterogeneous wireless network (HWN) consisting of two IEEE 802.11 Access Points (APs) providing local coverage, two IEEE 802.16 Base Stations (BSs) providing metropolitan coverage and a satellite network for global scenario coverage. All radio access technologies overlap in order to enable increased connectivity options. There are a varying number of

mobile nodes in the scenario communicating with the Correspondent Nodes (CNs) located within the network infrastructure. The users are allowed to move according to the random waypoint mobility model and they exhibit frequent vertical handovers in the scenario, i.e. changing of the Point of Attachment (PoA) which serves the users. All mobile nodes have active constant bit rate (CBR) applications of 64 kbps (voice application) or 2 Mbps (video application). The main performance parameter of interest that is crucial for the entire reconfigurable interoperability paradigm is the Vertical Handover Latency (VHL). The minimization of the VHL value allows seamless vertical handovers in heterogeneous wireless networks and facilitates the RM problem substantially. Its calculation must take into consideration the end-to-end delay of the new serving network and the receiver packet latency once a vertical handover occurs. The former parameter is calculated as:

$$\tau_{e2e}^{i} = t_{rcv}^{i} - t_{sent}^{i} \tag{1}$$

where t_{rcv}^{i} is the packet receive time and t_{sent}^{i} is the packet sent time on the new service network (denoted as network i) when a vertical handover happens. Using (1), the VHL value can be easily derived as:

$$\tau_{VHL}^{ij} = \tau_{L}^{ij} - \tau_{e2e}^{i} \tag{2}$$

where τ_{L}^{ij} is the receiver packet latency (defined as the time difference between the last successfully received packet from the old service network j and the first successfully received packet from the new service network i).

Fig.2a and Fig.2b depict the average VHL value for 30 and 90 mobile nodes in the scenario, respectively. It is evident that the introduction of the IEEE 802.21 functionalities provides significantly lower VHL values, a smaller increase of the VHL value with the increase of the number of mobile nodes and almost constant VHL values regardless of the users' mobility pattern.

Fig.2c depicts a comparison of the average VHL for different application bit rates, low mobility and high number of nodes. The introduction of IEEE 802.21 exhibits lower VHL values for all analyzed scenarios proving its superior performance.

The validation results clearly show that the IEEE 802.21 standard can be foreseen as a suitable reference for enabling seamless vertical handovers, providing user transparent interoperability. In some use cases like high speed mobility or large number of users, the VHL exceeds certain delay requirements for real time applications e.g. conversational video. This negative effect is a result of the congested access networks in the case of large number of users or the use of satellite links in the case of high speed mobility which introduces substantial round trip time.

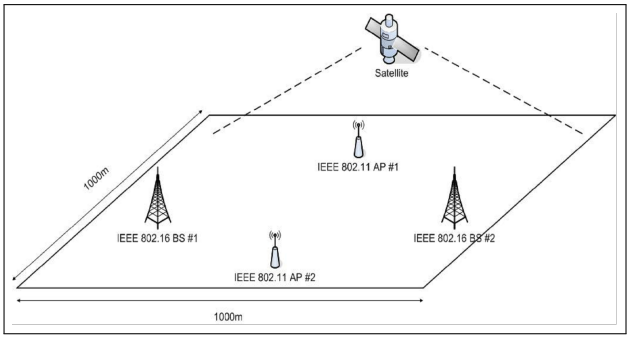


Figure 1: Simulation scenario

In contrast to the simulation scenarios presented here, any real-world system implementations would require a coexistence and interworking of IEEE 802.21 with a number of contemporary

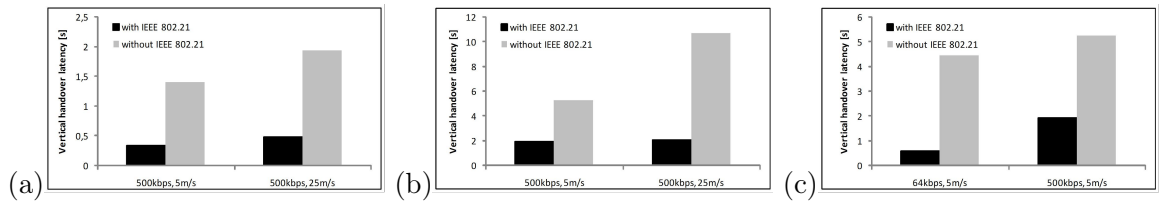


Figure 2: Simulation results: (a) VHL for 30 mobile nodes, (b) VHL for 90 mobile nodes, (c) VHL for different application bit rates

networking solutions like, SIP/SDP signaling, IMS procedures, EPS bearer concepts, etc. In most of the cases the above mentioned coexisting network solutions can be utilized as MIH users and provide the required inter-working capabilities between them and IEEE 802.21. Because the IEEE 802.21 standard envisions a prediction mechanism, all MIH users reconfigure their parameters of interest before the HO occurs, providing a seamless VHO execution that results in a VHL that is impacted only by the IEEE 802.21 HO processes. Hence, the performance of a real-work implementation of the IEEE 802.21 standard is not expected to dramatically defer from the results elaborated in Fig.2.

The following section provides extensive details on how IEEE 802.21 standard is customized and fitted into a novel architecture for providing reconfigurable interoperability of wireless communications systems combining it with an intelligent resource management.

3 IEEE 802.21 Based Radio Resource Management

There are several ideas found in the literature that propose IEEE 802.21 usage for mobility and QoS management support in heterogeneous wireless networks [16]. There are examples of SIP and IEEE 802.21 convergence as a powerful tool for soft vertical handover execution, as well as proposals to use IEEE 802.21 for integration of multimedia broadcast technologies (DVB-H) with other terrestrial access networks. Some works utilize IEEE 802.21 for QoS provisioning in IEEE 802.16 - IEEE 802.11 environment, showing that the assistance of IEEE 802.21 contributes in decreasing the effects of handover latency, jitter and packet loss, thus improving the user perception. Additionally, IEEE 802.21 can be used in a hybrid Satellite - Terrestrial access networks or enable the so called Knowledge Based (KB) mechanism for network selection.
But, a fairly small amount of research work specifically addresses the problem of RM in heterogeneous wireless environment. There are some approaches which are based on Joint Radio RM (JRRM) framework for beyond 3G wireless heterogeneous systems capable of adapting to the resource assignments of the specific system conditions and QoS demands. Unlike related work in the field, the approach presented here enables a fully functional RM mechanism for heterogeneous wireless networks and possesses unique features allowing maximum user servicing and maximum network utilization [16]. In particular, this platform combines RM and reconfigurable interoperability within the principles defined by the IEEE 802.21 standard. As depicted in Fig.3, the MIH User (in this case a RM) resides above the lower MIHF (here denoted as interoperability module - IM). Both components, RM and IM, are located in the end-user equipment. The platform architecture comprises several functional blocks.

Application block: This block presents the user application that requires network resources. Different applications have different needs in terms of bit rate and delay. Furthermore, when making RM decision it is also important what type of user started that application. This message is transferred to the *User and Application Profile* entity (U&AProf). Here the *StartApp* message is processed in *U&A* message that identifies the user class and the application type. The *U&A*

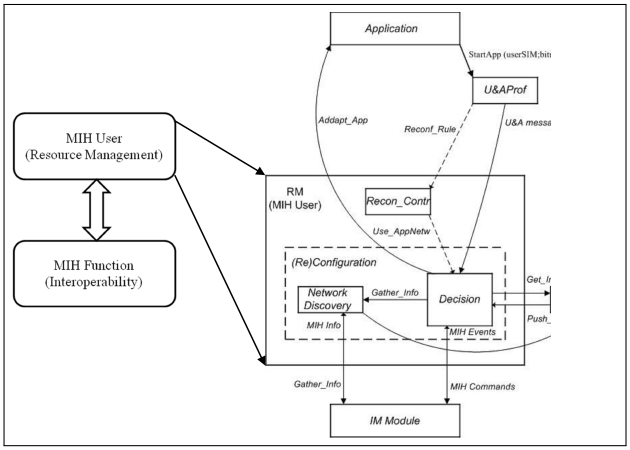


Figure 3: General architecture of the platform

message also presents energy (battery) consumption feature (with two modes of operation: save battery mode and normal battery mode). This information will be used further to select the mode in which the Decision block will operate. The *LRes* repository is a small database that stashes the information regarding the newly detected available networks.

Decision block: This block operates in several modes:

- Emergency Mode,
- BatteryLow Mode and
- Normal Mode.

In the Normal Mode, the Decision block performs simple switching of the user demands to what is available as a resource in the *LRes*. In the *BatteryLowMode*, the Decision block selects the technology that best fits the battery saving constraint, thus reducing QoS (the selected technology may not be best fitted for the application). In the Emergency mode, the RM module uses specially designed algorithm for sorting applications' serving priority.

Network Discovery block: This block has an interface towards the IM module for receiving MIH messages that carry relevant information about the networks in the users' vicinity. It uses *Store_ Info* message to fill the *LRes* database and its work is triggered by the Decision block with the *Gather_ Info* message.

Recon_Contr block: This block is left for additional reconfigurability constrains. It can learn and store consumers' behavior, thus providing further upgrade of the system (cognition). The ability of this platform to provide reconfigurable interoperability and efficient RM is validated on the same scenario as presented in Fig.1 The results compare the new architecture with the traditional way of serving consumers, i.e. the signal-to-noise ratio (SNR)-based serving policy, where mobile nodes connect to the PoA having the highest SNR value. The performance validation uses a novel performance metric, named *service retainability*. This Key Performance Indicator (KPI) parameter represents the average ratio of the dedicated bit rate to the required bit rate for mobile node, when exhibiting vertical handovers during the simulation scenario. The goal is to maximize the service retainability as its higher values mean that the mobile *node exhibits higher service retention in terms of dedicated bit rate*. Fig.4 presents the simulation results for the dependence of the service retainability on the number of mobile users for the new platform and the SNR based serving policies for 64 kbps and 2 Mbps application data rates. In particular, Fig.4a proves that the new architecture exhibits *increased service retainability* for 64 kbps

application data rates regardless of the number of mobile users and the users' mobility. Fig.4b depicts the new architecture behavior in terms of service retainability for 2 Mbps application data rates. It is evident that the proposed platform *outperforms* the SNR based serving policies and that maximum gains are achieved for low number of mobile users.

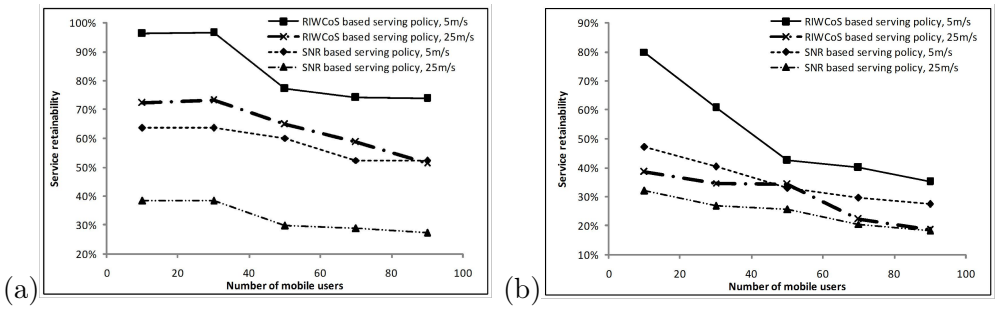


Figure 4: Simulation results: (a) Service retainability for 64 kbps application, (b) Service retainability for 2 Mbps application

Relying on the IEEE 802.21 standard, the proposed architecture is able to cope with RM issues in more efficient manner than the traditional methods proving its soundness for different consumer applications.

The next section will present the prototyping process of a developed consumer terminal which is capable of reconfigurable interoperability based on IEEE 802.21 principles [23]. The section explains different aspects of the reconfigurable terminal design and testing process.

4 Reconfigurable Terminal Prototyping

4.1 Terminal Architecture

In general, the terminal architecture overlays the Network Driver Interface Specification (NDIS) library. It is a standard application programming interface (API) for Network Interface Cards (NICs). A Media Access Controller (MAC) device driver wraps the details of NICs hardware implementation in such a way that all NICs for the same media can be accessed using a common programming interface. In this manner, the terminal architecture includes NDIS client sublayer, radio RM module (RRM) with MIHF and Video Data content used for transmission (Fig.5).

The NDIS library provides the features of NDIS protocol (NDISProt), which connects directly with the NIC hardware. The NDIS client sublayer coordinates the work of all entities, enabling the reconfiguration function through the Main Thread. The ReadHandler and the WriteHandler are the key features controlled by the Main Thread that provide data communication between the content located in the Content Block and the NICs hardware. The test scenario uses video data, where the Main Thread controls the video content in terms of segmentation of packets according to the transmission parameters. The WriteHandler additionally prepares the packets by adding data payload and appropriate Ethernet, IP, UDP and TFTP headers.

Before transmitting of any data, the Main Thread sends information about the network availability to the RRM through the MIHF. The logic of selection an appropriate communication interface depends on the decision made by the RRM module. The RRM and the MIHF are designed and functionally tested by using Specification and Description Language (SDL). The RRM module receives and sends MIH information from and to the NDIS client sublayer in order

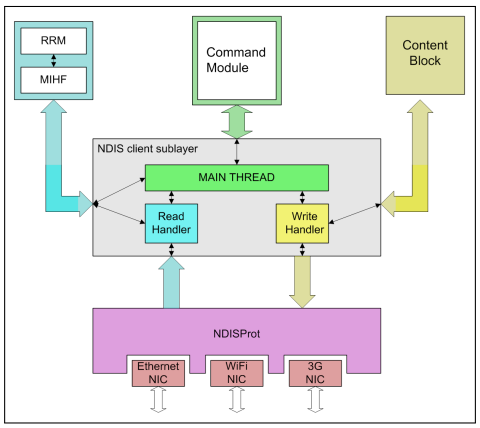


Figure 5: Terminal architecture

to enforce the usage of the most suitable network interface. The decision algorithm is simple and it ranks the three available technologies by priority (Ethernet, WLAN, UMTS). When there is no Ethernet connection available, before using the WLAN, the signal strength of the WLAN is inspected. If this value is lower than a predefined threshold, the WLAN interface is not selected and the packets are send on the UMTS communication interface. The RRM makes decision for using WLAN if the inspected signal strength is above the satisfactory level. The handover between the wireless networks (WLAN/UMTS) occurs when the WLAN signal drops under a predefined threshold set by the RRM.

A Command Module is specifically developed for the Ethernet/WLAN handover in order to overcome a possible blocking of communication during unplugging of the Ethernet cable.

4.2 Testing Scenario and Results

Fig.6 presents the testing scenario setup of the developed reconfigurable terminal. The Correspondent Node (CN) and the reconfigurable Mobile Node (MN) are both attached to the same local network. Consequently, the traffic flow when using Ethernet and WLAN does not rout outside the local network. However, when using UMTS the traffic flow is routed from UMTS towards the CN through Internet (as a global network). The testing scenario begins while the Ethernet cable is connected. When the cable is removed, the terminal is within WLAN coverage and performs vertical handover to the available WiFi access point (AP). As the terminal moves away from the AP, the WLAN signal level drops. When the signal is under a predefined threshold, the MN conducts a vertical handover towards the UMTS network.

The communication performance generally depends on the networks configuration (radio parameters, local or global networks) and hardware configuration (NIC, operating system, etc.) and is denoted by the achieved throughput and inter-packet delay. Fig.7a and Fig.7b depict the measured throughput for the case of WLAN to UMTS handover, while Fig.7c and Fig.7d present the packet delay.

The measurement results reveal a serious problem regarding the WLAN/UMTS handover performance. As the WLAN is a local network and the UMTS is a global network, the handover suffers from a high communication-break period followed by a low performance period (named *slow start*) for the UMTS throughput. These periods are denoted as T_{break} and T_{slow} in Fig.7a. The described problem is not particularly related to the WLAN and UMTS networks. The very same problem will occur for every transition from a local to a global serving network (e.g. from

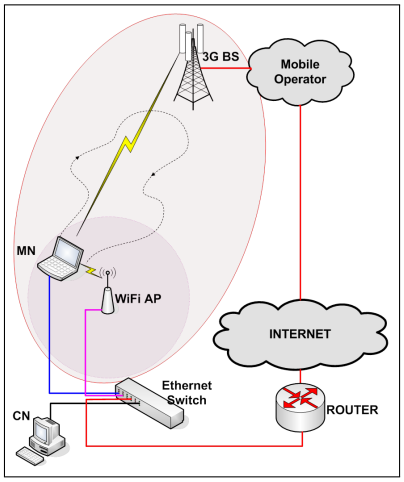


Figure 6: Testing scenario

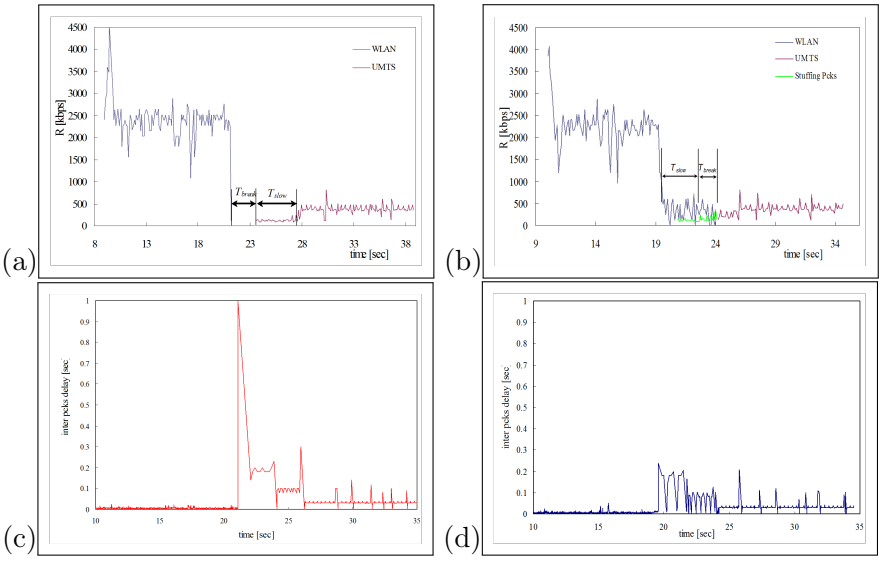


Figure 7: Handover performance: (a) Measured throughput before terminal upgrade, (b) Measured throughput after terminal upgrade, (c) Measured delay before terminal upgrade, (d) Measured delay after terminal upgrade

Ethernet to WLAN-based metropolitan network, from local WLAN to WiMAX network, from a terrestrial network to a satellite network, etc.). In order to overcome this problem, additional features are implemented in the terminal, enhancing the capabilities of its MIHF and NDIS sublayer.

4.3 Handover Prediction and Terminal Reconfiguration

The MIHF is extended with a *handover prediction* capabilities, whereas the NDIS sublayer is enhanced with additional reconfigurability control capabilities. The *linear predictor* implemented within the MIHF provides prediction of the near future WLAN received signal strength (RSS) values, presented by the RSS indicator (RSSI). The RSSI values $y_i, i = 1, 2, , N$ are measured at equally spaced time intervals. The predictor takes the last M RSSI values $y_i, i = 1, 2, .., M$ and predicts the following $(M + 1)$-th RSSI value. A stationarity is assumed, meaning that the autocorrelation function $R = y_j, y_k$ depends only on the difference $|jk|$ and not on the particular j and/or k values. Hence, the autocorrelation R has only single index and can be calculated as:

$$R_j = \langle y_j y_{i+j} \rangle \approx \frac{1}{N-j} \sum_{i=1}^{N-j} y_i y_{i+j} \tag{3}$$

Under the given assumptions, the estimation of the next RSSI value is:

$$\hat{y}_n = \sum_{j=1}^{M} d_j y_{n-j} + x_n \tag{4}$$

where d_j are linear prediction (LP) coefficients, obtained from the following set of M equations:

$$\sum_{j=1}^{M} R_{|j-k|} d_j = R_k, (k = 1, ..., M) \tag{5}$$

In (4), x_n is positive root of the mean square discrepancy $\left\langle x_n^2 \right\rangle$, estimated with the following equation:

$$\left\langle x_n^2 \right\rangle = R_0 - R_1 d_1 - R_2 d_2 - ... - R_M d_M \tag{6}$$

Based on this theoretical approach, the predictor can provide the MIHF with one predicted RSSI value $\hat{y}(t)$, based on the previous M received values:

$$\hat{y}(t) = f(y(t-1), y(t-2), .., y(t-M)) \tag{7}$$

For the prediction of the next RSSI values, the predicted output $\hat{y}(t)$ is used as a current input, coupled with $(n-1)$ delayed measured RSSI samples to provide the next RSSI sample $\hat{y}(t+1)$. This can be repeated N times, to predict the RSSI values sufficiently in advance:

$$\begin{aligned} \hat{y}(t+N) = f(&\hat{y}(t+N-1), \hat{y}(t+N-2), .., \hat{y}(t), \\ &y(t-1), y(t-2), .., y(t-M-N)) \end{aligned} \tag{8}$$

The prediction method is implemented within MIHF as a standalone procedure which requires two different parameters, i.e. the values of M and N.

The second enhancement of the architecture, i.e. the modification of the NDIS sublayer, involves introduction of two new operating modes: *mixed mode* and *stuffing mode*. When in mixed mode, each cycle of the Main Thread invokes the WriteHandler with interleaved interface

handlers (WLAN and UMTS). This enables the MN to simultaneously send packets on WLAN and UMTS by switching. The stuffing mode is similar to the mixed mode, but the UMTS packets do not carry information bits. The mixed mode is used to overcome the T_{break} period by sending UMTS packets before the actual dropping of the WLAN signal. This compensates the initial delay introduced by the UMTS technology that results in the T_{break} period. The mixed mode needs to be activated T_{break}-time before the communication break. In addition, the stuffing mode should be introduced prior to the mixed mode in order to overcome the slow start of the UMTS network.

A total overcoming of the WLAN/UMTS handover problem requires a prediction of WLAN signal dropping for $T_{break} + T_{slow}$ time before its actual occurrence. The MIHF informs the RRM when predicting such situation. The RRM sends a command to the Main Thread to operate for T_{slow} seconds in stuffing mode and for the next T_{break} seconds in mixed mode. The network throughput in this case is depicted in Fig.7b). Regarding the inter-packet delay, the results show that the predictor implementation significantly improves the handover performance, resulting in seamless vertical handover and session continuity without a delay peak (see Fig.7d).

The WLAN throughput drops during the stuffing and mixed mode, but the communication continues seamlessly during the handover. It is important to note that the bitrates of WLAN and UMTS during the mixed period can be controlled according to the following analysis.

When only the WLAN interface is used for communication, the achieved interface capacity utilization (R_W) can be calculated as:

$$R_W = \frac{BS_W \cdot T_W}{BS_W \cdot T_W + TC} \tag{9}$$

where BSW represents the optimal burst size defined as the number of packets sent through the WLAN interface in one Main Thread cycle, T_W is the amount of time spent for sending single packet through the WLAN interface, and TC is the delay introduced by the Main Thread while checking all control information in single cycle. In the same manner, we can calculate the UMTS interface capacity utilization R_U for the period when only UMTS interface is used for communication (10), only this time we use the BS_U (optimal UMTS burst size) and T_U (UMTS packets sending time) parameters.

$$R_U = \frac{BS_U \cdot T_U}{BS_U \cdot T_U + TC} \tag{10}$$

During the mixed period the Main Thread uses both interfaces for communication in one cycle. Consequently the interface capacity utilization is different and can be calculated as:

$$R_{WM} = \frac{BS_{WM} \cdot T_W}{BS_{WM} \cdot T_W + BS_{UM} \cdot T_U + TC} \tag{11}$$

$$R_{UM} = \frac{BS_{UM} \cdot T_U}{BS_{UM} \cdot T_U + BS_{WM} \cdot T_W + TC} \tag{12}$$

where BS_{WM} and BS_{UM} represent the values of the burst size for WLAN and UMTS during the mixed period, respectively. By changing the BS_{WM} and BS_{UM} parameters as independent variables, the R_{WM} and R_{UM} can be controlled within the following boundaries:

$$\frac{T_W}{T_W + BS_{UM} \cdot T_U + TC} \leq R_{WM} \leq \frac{BS_{WM} \cdot T_W}{BS_{WM} \cdot T_W + T_U + TC} \tag{13}$$

$$\frac{T_U}{BS_{WM} \cdot T_W + T_U + TC} \leq R_{UM} \leq \frac{BS_{UM} \cdot T_U}{T_W + BS_{UM} \cdot T_U + TC} \tag{14}$$

The results presented in Fig.7b and Fig.7d use $BS_{WM} = BS_{UM} = 1$. The results show that the implementation of handover predictor and terminal reconfiguration (i.e. the introduction of mixed and stuffing modes) successfully overcomes the problems related to MN's transition from a local to a global serving network.

5 Conclusions and Future Works

The concept of reconfigurable interoperability becomes a cornerstone of future communication systems under heterogeneous access scenarios. The ability to choose among different access networks with user transparent and seamless vertical handovers and the ability to reconfigure and sustain the required QoS levels is a quintessential aspect of the development towards modern communications.

The emerging IEEE 802.21 standard represents a promising effort in this direction aiming at provisioning a global standard for media independent handovers. This paper shows the quantitative benefits and the corresponding potentials of the standard, providing case studies involving simulation and demo-testing platforms. The simulations have proved that the introduction of IEEE 802.21 exhibits lower vertical handover latency values for all analyzed scenarios. Furthermore, the results revealed that the proposed platform that combines IEEE 802.21 standard and intelligent resource management outperforms the SNR based serving policies and especially when the scenario involves low number of mobile users. Finally, the paper presented a reconfigurable terminal prototype, which is capable of signal prediction, fast interface switching, combined transmission, etc. The future of communications will bring intelligent devices which should be able to combine different data flows coming from different communication technologies with high rates, while allowing the user to freely traverse through heterogeneous environments with high speeds.

Acknowledgment

This research was sponsored by NATO's Public Diplomacy Division in the framework of "Science for Peace" through the SfP-982469 "Reconfigurable Interoperability of Wireless Communications Systems (RIWCoS)" project [24].

Bibliography

[1] Ericsson (2012); Ericsson Mobility Report, http://www.ericsson.com/res/docs/2012/ericsson-mobility-report-november-2012.pdf

[2] ABI Research (2012); Worldwide Mobile Data Traffic Will Exceed 107 Exabytes in 2017, But That Doesnt Imply a Data Tsunami, http://www.abiresearch.com/press/worldwide-mobile-data-traffic-will-exceed-107-exab.

[3] 3GPP TS 23.402 (2008); Architecture enhancements for non-3GPP accesses - Release 8.

[4] Holma H. (2010); *LTE for UMTS: Evolution to LTE-Advanced*, Wiley & Sons.

[5] Holma H. (2007); *WCDMA for UMTS: HSPA evolution and LTE*, Wiley & Sons.

[6] HP (2013); HP Offers Mobile Network Operators Clear Path to Heterogeneous Networking, http://www8.hp.com/uk/en/hp-news/press-release.html?id=1375527.

[7] Nokia Siemens Networks (2013); Nokia Siemens Networks Smart Wi-Fi makes mobile integration seamless, http://www.nokiasiemensnetworks.com/portfolio/products/small-cells/smart-wi-fi.

[8] *IEEE Standard for Local and Metropolitan Area Networks, Media Independent Handover Services, IEEE 802.21*, (2008).

[9] Zhang J.; de la Roche G. (2010); *Femtocells Technologies and Deployments*, Wiley & Sons.

[10] Sounders S.R.; Carlaw S.; Giustina A.; Bhat R. R.; Rao V.S.; Siegberg R. (2009); *Femtocells Opportunities and Challenges for Business and Technology*, Wiley & Sons.

[11] Nokia Siemens (2012); Deployment strategies for heterogeneous networks, *White paper*.

[12] 3GPP TR 36.913 V10.0.0 (2011); Requirements for Further Advancements for Evolved Universal Terrestrial Radio Access, LTE-Advanced - Release 10, *3GPP*.

[13] Nakamura T. et al (2013); Trends in small cells enhancements in LTE-Advanced, *IEEE Communications Magazine*, 51(2), 98-105.

[14] Georganopoulos N. et al. (2004); Terminal-Centric View of Software Reconfigurable System Architecture and Enabling Components and Technologies, *IEEE Communications Magazine*, 42(5), 100-110.

[15] Hossain E. (2008); Heterogeneous Wireless Access Networks: Architectures and Protocols, *Springer*.

[16] Atanasovski V.; Rakovic, V.; Gavrilovska, L. (2010); Efficient Resource Management in Future Heterogeneous Wireless Networks: the RIWCoS Approach, *IEEE Military Communications Conference (MILCOM)*, San Jose, CA, USA.

[17] Zhu Y.; Ni L.; Li L. (2013); Exploiting mobility patterns for inter-technology handover in mobile environments, *Computer Communications*, ISSN 0140-3664, 36(2), 203-210.

[18] 3GPP TS 23.203 Tech. Spec. (2008); Policy and Charging Control Architecture.

[19] 3GPP TS 23.216 V8.6.0 (2009-12); Single Radio Voice Call Continuity, SRVCC -Release 8.

[20] 3GPP TR 36.938 (2009); Improved Network Controlled Mobility between E-UTRAN and 3GPP2/Mobile WiMAX Radio Technologies.

[21] Ali I. et al. (2009); Network-based mobility management in the evolved 3GPP core network, *IEEE Communications Magazine*, 47(2), 58-66.

[22] Knaesel F.J.; Neves P.; Sargento S. (2011); IEEE 802.21 MIH-enabled Evolved Packet System Architecture, *Third Int. ICST Conference, MONAMI 2011*, Aveiro, Portugal.

[23] Latkoski P.; Ognenoski O.; Rakovic V.; Gavrilovska L. (2010); Prototyping and Optimization of IEEE 802.21-based Reconfigurable Mobile Terminal, *Military Communication Conference - MILCOM*, California, USA.

[24] NATO SfP-982469; Reconfigurable Interoperability of Wireless Communications Systems (RIWCoS), http://riwcos.comm.pub.ro

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):333-347, June, 2015.

Determining Basic Probability Assignment Based on the Improved Similarity Measures of Generalized Fuzzy Numbers

W. Jiang, Y. Yang, Y. Luo, X.Y. Qin

Wen Jiang*, Yan Yang, Yu Luo, Xiyun Qin
Northwestern Polytechnical University
School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
jiangwen@nwpu.edu.cn, yangyan7003@nwpu.edu.cn
345255046@qq.com, 945766782@qq.com
*Corresponding author

Abstract: Dempster-Shafer theory of evidence has been widely used in many data fusion application systems. However, how to determine basic probability assignment, which is the main and the first step in evidence theory, is still an open issue. In this paper, an improved method to determine the similarity measure between generalized fuzzy numbers is presented. The proposed method can overcome the drawbacks of the existing similarity measures. Then, we propose a new method for obtaining basic probability assignment (BPA) based on the proposed similarity measure method between generalized fuzzy numbers. Finally, the efficiency of the proposed method is illustrated by the classification of Iris data.
Keywords: data fusion, dempster-Shafer evidence theory, basic probability assignment (BPA), generalized fuzzy numbers, similarity measures

1 Introduction

Dempster-Shafer (D-S) theory of evidence is widely used in many fields of information fusion due to its efficiency in dealing with uncertain information. In real data fusion application systems based on DS theory, the basic probability assignment function should be given so that the combined BPA can be obtained through Dempster's rule of combination [1, 2]. However, how to determine basic probability assignment, which is the main and the first step in evidence theory, is still an open issue. A number of authors have addressed this problem using different approaches. Zhu et al. proposed a method to derive mass values from fuzzy membership degrees. For this purpose, fuzzy c-means (FCM) clustering is used to represent the grey levels as fuzzy sets [3]. Bendjebbour et al. [4] proposed a probabilistic model where the frame of discernment contained the individual clusters and the ignorance that was modeled by the union of all individual clusters. In that work, the authors derived the mass value of ignorance from the mixture of distributions of the individual clusters composing it. Guan et al. [5] came up with three methods to construct the BPA function based on gray correlation analysis, fuzzy sets, and attribute measure respectively. Chen et al. [6] and Li et al. [7] used back-propagation (BP) neural network to obtain basic probability assignment. According as neural network can gain stronger generalization ability, the measured data being processed by neural network can be used as the BPA value of every sensor. Xu et al. [8] used the difference matrix of deviation degree to represent quantitatively the degree of similarity between interval numbers, and constructed an expression of basic probability assignment function. Since the basic probability assignment of evidence theory obtained by using neural network has high computational complexity, Zuo et al. [9] put forward a method of rough set theory based on random set and BP neural network to obtain the basic probability assignment. In the framework of random set, the ability of attribute reduction of rough set was made use of to reduce the neural network input dimension. In papers [10-13], in order to solve the different practical problems, we proposed several different

approaches to obtain BPA. These more pragmatic methods are proposed to generate BPAs from uncertain information.

In this paper, an improved method to determine the similarity measure between generalized fuzzy numbers is presented. Then, a new method to obtain basic probability assignment (BPA) is proposed based on the improved similarity measure between generalized fuzzy numbers. An experiment of Iris data classification is used to illustrate the efficiency of our method.

2 Preliminaries

2.1 Dempster Shafer Evidence Theory.

Evidence theory first supposes the definition of a set of hypotheses $\Theta = \{H_1, H_2, \cdots, H_N\}$ called the frame of discernment. It is composed of N exhaustive and exclusive hypotheses. Let us denote $P(\Theta)$, the power set composed with the 2^N propositions A of Θ:

$$P(\Theta) = \{\emptyset, \{H_1\}, \{H_2\}, \cdots, \{H_N\}, \{H_1 \cup H_2\}, \{H_1 \cup H_3\} \cdots, \Theta\} \tag{1}$$

Where $\emptyset$ denotes the empty set. The N subsets containing only one element are called singletons. A key point of evidence theory is the basic probability assignment (BPA). A BPA is a function from $P(\Theta)$ to [0, 1] ,and which satisfies the following conditions:

$$\sum_{A \in P(\Theta)} m(A) = 1; m(\emptyset) = 0, \tag{2}$$

Dempster's rule of combination (also called orthogonal sum), noted by $m = m_1 \oplus m_2$, is the first one within the framework of evidence theory which can combine two BPAs m_1 and m_2 to yield a new BPA:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - k} \quad \text{and} \quad k = \sum\nolimits_{B \cap C = \emptyset} m_1(B) m_2(C) \tag{3}$$

Where k is a normalization constant, called conflict because it measures the degree of conflict between m_1 and m_2. $k = 0$ corresponds to the absence of conflict between m_1 and m_2, whereas $k = 1$ implies complete contradiction between m_1 and m_2. The belief function resulting from the combination of J information sources S_J defined as

$$m = m_1 \oplus m_2 \cdots \oplus m_j \cdots\cdots \oplus m_J, \tag{4}$$

As can be seen from above, multi source information can be easily fused in the framework of evidence theory, if we can obtain the BPA functions.

2.2 Generalized Fuzzy numbers

A generalized fuzzy number $\widetilde{A} = (a, b, c, d; w)$ is described as any fuzzy subset of the real line R with membership function $\mu_{\widetilde{A}}$ which has the following properties [14]:

(1) $\mu_{\widetilde{A}}$ is a continuous mapping from R to the closed interval in $[0, w]$, $0 \leq w \leq 1$;
(2) $\mu_{\widetilde{A}}(x) = 0$ for all $x \in (-\infty, a]$;
(3) $\mu_{\widetilde{A}}(x)$ is strictly increasing in $[a, b]$;
(4) $\mu_{\widetilde{A}}(x) = w$ for all $x \in [b, c]$, where w is a constant and $0 < w \leq 1$;
(5) $\mu_{\widetilde{A}}(x)$ is strictly decreasing in $[c, d]$;
(6) $\mu_{\widetilde{A}}(x) = 0$ for all $x \in [d, +\infty)$.

Where $0 < w \leq 1$, a, b, c and d are real numbers. Especially, a generalized trapezoidal fuzzy number can be defined as $\widetilde{A} = (a, b, c, d; w)$, where $a \leq b \leq c \leq d$, $0 \leq w \leq 1$, its membership function is defined by

$$\mu_{\widetilde{A}}(x) = \begin{cases} \dfrac{(x-a)}{b-a} & a \leq x \leq b \\ w & b \leq x \leq c \\ \dfrac{(x-c)}{d-c} & c \leq x \leq d \\ 0 & \text{else} \end{cases} \tag{5}$$

If $w = 1$, then the generalized fuzzy number $\widetilde{A}$ is called a normal trapezoidal fuzzy number, denote as $\widetilde{A} = (a, b, c, d)$. If $a = b$ and $c = d$, then $\widetilde{A}$ is called a crisp interval. If $b = c$, then $\widetilde{A}$ is called a generalized triangular fuzzy number. If $a = b = c = d$, then $\widetilde{A}$ is called a real number.

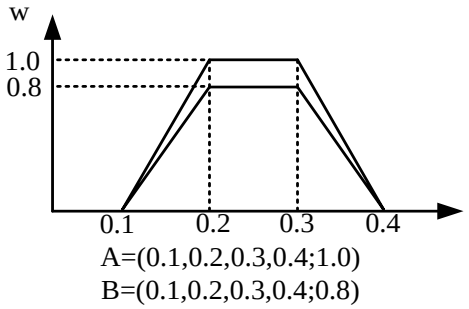


Figure 1: Two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$

Figure 1 shows two different generalized trapezoidal fuzzy numbers $\widetilde{A} = (0.1, 0.2, 0.3, 0.4; 1.0)$ and$\widetilde{B} = (0.1, 0.2, 0.3, 0.4; 0.8)$. Compared with normal fuzzy numbers, the generalized fuzzy numbers can deal with uncertain information in a more flexible manner. For example, in decision making situation, the values w_1 and w_2 represent the degree of confidence of the opinions of the decision-makers' $\widetilde{A}$ and$\widetilde{B}$, respectively, where $w_1 = 1$ and $w_2 = 0.8$.

2.3 A Review of the Existing Similarity Measures between Fuzzy Numbers

In this section, we briefly introduce some existing similarity measures between fuzzy numbers from Chen [15], Lee [16], Chen and Chen [17], Wei & Chen [18] and Hejazi, et al. [19].

Let $\widetilde{A}$ and $\widetilde{B}$ be two trapezoidal fuzzy numbers, where $\widetilde{A} = (a_1, a_2, a_3, a_4)$ and $\widetilde{B} = (b_1, b_2, b_3, b_4)$, as shown in Figure 2. Chen [15] presented a similarity measure between fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ based on the geometric distance, where the degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = 1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}. \tag{6}$$

Where $S(\widetilde{A}, \widetilde{B}) \in [0, 1]$. The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity between the fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$.

If $\widetilde{A}$ and $\widetilde{B}$ are two triangular fuzzy numbers, where $\widetilde{A} = (a_1, a_2, a_3)$ and $\widetilde{B} = (b_1, b_2, b_3)$. The degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the triangular fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as

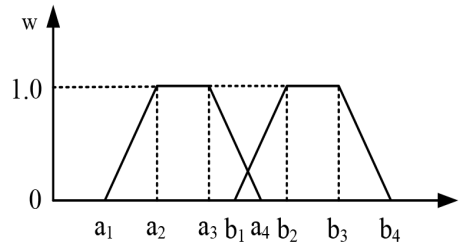


Figure 2: Trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$

follows [15]:

$$S(\widetilde{A},\widetilde{B}) = 1 - \frac{\sum_{i=1}^{3}|a_i - b_i|}{3}. \tag{7}$$

Where$S(\widetilde{A},\widetilde{B}) \in [0,1]$. The larger the value of $S(\widetilde{A},\widetilde{B})$, the more the similarity between the fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$.

Lee [16] presented a similarity measure between trapezoidal fuzzy numbers, where the degree of similarity $S(\widetilde{A},\widetilde{B})$ between the trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A},\widetilde{B}) = 1 - \frac{\left\|\widetilde{A} - \widetilde{B}\right\|_{l_p}}{\|U\|} \times 4^{-1/p}. \tag{8}$$

Where

$$\left\|\widetilde{A} - \widetilde{B}\right\|_{l_p} = \left(\sum_{i=1}^{4}(|a_i - b_i|)^p\right)^{1/p}. \tag{9}$$

and

$$\|U\| = \max(U) - \min(U). \tag{10}$$

In order to optimally aggregate experts' fuzzy opinions, Chen and Chen [17] presented a similarity measure between generalized trapezoidal fuzzy numbers. First, they calculate the center-of-gravity (COG) point $(x^*_{\widetilde{A}}, y^*_{\widetilde{A}})$ and $(x^*_{\widetilde{B}}, y^*_{\widetilde{B}})$ of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, respectively. The COG point $(x^*_{\widetilde{A}}, y^*_{\widetilde{A}})$ of the generalized trapezoidal fuzzy numbers $\widetilde{A} = (a_1, a_2, a_3, a_4; w_{\widetilde{A}})$ is calculated as follows:

$$y^*_{\widetilde{A}} = \begin{cases} \dfrac{w_{\widetilde{A}} \times \left(\dfrac{a_3 - a_2}{a_4 - a_1} + 2\right)}{6} & \text{if} \quad a_1 \neq a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \leq 1, \\ \dfrac{w_{\widetilde{A}}}{2} & \text{if} \quad a_1 = a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \leq 1, \end{cases} \tag{11}$$

$$x^*_{\widetilde{A}} = \frac{y^*_{\widetilde{A}}(a_3 + a_2) + (a_4 + a_1)(w_{\widetilde{A}} - y^*_{\widetilde{A}})}{2w_{\widetilde{A}}}. \tag{12}$$

Then the degree of similarity $S(\widetilde{A},\widetilde{B})$ between the two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ can be calculated as follows [17]:

$$S(\widetilde{A},\widetilde{B}) = \left[1 - \frac{\sum_{i=1}^{4}|a_i - b_i|}{4}\right] \times \left(1 - \left|x^*_{\widetilde{A}} - x^*_{\widetilde{B}}\right|\right)^{B(S_{\widetilde{A}}, S_{\widetilde{B}})} \times \frac{\min(y^*_{\widetilde{A}}, y^*_{\widetilde{B}})}{\max(y^*_{\widetilde{A}}, y^*_{\widetilde{B}})} \tag{13}$$

Where $B(S_{\widetilde{A}}, S_{\widetilde{B}})$ are defined as follows:

$$B(S_{\widetilde{A}}, S_{\widetilde{B}}) = \begin{cases} 1 & S_{\widetilde{A}} + S_{\widetilde{B}} > 0 \\ 0 & S_{\widetilde{A}} + S_{\widetilde{B}} = 0 \end{cases} \tag{14}$$

Where $S_{\widetilde{A}} = a_4 - a_1$ and $S_{\widetilde{B}} = b_4 - b_1$ are the lengths of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$. The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity measure between two fuzzy numbers.

Wei & Chen [18] proposed a method for calculating the similarity of two fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, where $\widetilde{A} = (a_1, a_2, a_3, a_4; w_{\widetilde{A}})$ and $\widetilde{B} = (b_1, b_2, b_3, b_4; w_{\widetilde{B}})$. If $0 \le a_1 \le a_2 \le a_3 \le a_4 \le 1$ and $0 \le b_1 \le b_2 \le b_3 \le b_4 \le 1$, then the degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = \left[1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right] \times \frac{\min(P(\widetilde{A}), P(\widetilde{B})) + \min(w_{\widetilde{A}}, w_{\widetilde{B}})}{\max(P(\widetilde{A}), P(\widetilde{B})) + \max(w_{\widetilde{A}}, w_{\widetilde{B}})} \tag{15}$$

Where $S(\widetilde{A}, \widetilde{B}) \in [0, 1]$; $P(\widetilde{A})$ and $P(\widetilde{B})$ are defined as follows:

$$P(\widetilde{A}) = \sqrt{(a_1 - a_2)^2 + w_{\widetilde{A}}^2} + \sqrt{(a_3 - a_4)^2 + w_{\widetilde{A}}^2} + (a_3 - a_2) + (a_4 - a_1). \tag{16}$$

$$P(\widetilde{B}) = \sqrt{(b_1 - b_2)^2 + w_{\widetilde{B}}^2} + \sqrt{(b_3 - b_4)^2 + w_{\widetilde{B}}^2} + (b_3 - b_2) + (b_4 - b_1). \tag{17}$$

$P(\widetilde{A})$ and $P(\widetilde{B})$ are the perimeters of generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, respectively. The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity measure between two fuzzy numbers.

Hejazi, etc.[19] presented an improved similarity measure between two fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ combining the concept of geometric distance, height, areas and perimeters of generalized fuzzy numbers. The degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = \left[1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right] \times \frac{\min(P(\widetilde{A}), P(\widetilde{B}))}{\max(P(\widetilde{A}), P(\widetilde{B}))} \times \frac{\min(A(\widetilde{A}), A(\widetilde{B})) + \min(w_{\widetilde{A}}, w_{\widetilde{B}})}{\max(A(\widetilde{A}), A(\widetilde{B})) + \max(w_{\widetilde{A}}, w_{\widetilde{B}})} \tag{18}$$

$P(\widetilde{A})$ and $P(\widetilde{B})$ are the perimeters of two generalized trapezoidal fuzzy numbers which are calculated by Eqs.(16),(17).On the other hand they have $A(\widetilde{A})$ and $A(\widetilde{B})$ which are the areas of the two fuzzy numbers and that are calculated as follows:

$$A(\widetilde{A}) = \frac{1}{2} w_{\widetilde{A}} (a_3 - a_2 + a_4 - a_1). \tag{19}$$

$$A(\widetilde{B}) = \frac{1}{2} w_{\widetilde{B}} (b_3 - b_2 + b_4 - b_1). \tag{20}$$

The larger the value of $S(\widetilde{A}, \widetilde{B})$, the more the similarity measure between two fuzzy numbers.

3 An Improved Similarity Measure of Generalized Fuzzy Numbers

Many similarity measures between fuzzy numbers have been proposed [15-19]. However, it has been found that the existing methods cannot correctly calculate the degree of similarity between

two generalized fuzzy numbers in some situations. In this section, we present an improved method to calculate the degree of similarity between generalized fuzzy numbers[20],which gives consideration to the horizontal center of gravity, the perimeter, the height and the area of the two fuzzy numbers. The proposed similarity measure can overcome the drawbacks of the existing methods.

Assume there are two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, where $\widetilde{A} = (a_1, a_2, a_3, a_4; w_{\widetilde{A}})$ and $\widetilde{B} = (b_1, b_2, b_3, b_4; w_{\widetilde{B}})$, $0 \le a_1 \le a_2 \le a_3 \le a_4 \le 1$ and $0 \le b_1 \le b_2 \le b_3 \le b_4 \le 1$. Then the degree of similarity $S(\widetilde{A}, \widetilde{B})$ between the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$S(\widetilde{A}, \widetilde{B}) = \left[1 - \left|x^*_{\widetilde{A}} - x^*_{\widetilde{B}}\right|\right] \times \left[1 - \left|w_{\widetilde{A}} - w_{\widetilde{B}}\right|\right] \times \frac{\min(P(\widetilde{A}), P(\widetilde{B})) + \min(A(\widetilde{A}), A(\widetilde{B}))}{\max(P(\widetilde{A}), P(\widetilde{B})) + \max(A(\widetilde{A}), A(\widetilde{B}))} \quad (21)$$

Where $x^*_{\widetilde{A}}$ and $x^*_{\widetilde{B}}$ are the horizontal center-of-gravity (COG) of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$, respectively. The COG point $(x^*_{\widetilde{A}}, y^*_{\widetilde{A}})$ of the generalized trapezoidal fuzzy numbers $\widetilde{A}$ is calculated as follows:

$$y^*_{\widetilde{A}} = \begin{cases} \dfrac{w_{\widetilde{A}} \times \left(\dfrac{a_3 - a_2}{a_4 - a_1} + 2\right)}{6} & \text{if} \quad a_1 \neq a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \le 1, \\ \dfrac{w_{\widetilde{A}}}{2} & \text{if} \quad a_1 = a_4 \quad \text{and} \quad 0 < w_{\widetilde{A}} \le 1, \end{cases} \quad (22)$$

$$x^*_{\widetilde{A}} = \frac{y^*_{\widetilde{A}}(a_3 + a_2) + (a_4 + a_1)(w_{\widetilde{A}} - y^*_{\widetilde{A}})}{2w_{\widetilde{A}}}, \quad (23)$$

$P(\widetilde{A})$ and $P(\widetilde{B})$ are the perimeters of two generalized trapezoidal fuzzy numbers which are calculated as follows:

$$P(\widetilde{A}) = \sqrt{(a_1 - a_2)^2 + w^2_{\widetilde{A}}} + \sqrt{(a_3 - a_4)^2 + w^2_{\widetilde{A}}} + (a_3 - a_2) + (a_4 - a_1) \quad (24)$$

$$P(\widetilde{B}) = \sqrt{(b_1 - b_2)^2 + w^2_{\widetilde{B}}} + \sqrt{(b_3 - b_4)^2 + w^2_{\widetilde{B}}} + (b_3 - b_2) + (b_4 - b_1) \quad (25)$$

$A(\widetilde{A})$ and $A(\widetilde{B})$ are the areas of two generalized trapezoidal fuzzy numbers which are calculated as follows:

$$A(\widetilde{A}) = \frac{1}{2} w_{\widetilde{A}}(a_3 - a_2 + a_4 - a_1), \quad (26)$$

$$A(\widetilde{B}) = \frac{1}{2} w_{\widetilde{B}}(b_3 - b_2 + b_4 - b_1), \quad (27)$$

The larger the value of $S(\widetilde{A}, \widetilde{B})$ is, the more the similarity measure between two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ will be.

In the following sections, we will introduce some of properties that our model has:

Theorem 3.1. Two generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ are identical if and only if $S(\widetilde{A}, \widetilde{B}) = 1$.

Proof:

(i) If $\widetilde{A}$ and $\widetilde{B}$ are identical, $x^*_{\widetilde{A}} = x^*_{\widetilde{B}}$, $w_{\widetilde{A}} = w_{\widetilde{B}}$, $\min(P(\widetilde{A}), P(\widetilde{B})) = \max(P(\widetilde{A}), P(\widetilde{B}))$, $\min(A(\widetilde{A}), A(\widetilde{B})) = \max(A(\widetilde{A}), A(\widetilde{B}))$.

The degree of similarity between two generalized trapezoidal fuzzy numbers is calculated as follows:

$$\begin{aligned} S(\widetilde{A},\widetilde{B}) &= \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right]\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))} \\ &= [1-0]\times[1-0]\times 1=1 \end{aligned} \tag{28}$$

(ii) If $S(\widetilde{A},\widetilde{B})=1$, then

$$\begin{aligned} S(\widetilde{A},\widetilde{B}) &= \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right] \\ &\quad\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))}=1 \end{aligned} \tag{29}$$

It implies that $x^*_{\widetilde{A}}=x^*_{\widetilde{B}}$, $w_{\widetilde{A}}=w_{\widetilde{B}}$, $\min(P(\widetilde{A}),P(\widetilde{B}))=\max(P(\widetilde{A}),P(\widetilde{B}))$ and $\min(A(\widetilde{A}),A(\widetilde{B}))=\max(A(\widetilde{A}),A(\widetilde{B}))$. Therefore, the generalized trapezoidal fuzzy numbers $\widetilde{A}$ and $\widetilde{B}$ are identical. □

Theorem 3.2. $S(\widetilde{A},\widetilde{B})=S(\widetilde{B},\widetilde{A})$.

Proof: Because

$$S(\widetilde{A},\widetilde{B})=\left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right]\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))} \tag{30}$$

$$S(\widetilde{B},\widetilde{A})=\left[1-\left|x^*_{\widetilde{B}}-x^*_{\widetilde{A}}\right|\right]\times\left[1-\left|w_{\widetilde{B}}-w_{\widetilde{A}}\right|\right]\times\frac{\min(P(\widetilde{B}),P(\widetilde{A}))+\min(A(\widetilde{B}),A(\widetilde{A}))}{\max(P(\widetilde{B}),P(\widetilde{A}))+\max(A(\widetilde{B}),A(\widetilde{A}))} \tag{31}$$

We can see that $\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|=\left|x^*_{\widetilde{B}}-x^*_{\widetilde{A}}\right|$, $\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|=\left|w_{\widetilde{B}}-w_{\widetilde{A}}\right|$, $\min(P(\widetilde{A}),P(\widetilde{B}))=\min(P(\widetilde{B}),P(\widetilde{A}))$, $\min(A(\widetilde{A}),A(\widetilde{B}))=\min(A(\widetilde{B}),A(\widetilde{A}))$, and $\max(A(\widetilde{A}),A(\widetilde{B}))=\max(A(\widetilde{B}),A(\widetilde{A}))$.

Therefore, $S(\widetilde{A},\widetilde{B})=S(\widetilde{B},\widetilde{A})$. □

Theorem 3.3. If $\widetilde{A}=(a_1,a_2,a_3,a_4;w_{\widetilde{A}})$ and $\widetilde{B}=(b_1,b_2,b_3,b_4;w_{\widetilde{B}})$ are two generalized trapezoidal fuzzy numbers with the same geometric shape and height, then $S(\widetilde{A},\widetilde{B})=1-d$, where $d=\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|$ is the offset between $\widetilde{A}$ and $\widetilde{B}$.

Proof: Because $w_{\widetilde{A}}=w_{\widetilde{B}}$, and based on Eq.(24) - Eq.(27),

we can get $\min(P(\widetilde{A}),P(\widetilde{B}))=\max(P(\widetilde{A}),P(\widetilde{B}))$ and $\min(A(\widetilde{A}),A(\widetilde{B}))=\max(A(\widetilde{A}),A(\widetilde{B}))$;

Therefore, due to Eq.(21), the degree of similarity between $\widetilde{A}$ and $\widetilde{B}$ is calculated as follows:

$$\begin{aligned} S(\widetilde{A},\widetilde{B}) &= \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times\left[1-\left|w_{\widetilde{A}}-w_{\widetilde{B}}\right|\right]\times\frac{\min(P(\widetilde{A}),P(\widetilde{B}))+\min(A(\widetilde{A}),A(\widetilde{B}))}{\max(P(\widetilde{A}),P(\widetilde{B}))+\max(A(\widetilde{A}),A(\widetilde{B}))} \\ &= \left[1-\left|x^*_{\widetilde{A}}-x^*_{\widetilde{B}}\right|\right]\times[1-0]\times 1=1-d \end{aligned} \tag{32}$$

□

4 A Comparison of The Similarity Measures

In this section, we extend 15 sets of fuzzy numbers presented by Wei & Chen [18] into 18 sets of fuzzy numbers, as shown in Figure 3, and compare the calculation results of the proposed method with the results of the existing similarity measures, as shown in Table 1. From Figure 3 and Table 1, we can see the drawbacks of the existing similarity measures:

(1) From Figure 3, we can see that Set 3 and Set 4 are different sets of fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996) and Lee's method (Lee, 2002), Set 3 and Set 4 get the same degree of similarity, respectively.

(2) From Set 5 of Figure 3, we can see that $\widetilde{A}$ and $\widetilde{B}$ are different generalized fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996) and Lee's method (Lee, 2002), their result is a degree of similarity equal to 1, respectively, which is an incorrect result.

(3) From Set 6 of Figure 3 and Table 1, we can see that if we apply Lee's method (Lee, 2002), we cannot calculate the degree of similarity between two identical real values due to the fact that the denominator will become zero, such that $S(\widetilde{A}, \widetilde{B}) = \infty$, which is an incorrect result.

(4) From Set 7 of Figure 3 and Table 1, we can see that if we apply Lee's method (Lee, 2002), we can see that Lee's method cannot correctly calculate the degree of similarity between two identical real values due to the fact that the degree of similarity of the real values become zero, which is an incorrect result.

(5) From Set 8 and Set 9 of Figure 3, we can see that they are two different sets of fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996), they get the same degree of similarity, respectively, which does not coincide with the intuition of human being.

(6) From Set 10, Set 11 and Set 12 of Figure 3, we can see that they are different sets of generalized fuzzy numbers. However, from Table 1, we can see that if we apply Chen's method (Chen, 1996), they get the same degree of similarity, respectively, which does not coincide with the intuition of human being.

(7) From Set 7, Set 9 and Set 13 of Figure 3, we can see that $\widetilde{A}$ and $\widetilde{B}$ have the same shape and the offset $d = 0.1$ in the X-axis, respectively. By applying the proposed method, we can see that the proposed method has the good property that the degree of similarity between $\widetilde{A}$ and $\widetilde{B}$ is equal to$1 - |d| = 1 - 0.1 = 0.9$. However, from Table 1, we can see that if we apply Chen and Chen's method (Chen & Chen, 2003), the degree of similarity is equal to 0.81, which is an incorrect result.

(8) From Set 14 of Figure 3, using Chen's Method (Chen, 1996) and Lee's Method (Lee, 2002), the result is a degree of similarity equal to 1, respectively, which is an incorrect result.

(9) From Set 14 and Set 15 of Figure 3, we can see that Set 14 is more similar than Set 15 by the intuition of human being. However, from Table 1, we can see that if we apply Chen and Chen's method (Chen & Chen, 2003), we can see that it gets an incorrect result.

(10) From Figure 3, we can see that Set 10 and Set 16 are different sets of generalized fuzzy numbers and Set 10 is more similar than set 16 by the intuition of human being. However, from Table 1, we can see that if we apply the methods presented by Chen (1996), Lee (2002) and Hejazi et al. (2011), Set 10 and Set 16 get the same degree of similarity, respectively, and if we apply the method presented by Wei & Chen (2009), the result shows that Set 16 is more similar than Set 10. They are not the correct results.

(11) From Figure 3, we can see that Set 11 and Set 17 are different sets of generalized fuzzy numbers and Set 11 is more similar than Set 17 by the intuition of human being. However, from Table 1, we can see that if we apply the methods presented by Chen (1996), Lee (2002) and Hejazi et al. (2011), Set 11 and Set 17 get the same degree of similarity, respectively, and if we

apply the method presented by Chen & Chen (2003) and Wei & Chen (2009), the result shows that Set 17 is more similar than Set 11. They are not the correct results.

(12) From Figure 3, we can see that Set 11 and Set 18 are different sets of generalized fuzzy numbers. However, from Table 1, we can see that if we apply the methods presented by Chen (1996), Lee (2002), Hejazi et al. (2011) and Wei & Chen (2009), Set 11 and Set 18 get the same degree of similarity, respectively.

In summary, from Figure 3 and Table 1, we can see that the proposed method can overcome the drawbacks of the existing similarity measures.

Table 1: The fuzzy model of attribute SL

Sets	Chen's method (1996)[15]	Lee's method (2002)[16]	Chen & Chen's method (2003)[17]	Wei & Chen's method (2009)[18]	Hejazi et al. method (2011)[19]	Proposed method
Set 1	0.9750	0.9617	0.8357	0.9500	0.9004	0.9473
Set 2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Set 3	**0.7000**	**0.5000**	0.4200	0.6820	0.6465	0.6631
Set 4	**0.7000**	**0.5000**	0.4900	0.700	0.7000	0.7000
Set 5	**1.0000**	**1.0000**	0.8000	0.8248	0.6681	0.6659
Set 6	1.0000	*	1.0000	1.0000	1.0000	1.0000
Set 7	0.9000	**0.0000**	**0.8100**	0.9000	0.9000	0.9000
Set 8	**0.9000**	0.5000	0.5400	0.8411	0.3700	0.3896
Set 9	**0.9000**	0.6667	**0.8100**	0.9000	0.9000	0.9000
Set 10	**0.9000**	**0.8333**	0.9000	**0.7833**	**0.6261**	0.7731
Set 11	**0.9000**	**0.7500**	**0.7200**	**0.8003**	**0.6448**	0.7938
Set 12	**0.9000**	0.8000	0.8325	0.8289	0.7361	0.7478
Set 13	0.9000	0.7500	**0.8100**	0.9000	0.9000	0.9000
Set 14	**1.0000**	1.0000	**0.7000**	0.7209	0.5113	0.5104
Set 15	0.9500	0.7500	**0.9048**	0.6215	0.3830	0.4242
Set 16	**0.9000**	**0.8333**	0.7425	0.8140	0.6261	0.7321
Set 17	**0.9000**	**0.7500**	**0.8911**	**0.8380**	**0.6448**	0.7432
Set 18	**0.9000**	**0.7500**	0.6976	**0.8003**	**0.6448**	0.7144

Note. "*" means that the similarity measure cannot calculate the degree of similarity between two generalized fuzzy numbers and the results that are not satisfactory are given in bold.

5 A New Method to Obtain BPA

In fact, some samples exist in many systems, which often approximatively submit the triangular distribution. Therefore, we use the existing sample data to build a triangular distribution to describe model of attribute categories, and then generate the BPA function based on the similarity between the collected attribute and the model attribute.

In order to be understood easily, the following Iris data classification problem shows the detailed approach of the proposed method.

In the Iris data, there are 3 species of Iris flower, i.e., *Setosa*, *Versicolor*, and *Virginica* [21]. The Iris data contain 150 instances, and each species contains 50 instances. There are four

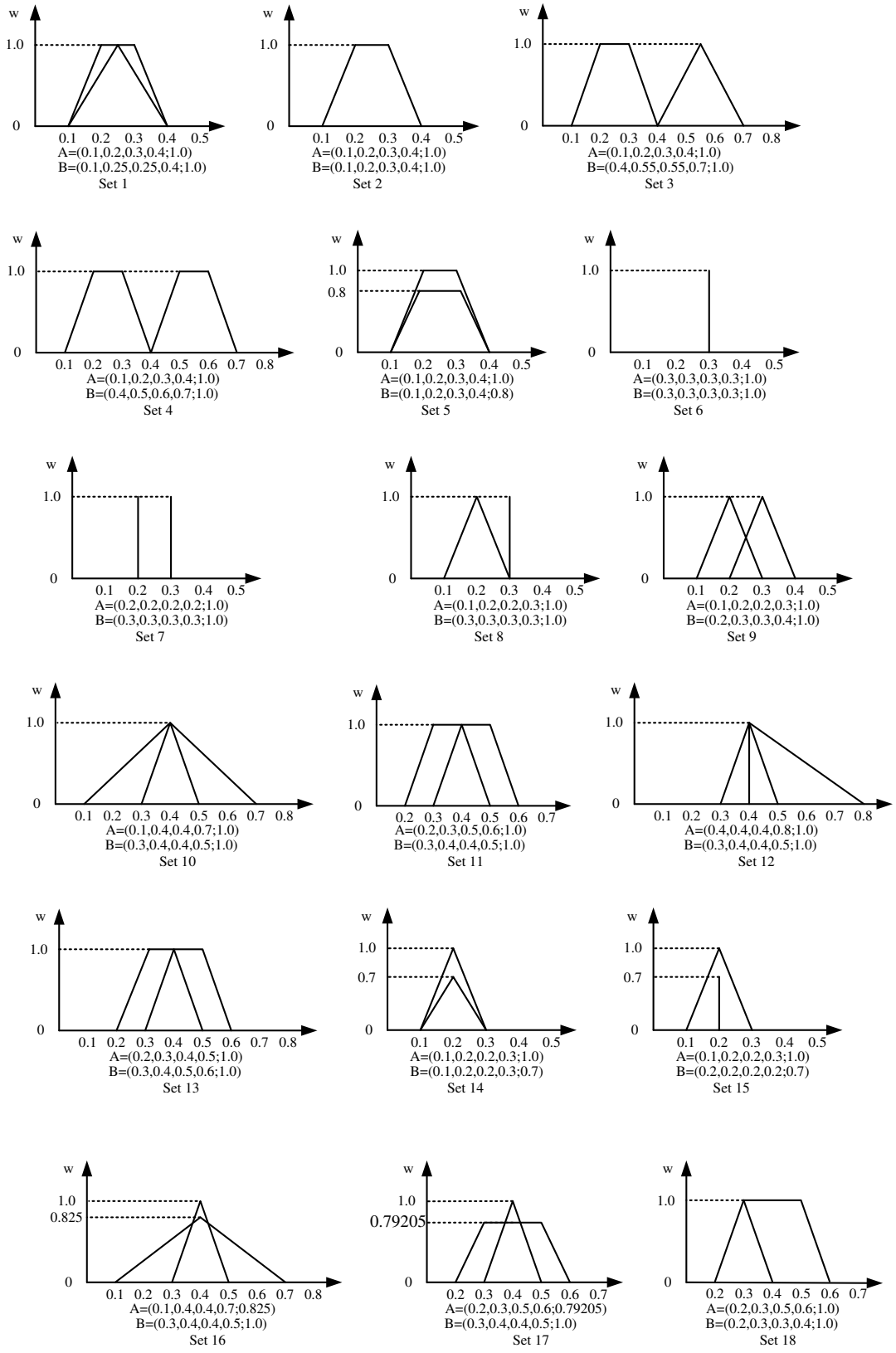


Figure 3: Eighteen sets of generalized fuzzy number $\widetilde{A}$ and $\widetilde{B}$

attributes in the Iris data, i.e., Sepal Length (SL), Sepal Width (SW), Petal Length (PL), and Petal Width (PW).

We randomly chose 40 instances from *Setosa*, the min(SL)=4.30; the average(SL)=5.04; the max(SL)=5.80 can be obtained. Hence, we can construct the fuzzy model of SL attribute of *Setosa* in Figure 4. In the same way, we can construct fuzzy models of Sepal Length(SL) of *Versicolor* and *Virginica*, as shown in Table 2 and Figure 4.

As can be seen from Figure 4, there are some crossing areas. For example, the crossing area of fuzzy number of *Setosa and Versicolour* can be shown in Figure 5. All the crossing areas can be modeled as generalized fuzzy numbers shown in Table 3.

Table 2: The fuzzy model of attribute SL

Species	Setosa(S)	Versicolor(C)	Virginica(V)
Attribute (SL)	(4.30,5.04,5.80;1.0)	(5.0,5.90,6.8;1.0)	(5.6,6.59,7.90;1.0)

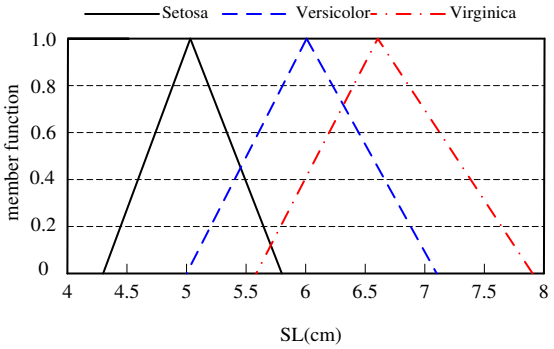


Figure 4: The fuzzy number representation of SL attribute of each species

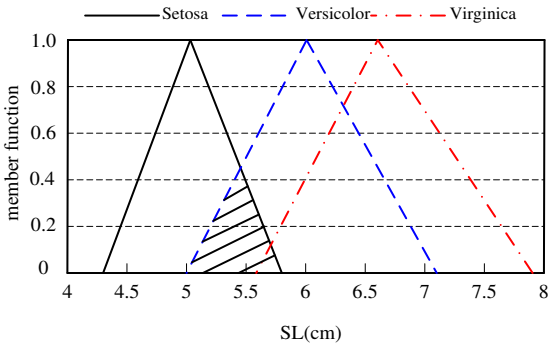


Figure 5: The generalized fuzzy number model of crossing area of *Setosa* and Versicolour

In a similar way, the fuzzy models of Sepal Width (SW) attribute, Petal Length (PL) attribute, and Petal Width (PW) attribute of each species can be constructed.

Table 3: The generalized fuzzy number model of crossing area of three species Iris flowers

Species	S&C	S&V	V&C	S&C&V
Attribute (SL)	(5.00,5.43,5.80;0.48)	(5.60,5.71,5.80;0.11)	(5.60,6.23,6.80;0.63)	(5.60,5.71,5.80;0.11)

We randomly chose a datum from Iris source; for example, a new instance (NI) of *Setosa* could be shown as (5.1cm, 3.5cm, 1.4cm, 0.2cm). In Figure 6, the relation between SL attribute of NI and the fuzzy number representation of SL attribute of each species is distinctly shown.

To calculate the similarities between NI and each generalized fuzzy number and normalize the obtained similarities, the BPA can be obtained, as shown in table 4.

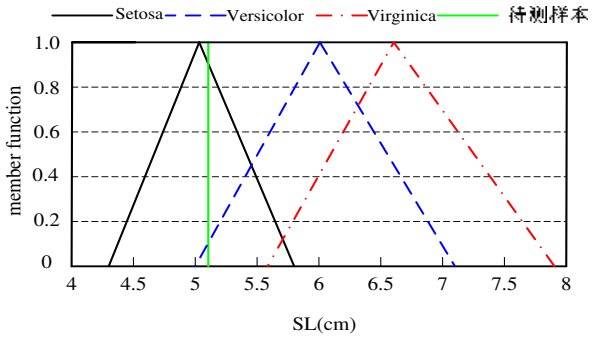


Figure 6: The relation between the new instance and the model of SL attribute of each species

Table 4: The BPA of an instance attribute SL

	m(S)	m(C)	m(V)	m(S,C)	m(V,S)	m(C,V)	m(S,V,C)
SL	0.2955	0.2605	0.2215	0.0842	0.0046	0.1291	0.0046

In the same way, the four attributes and their relative BPAs can be obtained, as shown in Table 5.

With the combination rule in Eq.(3) used, the fusion results can be shown as follows:

$$m(S) = 0.5324, \quad m(C) = 0.2607, \quad m(V) = 0.1879, \quad m(S,C) = 0.0112,$$
$$m(V,S) = 0.0015, \quad m(C,V) = 0.0063, \quad m(S,V,C) = 0.0001$$

Hence, the instance can be classed as *Setosa*. The result is consistent with the actual situation.

The algorithm of our proposed method can be listed step by step as follows.

Step 1: Use the existing sample data to obtain the min, average and max value to construct the triangular fuzzy models, which describe the model attributes of instances.

Step 2: Calculate the similarity between the collected attribute and the model attribute.

Step 3: Normalize the similarity measure to obtain the BPA function.

We randomly selected 120 instances, 40 instances for each of the 3 species, to construct species models. The remaining 30 instances, 10 instances for each of the 3 species, were used as collected instances whose class was unknown. By applying the method proposed in the above section to classify the 30 instances, the correct rate of Iris data classification was calculated as 96.67%.

Table 5: The BPA of an instance

	m(S)	m(C)	m(V)	m(S,C)	m(V,S)	m(C,V)	m(S,V,C)
SL	0.2955	0.2605	0.2215	0.0842	0.0046	0.1291	0.0046
SW	0.1727	0.1614	0.1687	0.0925	0.1510	0.1611	0.0925
PL	0.2198	0.1193	0.0731	0.1943	0.1943	0.0049	0.1943
PW	0.1921	0.0988	0.0499	0.2143	0.2143	0.0162	0.2143

Further, we applied the proposed method 10 times; the average correct rate of Iris data classification was up to 95.67%. It can be seen that our proposed method has good results in classification problem.

6 Conclusions

The estimation of BPA plays a very important role in the application of Dempster-Shafer theory in complex uncertain problems. The fusion performance depends on the method of BPA construction. To solve this problem, firstly, the paper presents an improved method to calculate the degree of similarity between generalized fuzzy numbers, which gives consideration to the horizontal center of gravity, the perimeter, the height and the area of the two fuzzy numbers. The method can overcome the drawbacks of the existing similarity measures. Then, a new method to obtain BPA is proposed based on the improved similarity measure between generalized fuzzy numbers. The proposed method to obtain BPA can effectively overcome the problem of subjectivity, which also has strong generality. The classification of Iris data is used to illustrate the efficiency and the low computational complexity of the proposed method. The proposed method provides a simple technique that will help to use the classical Dempster combination rule effectively.

Acknowledgment

This work was supported in part by a grant from National Natural Science Foundation of China (No. 61104214) and Foundation for Fundament Research of Northwestern Polytechnical University, Grant No.JC20120235.

Bibliography

[1] A. Dempster(1967), Upper and lower probabilities induced by multivalued mapping, *Annals of Mathematical Statistics*, ISSN 0003-4851, 38(2): 325-339.

[2] G. Shafer(1976), *A mathematical theory of evidence*, Princeton University Press, ISBN 978-069-11-0042-5.

[3] Y.M. Zhu, L. Bentabet, M. Rombaut, O. Dupuis, V. Kaftandjian, D. Babot(2002), Automatic determination of mass functions in DS theory using FCM and spatial neighbourhood information for image segmentation, *Optical Engineering*, ISSN 0091-3286, 41(4): 760-770.

[4] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski(2001), Multisensor image segmentation using DS fusion in Markov fields context, *IEEE Trans. Geosci. Remote Sensing*, ISSN 0196-2892, 39(8): 1789-1798.

[5] X. Guan, X. Yi, Y. He(2008), Study on algorithms of determining basic probability assignment function in Dempster-Shafer evidence theory, *Proc. of the 7th Int. Conf. on Machine Learning and Cybernetics*, 121-126.

[6] B. Chen, J.F. Wang, S.B. Chen(2010), Prediction of pulsed GTAW penetration status based on BP neural network and D-S evidence theory information fusion, *International Journal of Advanced Manufacturing Technology*, ISSN 0268-3768, 48(1-4): 83-94.

[7] X.M. Li, L.X. Ding, Y. Li, G. Xu, J.B. Li(2009), HVAC fan mechinery fault diagnosis based on ANN and D-S evidence theory, *IITA Int. Conf. on Control, Automation and Systems Engineering*, Zhangjiajie, China, 603-606.

[8] Z. Xu, M. Liu, G. Yang, N. Li(2009), Application of interval analysis and evidence theory to fault location, *IET Electric Power Application*, ISSN 1751-8660, 3(1): 77-84.

[9] Z.Y. Zuo, Y.F. Xu, G.C. Chen(2009), A new method of obtaining BPA and application to the bearing fault diagnosis of wind turbine, *Proc. of the 2009 Int. Symposium on Information Processing*, Huangshan, China, 368-371.

[10] W. Jiang, J.Y. Peng, Y. Deng(2011), A new method to determine BPA in evidence theory, *Journal of Computers*, ISSN 1796-203X, 6(6): 1162-1167.

[11] Y. Deng, W. Jiang, R. Sadiq(2011), Modeling contaminant intrusion in water distribution networks: A new similarity-based DST method, *Expert Systems with Applications*, ISSN 0957-4174, 38(1): 571-578.

[12] Y. Deng, R. Sadiq, W. Jiang, S. Tesfamariam(2011), Risk analysis in a linguistic environment: A fuzzy evidential reasoning-based approach, *Expert Systems with Applications*, ISSN 0957-4174, 38(12): 15438-15446.

[13] Y. Deng, W. Jiang, X.B. Xu(2009), Determinging BPA under uncertainty environments and its application in data fusion, *Journal of Electronics (China)*, ISSN 0217-9822, 26(1): 13-17.

[14] W. Jiang, Y. Luo, X.Y. Qin, J. Zhan(2015), An Improved Method to Rank Generalized Fuzzy Numbers with Different Left Heights and Right Heights, *Journal of Intelligent and Fuzzy Systems*, Accepted.

[15] S. M. Chen(1996), Foreword, New methods for subjective mental workload assessment and fuzzy risk analysis, *Cybernetics and Systems: An International Journal*, ISSN 0196-9722, 27(5): 449-472.

[16] H.S. Lee(2002), Optimal consensus of fuzzy opinions under group decision making environment, *Fuzzy Sets and Systems*, ISSN 0196-9722, 132(3): 303-315.

[17] S.J. Chen, S.M. Chen(2003), Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *IEEE Transaction on Fuzzy Systems*, ISSN 1063-6706, 11(1): 45-56.

[18] S.H. Wei, S.M. Chen(2009), A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *Expert Systems with Applications*, ISSN 0957-4174, 36(1): 589-598.

[19] S.R. Hejazi, A. Doostparast, S.M. Hosseini(2011), An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers, *Expert Systems with Applications*, ISSN 0957-4174, 38(8): 9179-9185.

[20] W. Jiang, X. Fan, D.J. Duanmu, Y.Deng(2011), A Modified Similarity Measure of Generalized Fuzzy Numbers, *2011 Int. Conf. on Advanced in Control Engineering and Information Science*, 2773-2777.

[21] R.A. Fisher(1936), The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, ISSN 1469-1809, 7(2): 179-188.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):348-356, June, 2015.

Collaborative Data Processing in WSN Using Voronoi Fuzzy Clustering

S. Nithya Kalyani, E. Sasikala B. Gopinath

S. Nithya Kalyani*
Department of Information Technology
K.S.R College of Engineering
Tiruchengode, India
*Corresponding author: nithyakalyani.me@gmail.com

E. Sasikala
Department of Information Technology
K.S.R College of Engineering
Tiruchengode, India

B. Gopinath
Department of Electrical and Electronic Engineering
Vivekanandha Institute of Engineering and Technology for Women
Tiruchengode, India

Abstract: In this paper, developed a novel Voronoi Fuzzy Clustering (VF) algorithm for energy efficient collaborative data aggregation in wireless sensor network. VF algorithm is fusion of Voronoi diagram and modified Fuzzy C- Means with respect to distance and Quality of Service. Here throughput, delay time and delivery ratio are considered as QOS parameters. Once clustering of sensor nodes is completed then data management technique such as data aggregation or compression is done for further decision making in sink node. Data mining clustering algorithm reduces overall transmission of data from each sensor to the sink node thus energy spent by individual sensor node is minimized. The cluster heads collects all sensed information from their respective cluster members and performs data aggregation or compression before transmitting the data to the sink node. Finally, the simulations are performed and the results are analyzed within the simulation set up to determine performance of the proposed algorithm in the sensor network. Our proposed approach has achieved 60% efficiency when compare with the K means algorithm.
Keywords: Clustering, Data aggregation, Compression, Voronoi fuzzy clustering algorithm, Energy, QOS, Throughput, Delivery ratio.

1 Introduction

Cluster aggregation is essential techniques that naturally reduce energy costs in wireless sensor networks (WSN) without compromising the quality of data delivery. The process of separating the sensor nodes into groups are called as clustering. A number of challenges involved in clustering. At first, the clusters themselves have to be identified. Secondly, cluster heads have to be chosen. Thirdly, routes have to be discovered from every node to their cluster head, and finally, the cluster heads have to proficiently relay the data to the sink node. This paper focuses all the later three problems. The foremost problem is defined by the application domain.

Data aggregation is another vital function in WSN to reduce the consumption of energy due to the number of limitation of WSN. The key idea of this process is to eliminating redundancy in data, minimizing the number of transmissions via integrate all the incoming data in cluster head from diverse sources and enroute it to the sink. This focuses in data-centric approach. Aggregation algorithms are limited to application requirement that is either in time or energy performance

A WSN consists of tiny sensing devices, which normally run on battery power. Sensor nodes are densely deployed in the region of interest. Each device has sensing and wireless communication capabilities, which enable it to sense and gather information from the environment and then send the data and messages to other nodes in the sensor network or to the remote base station[4].Wireless sensor networks have been envisioned to have a wide range of application in both military and civilian domains [1]. Due to the limitation of sensor node researchers have designed lots of energy-efficient routing protocols to prolong the lifetime of sensor networks [2]. In this paper, the Fuzzy C-Means (FCM) clustering algorithm is modified by incorporating with Voronoi diagram for clustering the sensor node based on the distance and the Quality of Services.

2 VF Algorithm for Cluster Data Aggregation

The ultimate aim of the proposed approach is to reduce the consumption of the energy and increase the life time of WSN using collaborative data processing. In order to reduce the energy of the every node, clustering algorithm is used where the sensor nodes are clustered and subsequently, the cluster heads are selected. The cluster head is used to transmit the sensed data of sensor nodes to the base station (sink node), subsequently from the cluster head, the data is processed by aggregation or compression. In this approach, the process is separated into two phases.

2.1 VF algorithm to select the cluster head

The wireless sensor nodes are located at different places and the sensors are clustered using the Voronoi fuzzy clustering algorithm. Initially, apply the Voronoi tessellation to the sensor nodes based on the position and energy of the sensor. Subsequently use fuzzy clustering algorithm to cluster the nodes as it's easy to cluster and select the cluster head after applying the Voronoi. For finding the membership function, select any number of Voronoi cells as a cluster head and calculate the membership for every Voronoi cells with the assumed cluster head. The Voronoi cell goes to cluster which has highest value membership function. With the help of the membership function, the cluster head among the clusters are calculated.

Pseudo code:

Input:

No. of sensor $S = \{s_1, s_2, ...s_i\}$, Position and energy of every sensors$s_i = (x_i, y_j, z_k)$

Algorithm:

Step 1: Find the distance for each s_i with r: $d(s_i, r) = \sqrt{(x_s - x_r)^2 + (y_s - y_r)^2 + (z_s - z_r)^2}$

Step 2: $\{r \,|d(s_i, r) \leq d(s_j, r), i \neq j\}$,where r = set of neighboring sensor

Step 3: $r = (r_1, r_2....r_n)$

Step 4: $V(s_i) = \frac{d(s_i, r_j)}{2}$

Step 5: Get the value of "c"

Step 6: Select cluster heads ch_y

Step 7: Find the member ship function $\mu_{xy} = \frac{1}{\alpha+\beta}[\alpha\ (D_{xy}) + \beta\ (Q_{xy})]$

Step 8: Find distance membership function using $D_{xy} = \left(\frac{\|v_x - ch_y\|}{\sum_{k=1}^{c}\|v_x - ch_k\|}\right)$

Step 9: Find QOS membership function using $Q_{xy} = \frac{\sum_{z=1}^{Q} v_x^y(z)}{\left(\sum_{k=1}^{C}\prod_{z=1}^{Q} v_x^k(z)\right)}$

Step 10: Find max of μ_{xy} for each s_i

Step 11: s_i max (μ_{xy})

Step 12: Find $ch_y = \frac{\sum_{x=1}^{N} \mu_{xy}^m . s_i}{\sum_{x=1}^{N} \mu_{xy}^m}$

Step 13: Update ch_y
Step 14: Go to step 5 until $ch_y$is placed in same $V(s_i)$
Step 15: Each s_i send data $s_i(d)$ to corresponding cluster head ch_i
Step 16: The cluster head ch_i receives $D = (d_1, d_2, ..., d_i)$
Step 17: Select Option 1. Data aggregation 2. Data compression
Step 18: If (1)
Select option a. Avg b. Max c. Min
a. $avg = \frac{\sum_{n=1}^{n}(d_i)}{N}$
b. $\max = \max(d_i)$
c. $\min = \min(d_i)$
Step 19: If (2)
a. Find probability of received data $P(d_i) = \frac{F(d_i)}{N}$
b. Sort the data into ascending order $P(D) = (P(d_1), P(d_2), ..., P(d_i))$
c. Add the first two data $(P(d_1) + P(d_2))$
d. Merge the id of the data
e. Repeat the process 2 and 3 until getting only 2 probability of data
Step 20: The result of (18) and (19) goes to sink node.

Let S sensor nodes is defined by $S = \{s_1, s_2, ...s_i\}$ and $(1 \leq i \leq n)$ is the total number of sensors in the wireless network. Each sensor node is defined by the position values (x coordinate and y coordinate) and the energy value (z). The position and the energy of the sensor nodes are given by $s_n = (x_i, y_j, z_k)$. A Voronoi diagram is a set of triangles, called Voronoi triangles. The Voronoi triangles generation is based on the distance and the energy of the sensor node. Find r the neighbors of the each sensor node based on the minimum distance as in Step 2,3,4, where $d(s_i, r)$ is the distance from point s_i to r, predicted using Step 1. That is r is the set of sensors closer to s_i than to any other s_j.

2.2 Calculation of membership function

From the entire sensor nodes, clustering is performed in the sensor nodes based on the membership function. To find the membership function of each sensor node, let assume certain the number of sensor node as cluster head. In order to find the membership function, first choose 'c' cells as a Cluster Head (CH) that is $c = 2$ and assume any two cells as cluster head $y_1 = ch_1 = (x_{i'1}, y_{j1}, z_{k1})$ and $y_2 = ch_2 = (x_{i2}, y_{j2}, z_{k2})$. With the help of assumed cluster head calculate the membership function of each sensor point. Based on the membership functions the nodes are clustered. In this paper, the membership function is attained based on two functions first one is based on the distance and second one is based on the QOS values. The equation in Step 7 is use for finding the membership function, where μ_{xy} is the membership function of x^{th} sensor node with respect to the y^{th} cluster head node, D_{xy} is the distance between x^{th} sensor node to the y^{th} cluster head node, Q_{xy} is the value of the QOS of x^{th} sensor node with respect to the y^{th} cluster head node, α is the weightage of distance (user defined), β is the weightage of QOS (user defined).

2.3 Distance based membership function

Based on the Euclidian distance between the cluster head and sensor node distance between the two points are calculated. This calculation describes, how much faraway the each sensor nodes from the each cluster head node. The Step 8 is used for calculating the distance based membership function.

Consider Table 1, which describes the positions of the each sensor node, form that select any number of the nodes as a cluster head first. Based on the selected cluster head, distance based membership function is found. Here, C1= (9.5, 1.3), C2= (8.8, 2.5) are select as a cluster head, based on that the distance based membership functions are calculated for other sensors. The distance based membership function using Step 8 for each sensor with respect to the C1 and C2 is calculated and the results are provided in Table 2.

Table 1: The positions of the sensor nodes

	x	y
V1	2.5	2
V2	4.1	1.3
V3	5.4	0.8
V4	9.5	1.3
V5	8.8	2.5

Table 2: Distance based membership of the sensor nodes

	V1(2.5,2)	V2(4.1,1.3)	V3(5.4,0.8)	V4(6.8,3.0)	V5(8.8,2.5)
V1(2.5,2)	0	0.2	0.3	0.4,	0.6
V2(4.1,1.3)	0.2	0	0.1	0.2	0.4
V3(5.4,0.8)	0.3	0.1	0	0.1	0.3
V4(6.8,3.0)	0.4	0.2	0.1	0	0.2
V5(8.8,2.5)	0.6	0.4	0.3	0.2	0

2.4 QOS based membership

Consider the following Table 3 that contains the values of the QOS of the each node with respect to each cluster head. Each of the sensor nodes has different quality parameter with respect to application where the sensors are applied. For this work QOS functions are throughput, delivery ratio, delay time are taken. Based on those QOS values of the nodes the cluster head can be obtained. Each of the node have the time slots at the beginning time, after generation of data packet the nodes are send to the cluster head within the next time slot, if the packets are not deliver in the particular time period then it marked as the delayed packet. The delay time is calculated by the ratio of the received time to the sending time of the node; the result of this subtracted by 1 that is a delay time.

Each of the nodes has the number of packets to send to the cluster head while sending the data packet to the cluster head some of the packets are not receive in the cluster head because of the interruption, noise. The ratio of number of received data packet in the cluster head to the number of sent data packet from the node.

Each of the nodes generates the data packets and sends it to the cluster head, flow of the data packet is varies from node to node with respect to the cluster head. The throughput of a node is calculated using number of packets sends to the cluster head in a particular time interval and it is different for each node with respect to the cluster.

The membership value based on the QOS values of the node with respect to the cluster heads are discovered by Step 9 and it is given in Table 4.

Table 3: The QOS values of each node with all possible cluster heads

	V1(2.5,2)	V2(4.1,1.3)	V3(5.4,0.8)	V4(6.8,3.0)	V5(8.8,2.5)
V1(2.5,2)	0,0,1	0.6,0.4,0.7	0.2,0.6,0.8	0.5,0.8,0.3	0.4,0.4,0.5
V2(4.1,1.3)	0.2,0.5,0.8	0,0,1	0.4,0.8,0.2	0.3,0.5,0.6	0.7,0.4,0.5
V3(5.4,0.8)	0.4,0.8,0.2	0.4,0.5,0.6	0,0,1	0.2,0.5,0.8	0.4,0.4,0.5
V4(6.8,3.0)	0.4,0.2,0.8	0.5,0.5,0.4	0.4,0.5,0.4	0,0,1	0.4,0.2,0.8
V5(8.8,2.5)	0.4,0.8,0.2	0.3,0.4,0.5	0.6,0.4,0.5	0.4,0.2,0.8	0,0,1

Table 4: The QOS membership of the nodes

	C1	C2
V1	0.55	0.44
V2	0.46	0.53
V3	0.48	0.51
V4	0.51	0.48
V5	0.45	0.54

With the help of D_{xy} and Q_{xy} values calculate the membership function using Step 7. Here equal weightage is given for both the distance and the QOS and calculated values are given in Table 5. The values of the weightage depends on the user. If the user wants to group the sensor nodes mainly based on distance then the user give the value of α is greater than β, if the user wants to group the sensor node based on the QOS then the value of β is greater than α.

Table 5: Sensor nodes membership function

	μ_{11}	μ_{12}
V1	0.535	0.455
V2	0.49	0.50
V3	0.50	0.48
V4	0.25	0.74
V5	0.72	0.27

Consider the two membership value μ_{xy_1}, μ_{xy_2}, where x be the sensor node, y_1 and y_2 be the cluster heads. If value of μ_{xy_1} minimum than μ_{xy_2} then the sensor node x moves to the cluster head y_2 else x move to y_1. In Figure 1 dark shaded cells illustrates the assumed cluster head in a wireless sensor network. After finding the membership value for all Voronoi cells, each and every cell is clustered based on the two assumed cluster heads. After the selection of the cluster head and its members, the cluster head transmits the received sensed data from the members to the sink and this will lead to more of time and energy consumption in CH. To solve this problem, data management process is achieved.

2.5 Data Management process

Data management process involves data aggregation or compression. Data aggregation is a method which aims to reduce the localized congestion problem. It tries to gather knowledge from the sensored data by applying the maximum, minimum, average functions using Step 18.

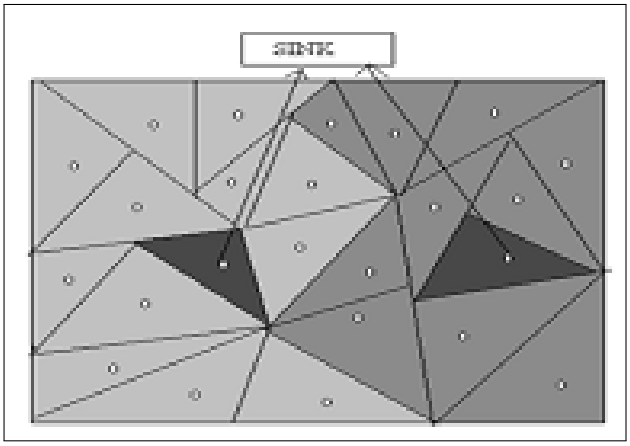


Figure 1: Transmission of aggregated or compressed data to the sink node

It then transmits only the useful information (aggregated data) to the sink node thus reducing congestion and its linked problems. Thus data aggregation helps the life of the wireless network to get increased. The cluster head sends the aggregated data to the sink node for the particular time interval. On the otherhand the collected sensed data are compressed in the cluster head by using the huffman code as in Step 19.

3 Experimental Results and Discussion

The experimental results of the proposed approach to cluster head selection is described in this section. The comparative analysis of the proposed cluster head selection approach and K-Means algorithm is presented. The proposed cluster head selection for data aggregation in wireless sensor network has been implemented using MATLAB and the performance of the proposed system is analyzed using the evaluation metrics including running time and size of received data in the sink node based on the number of cluster head in the wireless sensor network.

3.1 Evaluation of running time by varying the number of the sensor nodes

The VF algorithm is compared with the K-Means algorithm in order to evaluate the performance of the proposed VF algorithm. The following Figure 2 shows that the running time of the VF algorithm and K-Means algorithm. To evaluate the graph we change the number of sensor nodes at each execution and evaluate the time of the both algorithm, the value of the cluster head is constant for both algorithm to find the execution time. Here number of cluster heads is considered as 10 for all execution. The following Figure 2 shows that, VF algorithm needs less time to execute when compared with the K-Means algorithm. the reason behind is the if the membership function value of the sensor node is repeat in a single Voronoi cell then the execution of the membership function for that sensor node get stop since the running time of the VF algorithm get reduced.

3.2 Evaluation of running time by varying the number of cluster heads

Here, the running time is evaluated by keep the number of sensor nodes (1000) at constant and varying the number of cluster heads for each time. The following Figure 3 shows that the running time of VF algorithm and K-Means algorithm and when we see that graph we conclude the VF algorithm takes less to execute when compare to the K-Means algorithm. The running time of

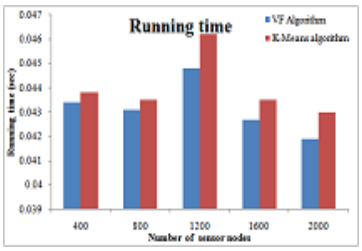


Figure 2: Running time VS No. of nodes

the both algorithm is directly proportional to the number of cluster heads. The calculation of the membership function for each point with respect to the each cluster head takes more time since the running time is increase when the number of cluster head increase.

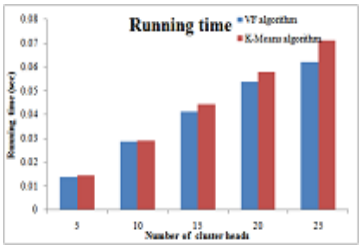


Figure 3: Running time VS No. of clustering head

3.3 Number of received data in sink node by varying the number of the sensor nodes

After the selection of cluster head, the each of the sensor node sends their sensed data to the cluster head. The cluster head received many data's from the sensor nodes, as per the user suggestion (data aggregation or data compression) the cluster head operate the received data and send that data to the sink node. The sink node receives the data from the cluster heads, the Figure 4 shows that the number of received data in a sink node for different number of sensor nodes. Based on the position of the cluster head, the received data of the sink node get varied since the position of the cluster head is different for both VF algorithm and K-Means algorithm. Hence, we take the number of received data in a sink node as a parameter. The following graph describes, the sink node receives more number of data's by using VF algorithm rather than using the K-Means algorithm.

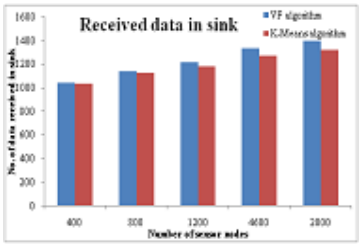


Figure 4: No. of data received in sink node Vs No. of nodes

3.4 Number of received data in sink node by varying the number of the cluster heads

The number of received data in a sink node is varies based on the number of cluster heads in a sensor network. Figure 5 shows that the number of received node in a sink node based on different number cluster heads for both VF algorithm and K-Means algorithm, the sink node receives more number of data's by using VF algorithm rather than using the K-Means algorithm.

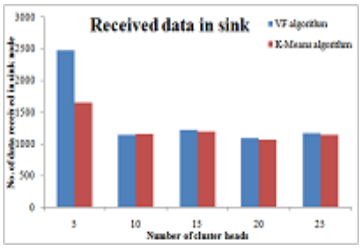


Figure 5: Size of data received in sink node Vs No. of cluster head

4 Conclusions

In this paper, we have presented the approach for energy efficient cluster algorithm for data summarization in wireless sensor network. In order to reduce the energy of the wireless sensor network, we have presented the efficient Voronoi fuzzy clustering algorithm. At first the Voronoi cells are applied for each node in the wireless sensor network based on the energy of the sensor node. Consequently, the cluster head is selected by the fuzzy clustering algorithm. Every node in the wireless sensor network sends their sensed data to the cluster head. The data aggregation and data compression process is done in the cluster head. The cluster head aggregate the received data by taking the value of minimum, maximum and average and uses the Huffman compression algorithm to compress the data and that compressed or aggregated data will send to the sink node. Finally, the experiment has carried out and results ensured that the efficiency (60%) is higher when compare with the K means algorithm.

Bibliography

[1] Kulik, J.; Heinzelman, W; Balakrishnan, H. (2002); Negotiation-based protocols for disseminating information in wireless sensor networks, *Wireless Networks*, 8:69-185.

[2] Xianghui Wang; Guoyin Zhang (2007); DECP: A Distributed Election Clustering Protocol for Heterogeneous Wireless Sensor Networks, *Computational Science*, 4489:105-108.

[3] Kim, J.M; Park,S.H; Han,Y.J; Chung,T.M (2008); CHEF: cluster head election mechanism using fuzzy logic in Wireless Sensor Networks, in *International Conference of Advanced Communication Technology*, 654-659.

[4] Mohammad Zeynali; Leili Mohammad Khanli; Amir Mollanejad (2009); TBRP: Novel Tree Based Routing Protocol in Wireless Sensor Network, *International Journal of Grid and Distributed Computing*, 2(4): 35-48.

[5] Ye, M.; Li, C.F.; Chen, G.; Wu, J.(2005); EECS: An Energy Efficient Clustering Scheme in Wireless Sensor Networks, In *Proc. of the IEEE International Performance Computing and Communications Conference*, 535-540.

[6] Tian, D.; Georganas, N.D.(2002); A Node Scheduling Scheme for Energy Conservation in Large Wireless Sensor Networks, *From Thesis: Multimedia Communications Research Laboratory, School of Information Technology and Engineering*, University of Ottawa.

[7] Guru, S.M.; Steinbrecher,M; Halgamuge,S; Kruse,R.(2007); *Multiple Cluster Merging and Multihop Transmission*, LNCS 4459: AGPC, Springer, 89-99.

[8] Qingchao Zheng; Liu,Z.; Liang Xue; Yusong Tan; Dan Chen; Xinping Guan(2010); An Energy Efficient Clustering Scheme with Self organized ID Assignment for Wireless Sensor Networks, *16th IEEE International Conference on Parallel and Distributed Systems*, DOI: 10.1109/ICPADS.2010.83, 635-639.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):357-369, June, 2015.

An Application of Latent Semantic Analysis for Text Categorization

G. Kou, Y. Peng

Gang Kou
School of Business Administration
Southwestern University of Finance and Economics, Chengdu, China
No.555, Liutai Ave, Wenjiang Zone
Chengdu, 611130, China
kougang@swufe.edu.cn

Yi Peng*
School of Management and Economics
University of Electronic Science and Technology of China, Chengdu, China
No.2006, Xiyuan Ave, West Hi-Tech Zone
Chengdu, 611731, China
*Corresponding author: pengyicd@gmail.com

Abstract: It is a challenge task to discover major topics from text, which provide a better understanding of the whole corpus and can be regarded as a text categorization problem. The goal of this paper is to apply latent semantic analysis (LSA) approach to extract common factors that representing concepts hidden in a large group of text. LSA involves three steps: the first step is to set up a term-document matrix; the second step is to transform the term frequencies into a term-document matrix using various weighting schemes; the third step performs singular value decomposition (SVD) on the matrix to reduce the dimensionality. The reduced-order SVD is the best k-dimensional approximation to the original matrix. The experiment uses more than fifteen hundreds research paper abstracts from a specific field. Because different factor solutions of the LSA suggest different levels of aggregation, this work examines thirteen solutions in the experiment. The results show that LSA is able to identify not only principle categories, but also major themes contained in the text.
Keywords: Latent Semantic Analysis, Topic extraction, Text Mining, Information Retrieval.

1 Introduction

Many multidisciplinary fields, such as data mining, bioinformatics, biochemistry, and neuroscience, emerge in the past several decades. Since multidisciplinary fields involve theories, methods, and techniques from multiple disciplines, it is not easy to comprehend all the research efforts in these fields. Text categorization, which organizes documents into groups based on their underlying structures, can help capturing the large amount of activities and diversity of a multidisciplinary field.

The goal of this paper is to apply latent semantic analysis (LSA) approach to detect major research topics and themes of a multidisciplinary field. In particular, it is intended to address three questions: what are the core research areas of the selected field, what are the major research themes, and what is the dynamics of the discipline? LSA is an automatic mathematical and statistical technique for uncovering common factors that representing concepts hidden in text[1,2,3,4]. Previous investigations in psychology and computer science have proved that LSA resembles the way the human brain distills meaning from text and is capable of inferring much deeper relations in the text data[3,5].

The rest of the paper is organized as follows. Section 2 describes the basic concepts of LSA. Section 3 presents the experimental study that was used to identify the core research areas

and themes. Section 4 discusses the results of this analysis, focusing on three important factor solutions of LSA. Section 5 summarizes the paper with conclusions and limitations.

2 Research method

Latent semantic analysis (LSA) is a theory of knowledge acquisition, induction and representation[2]. It was first introduced as an information retrieval (IR) technique by [1] and [6]. It is an automatic mathematical learning technique for analyzing the relationships and similarity structures among documents and terms, relying on no human experiences, prior theoretic models, semantic dictionaries, or knowledge bases[3].

Similar to factor analysis, principal components analysis, and linear neural networks, the main purpose of LSA is dimension reduction, which is realized through a matrix operation called singular value decomposition (SVD). SVD is a means of decomposing a matrix indo a product of three simpler matrices. By retaining the k largest singular values, the resulting reduced-order SVD provides the best k-dimensional approximation to the original matrix, in the least square error sense[7]. In the results of SVD, two sets of factor loadings, one for the words and one for the documents, are generated. Each term and document is represented as a k-dimensional vector in the same latent semantic space derived by the SVD. Thus each latent semantic factor is now associated with a collection of high-loading terms and high-loading documents[5]. High-loading terms and documents are used to interpret and label the corresponding factor. The number of factors is an input parameter that needs to be provided before SVD computation. As the number of factors changes, LSA groups key terms or documents into various levels of aggregation. When it is applied to identify important topics of a certain discipline using a collection of representative papers, a higher level of aggregation (e.g., 2 factors) indicates key research areas and a lower level of aggregation (e.g., 100 factors) represents general research themes[5].

The LSA analysis can be summarized in three main steps. The first step is to set up a term-document matrix in which each row stands for a key word or term and each column stands for a document or context in which the key word appears. An entry in the matrix is the frequency of a key word in the corresponding document. The second step is to transform the term frequencies in a term-document matrix using various weighting schemes. The third step is to perform SVD on the matrix to reduce the dimensionality, which is the key feature of the LSA method. In this step only the k largest singular values are retained. The reduced-order SVD is the best k-dimensional approximation to the original matrix[7].

Extensive experiments have demonstrated that the classification performance of LSA is robust[8] and it is capable of inferring relations in the text [3,5]. It can be used in information retrieval (IR), search optimization, classification, clustering, filtering and other IR-related applications[7]. Readers interested in mathematical details of the LSA approach can refer to [1].

3 Experimental study

This section describes the data source and the implementation details of LSA analysis that is utilized to identify the core research areas and research themes for the selected field.

3.1 Data sources

The field of Multiple Criteria Decision Making (MCDM) and multiattribute utility theory (MAUT) has grown exponentially and made remarkable progress since 1960s. As a multidisciplinary field, MCDM/MAUT has close collaboration with some neighboring disciplines, such as

mathematical programming, organizational behavior, engineering, decision analysis, and negotiation science[9]. During the past twenty years, extensive research papers have been published in MCDM, MAUT, and related disciplines. In the experiment, LSA is applied to a collection of MCDM/MAUT publications to extract major research topics and identify the trends of the field.

Since previous studies, such as [10], [11] and [12], have investigated the major areas and the evolution of MCDM/MAUT before 1990s, articles published before 1985 were not included in the analysis. A total of 1515 research abstracts published in 16 refereed MCDM-related journals in the English language during the period of 1985 to February 2009 that contain key words: multiple criteria and multicriteria, were collected. As the first and unique journal in multiple criteria decision analysis, articles published in the Journal of Multi-Criteria Decision Analysis were all collected (from 1992 through 2007).

The 16 refereed MCDM journals were selected according to two criteria: (1) journals appeared frequently in the Multiple Criteria Decision Aid bibliography on the International society on MCDM website[13]; (2) the most relevant and top-rated MCDM journals listed by [14] and [15]. Each article collected in the dataset is stored in Microsoft Excel as one row with five fields: article title, author(s), journal name, year of publication, and abstract.

Table 1 lists the journals and the number of abstracts included in the text data. About 34% of the articles were published in the *European Journal of Operational Research*, with about 19% in the *Journal of Multi-Criteria Decision Analysis* and 8.5% in the *Journal of the Operational Research Society*.

Table1. Refereed MCDM journal articles, 1985-2008

Journals	Number of Articles
European Journal of Operational Research (EJOR)	519
Journal of Multi-Criteria Decision Analysis (JMCDA)	292
Journal of the Operational Research Society (JORS)	130
Computers & Operations Research (C&OR)	88
Fuzzy Sets and Systems (FSS)	86
Computers & Industrial Engineering (C&IE)	70
Decision Analysis (DA)	64
Omega	64
Mathematical and Computer Modelling (M&CM)	41
Annals of Operations Research (AOR)	40
Decision Support Systems (DSS)	35
Management Science (MS)	31
Operations Research (OR)	18
Journal of Optimization Theory and Applications (JOTA)	16
Theory and Decision (TD)	11
Organizational Behavior and Human Decision Processes (OBHDP)	10

Figure 1 summarizes the number of publications in the field of MCDM from 1985 to February 2009. Because text data were retrieved in October 2008, the number of abstracts collected for the year of 2009 can not reflect the real publication trend and therefore is ignored in Figure 1. As seen in Figure 1, the MCDM publications have been increased rapidly since 1992 and the number of MCDM publications has increased 4.7 times from 1985 to 2008.

3.2 Text preprocessing

The initial step of LSA analysis is to represent the text as a term-document matrix in which each row stands for a term and each column stands for a document. In order to set up such a matrix, this study started the analysis with text preprocessing procedures that are popular in the information retrieval and text mining[16,17].

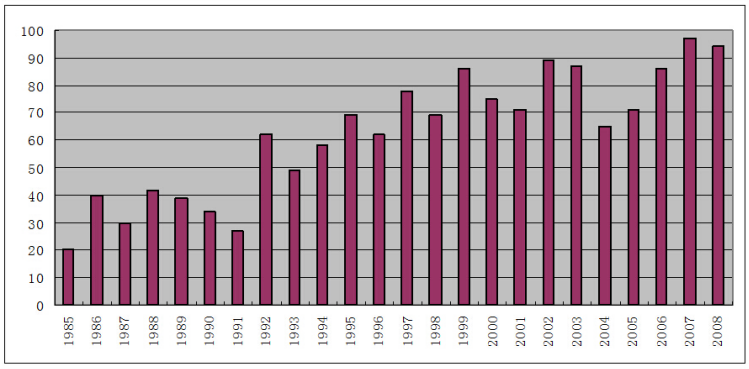


Figure 1: MCDM Publications from 1985 to 2008

The text preprocessing procedure consists of tokenization and term reduction. Tokenization divides documents into a set of terms. In this study, each article is represented by its title and abstract. Since titles are informative of research papers and normally contain pertinent key words, the weight of titles is set twice as much as abstracts. The 1,515 MCDM research papers generated a dictionary of 9,322 terms. Tokenization was implemented using a self-developed C++ program.

3.3 Term frequency matrix

Text preprocessing produced a term-frequency matrix with 1,515 columns (papers) and 3,299 rows (terms). Originally, an entry in the matrix contains the number of times a term occurs in a document. A term-frequency matrix measures the association of a term with respect to a given document[17]. There are many methods to define term weights. In this study, the *tf-idf*, a traditional term-frequency weighting, was used to transform the raw term frequencies in the matrix. The *tf-idf* weighting scheme combines term frequency (TF) and inverse document frequency (IDF) together:

$$W_{ij} = tf_{ij} \times idf_i \tag{1}$$

where tf_{ij} is term frequency and idf_i is the inverse document frequency of term i.

Inverse document frequency represents the importance of a term and is defined as:

$$idf_i = log_2(N/df_i) + 1 \tag{2}$$

N is the total number of documents and df_i is the document frequency of term i.

IDF implies that the discriminative power of a term will be decreased if it occurs in many documents. In other words, the importance of a term will increase if it appears in a limited number of documents. The reasoning behind the *tf-idf* weighting is that a term occurring frequently in a document but rarely in the rest of the collection is considered to be important. Experiments have shown that *tf-idf* measure works well in many applications[17,22]. The *tf-idf* weights were calculated using a linguistic analysis tool[23].

3.4 Latent semantic analysis

LSA can be considered as an application of reduced-order Singular Value Decomposition (SVD)[23]. SVD decomposes a term-document matrix X into the product of three other matrices:

$$X = W_0 S_0 C(prime)_0 \tag{3}$$

W_0 and C_0 are the matrices of left and right singular vectors and S_0 is the diagonal matrix of singular values. W_0 has the same number of rows as the original matrix and C_0 has the same number of columns as the original matrix. S_0 is a square matrix with non-zero entries only along one central diagonal and sorted in decreasing order[1]. The dimensionality of the original matrix can be reduced by keeping the first k largest coefficients in the diagonal matrix S_0 and setting the remaining smaller ones to zero. The zero rows and columns of S_0 can then be deleted to get a new diagonal matrix S. Similarly, the corresponding columns of W_0 and C_0 can be removed to obtain W and C respectively. The product of the simplified matrices is a new matrix $\hat{X}$:

$$\hat{X} = WSC(prime) \tag{4}$$

$\hat{X}$ is the k-rank matrix with the best possible least-squares-fit to X[1]. The results of SVD include one set of k-factor loading for the terms and one for the documents. High-loading terms and documents of a factor can then be used to interpret and label the factor. For mathematical and technical details of SVD, please refer to [1](p. 397-399).

The choice of k is a critical issue in SVD. An ideal value of k should be large enough to fit all the real structure in the data and small enough to avoid unimportant details[1]. Since solutions with different number of factors represent different levels of concept aggregation, we explored 2 through 13, and 100 factors respectively. Factor interpretation and labeling was conducted manually by two MCDM researchers. The high-loading terms and documents of 2 through 13 and 100 factor solutions were examined and labeled independently. The next section discusses the results of the LSA analysis.

4 Results and discussion

4.1 Different factor solutions

This work examined 13 solutions, including 2 through 13 and 100 factors, to identify key research areas and major research themes of MCDM. For the rest of the paper, factor x-y is used to indicate the y^{th} factor of the x-factor solution[5]. For example, factor 100-2 refers to the second factor of the 100-factor solution.

Different factor solutions of LSA show different levels of research themes of the MCDM discipline. The 6-factor and 11-factor solutions describe the evolution of these areas during the past twenty-four years and reveal major research areas of MCDM, including MAUT, ELECTRE methods, analytic network process (ANP), multicriteria decision support system (MCDSS), heuristics, preference learning, interactive multiple objective programming, MCDM applications, and goal programming.

As the number of factors increases, higher level research areas can be partitioned into subareas. For example, *Preference learning* (factor 6-4 from Table 2) in the 6-factor solution is represented by *Preference representation* (factor 11-5 from Table 3) and *Preference structure modeling* (factor 11-7 from Table 3) in the 11-factor solution; and *preference modeling* (factor 100-55), *preference elicitation support* (factor 100-77), and *preference ordering techniques* (factor 100-99) in the 100-factor solution.

Table2. Top 30 High-Loading Terms for the 6-Factor Solution

Factor	Factor Label	Top 30 Terms
6-1	Analytic network process (ANP)	pro,ecis,multi,decis,multipl,pre,riteria,criteria,roc, met,problem,ultipl, gener, rel,ref,model, criterion, valu,ram,set,risk,function, experi, probabl,appli, effect,approach,object, base, altern
6-2	Multicriteria decision support system	ecis,decis,riteria,criteria,pre,met,refer,ref,valu,method, multi,base, altern, risk,maker, group,appli, experi, prefer,model,multicriteria, util, function, result,theori, regret,multipl,analysi,rel,analys

6-3	Multi-Attribute Utility Theory (MAUT)	multipl,method,pro,ultipl,decis,probabl,task,learn,met,hypothes, gener, rel,effect,criterion, fuzzi, program,experi,problem,addit,ram, approach,ecis, linear,function,pre,paper,goal,set,risk,find
6-4	Preference learning	met,multi,multipl,ultipl,decis,method,object,line,linear, process, program, pre,prefer,refer, ecis,ref,singl,learn, evalu,effici,function, ram, roc, search,task,fuzzi,solut,multiobject,paramet,probabl
6-5	ELECTRE methods	riteria,criteria,method,ecis,decis,met,valu,multi,analysi, evalu,tri, analys, object,program,ram,multicriteria, perform,solut,linear,fuzzi, multipl, select,sel,algorithm, optim,risk,line,approach,paper,develop
6-6	Heuristics	riteria,pre,multi,refer,criteria,prefer,ref,multipl,ultipl, ram,met, problem,decis, function,optim,ecis, multicriteria, object,algorithm,program,pro, fuzzi,criterion, roc,solut,system,process,tri,plan,integ

The 100-factor solution presents a large variety of research themes studied during the last twenty years by the MCDM and related disciplines (see Table 4), including MCDM theories, algorithms, related areas of research, decision support systems, applications, and techniques. It also reveals important MCDM research topics that are not presented in the 6-factor and 11-factor solutions, such as data envelopment analysis (DEA) method, genetic algorithms, simulation, behavioral issues, theoretic foundation, and visual tools.

The 100-factor solution points out two notable trends in the MCDM publications. The first is the growth in applications of MCDM. In the 100-factor solution, 21 factors are related to MCDM applications. These applications cover not only traditional application areas, such as asset management[24], scheduling problem[25], assignment problem[26], questionnaire survey[27], credit scoring[28,29,30,31], and risk evaluation[32,33,34]; but also emerging novel areas, such as verbal data classification[35], Web-based decision support[36], habitual domains[37], electronic commerce systems[38,39], and e-participation[40]. The second trend is that MCDM has entered into some new research areas[41]. For example, Supply chain management has utilized MCDM methods to capture multicriteria decision making and decision-making under uncertainty[42]. Geographical Information Systems (GIS) and MCDM have been combined to aid spatial decisions[43]. These two results generally agree with [9], [12] and [28].

Table3. Top 30 High-Loading Terms for the 11-Factor Solution

Factor	Factor Label	Top 30 Terms
11-1	Goal programming	pro,ecis,multi,decis,multipl,pre,riteria,criteria, ultipl,met,problem,roc, gener,rel,ref,model,criterion, valu,ram,set,risk,function,experi, probabl,appli, effect,approach,object,base,altern
11-2	Multiple criteria sorting problem	ecis,decis,riteria,criteria,pre,met,refer,ref, prefer,method, multi, base,altern,risk,util,group, appli,maker,experi,valu,model,multicriteria, function, result,theori,regret,multipl,analysi,criterion,rel
11-3	Interactive fuzzy multiple objective decision making	multipl,method,pro,decis,ultipl,probabl,task,learn, met,gener,ram, hypothes,rel,effect,fuzzi,program, criterion,problem,approach,experi,linear, addit, set,function,goal,risk,ecis,pre,prefer,paper
11-4	Ranking alternatives	ram,met,multi,multipl,decis,ultipl,method,line,object,process,linear, program, pre,roc,ecis,prefer,refer,ref, evalu,singl,learn,fuzzi, effici, function, search,task,solut,multiobject,analyt,valu
11-5	Preference representation	riteria,criteria,ecis,method,decis,met,valu,multi,evalu, analysi, analys,object, program,tri,solut,ram,multipl, multicriteria,perform, optim, fuzzi, select,linear, risk,sel,ultipl,line,approach,algorithm,pape
11-6	Heuristic approach	riteria,pre,multi,refer,criteria,prefer,multipl,ref,ultipl,met, ram,decis, function, problem,optim,ecis,fuzzi, multicriteria,criterion,object,algorithm,pro, process, solut,system,program,singl,risk,prioriti,roc
11-7	Preference structure modeling	met,method,riteria,ref,refer,pre,criteria,prefer,model,multipl,function, ecis,process,roc,multicriteria,ultipl, weigh,regret,algorithm,decis, weight,group,prioriti, case,gener,goal,methodolog,learn,prior,sel
11-8	Machine learning and knowledge discovery	multi,multipl,ultipl,model,system,criteria,riteria,object, valu,attribut, analys,met,search,risk,function, research,regret,method,set, analysi,man, pre,problem, effici,theori,compar,ecis,line,polici,paper
11-9	Applications	tri,riteria,criteria,attribut,met,model,method,prefer,refer,multipl,system, ultipl,problem,evalu,valu,ref, util,ram,man,multi, multicriteria, multiattribut, term, ecis,log,pro,solut,criterion,manag,analys
11-10	Multiattribute utility theory	model,multi,analys,analysi,man,ultipl,problem,object, decis,solut, multipl, program,valu,manag,ram,ecis, refer,gener,prefer,multiobject, appli,method, maker, interact,system,rel,strateg,log,algorithm,paper

11-11	Interactive procedure for MCDM	model,tri,man,valu,function,attribut,weigh,line,linear,evalu,weight, criteria, manag,problem,multi,riteria, fuzzi,util,goal,system,plan, search, cost,ram,network, research,multiattribut,decis,altern,process

Table4. Factor Labels for the 100-Factor Solution

Factor Label		
Project selection and scheduling methodology	Multiple criteria decision making under uncertainty	Outranking relations
GIS and MCDM integration	Exact algorithms	Multiple criteria linear regression
Method for ranking alternatives	Dynamic consistency (DC) optimization techniques	Monte Carlo simulation
Multi-objective optimization	Evaluating decision alternatives	Electronic commerce
MCDM in data mining	Qualitative decision making	Portfolio selection and management
ELECTRE methods	Stochastic goal programming	Multiple objective ant colony optimization algorithms
Scheduling problems	Comparative study of MCDM methods	MAUT model
Multicriteria classification	Multiple criteria simulation optimization method	Bayesian approach
Heuristic algorithm	MAVT	Preference elicitation
Interactive multiple objective programming procedure	Artificial intelligence	Interactive multiobjective optimization
Manufacturing system	Neural network for MCDM	Alternative evaluation models
Interactive multi-objective sys.	AIM	MCDM in strategic energy policy making
Multiple criteria decision support system	Environmental planning assessment and decisions	Measures of interdependences between the objectives
Design problem	Attribute weights determination	Influence diagram
Multicriteria expert support system	Tchebycheff procedure for multiple objective decision making	Multiple criteria group decision making
Genetic algorithms	Tabu search	Dynamic programming
Multi-criteria production planning	Decision maker's utility function assessment	Group decision support system (GDSS)
Information systems	Multiple criteria ABC analysis	Knowledge discovery and MCDM (neural network)
Flow shop scheduling problem	Fuzzy set and approximate reasoning	Vector optimization
Genetic algorithms	Tabu search	Dynamic programming
Algorithm development	Internet and public decision making	MCDM in cellular manufacturing system
System performance measures	Preference modeling	System design problem
Case study	Web-based decision support and applications	Game theory approach
Industrial facilities layout planning and design	ANP technique	Discrete multiple criteria problems
Operations research	Multiobjective decision making in military applications	TOPSIS
Facility location problem	Zionts-Wallenius algorithm	Graphical display tools
MCDM and industrial engineering	Modeling interaction between criteria in MCDM	Fuzzy MCDM
DSS	Applications of heuristic approaches	Lexicographic goal programming
Data mining and ML	AHP improvements	DEA
Philosophy of MCDM	Parameter determination methods	Theoretic foundation
Goal programming	Metaheuristic algorithm	Behavioral issues
Multicriteria location problem	SMAA	Optimization algorithms and implementation of MCDM
Resource allocation model	Team decision making under uncertainty	
Visual tools	Simulation modeling	

Table5. Top 10 High-Loading Papers for the 6-Factor Solution

Factor	High-Loading Papers	Factor Loading
6-1	Jin Woo Lee, Soung Hie Kim, C&OR,2000	0.14
	J. M. Coutinho et al.,C&OR ,1999	0.11
	Wey, Wann-Ming,Wu, Kuei-Yang, M&CM,2007	0.11
	Behnam Malakooti, Jumah E. Al-alwani, C&OR, 2002	0.09
	Minghe Sun et al., C&OR, 2000	0.08
	Lorraine R. Gardiner, Ralph E. Steuer, EJOR, 1994	0.08
	Otto Rentz, FSS, 1996	0.08
	Taeyong Yang et al., FSS, 1991	0.08
	Bernard Roy, Roman Slowinski, AOR, 2006	0.07
	Mark A. Coffin, Bernard W. Taylor, C&OR, 1996	0.07
6-2	C. Zopounidis, Michael Doumpos, C&OR, 2000	0.14
	T. Terlaky, EJOR, 1985	0.11
	V. Mousseau et al., C&OR, 2000	0.11
	Taeyong Yang et al., FSS, 1991	0.11
	Otto Rentz, FSS, 1996	0.11
	Lorraine R. Gardiner, Ralph E. Steuer, EJOR, 1994	0.09
	N. M. Badra, FSS, 2002	0.09
	E. Melachrinoudis, Z. Xanthopulos, C&OR, 2003	0.09
	Masatoshi Sakawa, Hitoshi Yano, FSS, 1989	0.09
	John A. Aloysius, et al., EJOR, 2006	0.08
6-3	Jose Rui Figueira et al., EJOR, 2008	0.15
	Risto Lahdelma et al., EJOR, 2003	0.15
	David L. Olson, EJOR, 2001	0.15
	S. Greco, V. Mousseau, R. Slowinski, EJOR, 2008	0.11
	Stelios H. Zanakis, et al., EJOR, 1998	0.1
	Pekka J. Korhonen, Jukka Laakso, EJOR, 1986	0.1
	Gerard Colson, C&OR, 2000	0.09
	Silvia Angilella, EJOR, 2004	0.09
	An Ngo The, Vincent Mousseau, JMD,2002	0.09
	Risto Lahdelma, Pekka Salminen, EJOR, 2002	0.08
	Edmund Kieran Burke, Sanja Petrovic, EJOR, 2002	0.08
6-4	Jose Rui Figueira et al.,EJOR,2008	0.25
	Bernard Roy, Roman Slowinski, EJOR, 2008	0.19
	George Mavrotas, Panagiotis Trifillis, C&OR, 2006	0.19
	Theodor J. Stewart, EJOR, 1986	0.18
	Peter Muller, DA, 2006	0.13
	Kim Fung Lam, Eng Ung Choo, JORS, 1995	0.1
	Murat Koksalan, Ahmet Burak Keha, 2003	0.1
	Risto Lahdelma, Pekka Salminen, EJOR, 2002	0.09
	Salvatore Greco, et al., EJOR, 2002	0.09
	Gregory E. Kersten, DSS, 1988	0.08
6-5	J.C. Leyva-Lopez, E. Fernandez-Gonzalez, EJOR, 2003	0.28
	Salvatore Greco et al., EJOR, 2008	0.25
	Bernard Roy, Roman Slowinski, EJOR, 2008	0.12
	Risto Lahdelma, Pekka Salminen, EJOR, 2002	0.12
	J OS C. FODOR, MARC ROUBENS, JMCDA, 1997	0.12
	Silvia Angilella et al., EJOR, 2004	0.11
	Huseyin Cavusoglu, Srinivasan Raghunathan, DA, 2004	0.11
	Salvatore Greco, et al., EJOR, 2002	0.1
	Minghe Sun, EJOR, 2002	0.1
	Minghe Sun, EJOR, 2002	0.1
6-6	J. Gupta, Kruger, Lauff, Werner, Sotskov, C&OR, 2002	0.27
	J. Gupta, K. Hennig, F. Werner, C&OR, 2002	0.24
	Peter Muller et al., DA,2006	0.13
	Sandeep Purao et al., DSS, 1999	0.12
	Gregory E. Kersten, DSS, 1988	0.12
	Ilia Tsetlin, Robert L. Winkler, DA, 2006	0.12
	Jatinder N. D. Gupta, Johnny C. Ho, C&OR, 2001	0.11
	Vincent T'kindt et al., C&OR, 2003	0.11
	B. Malakooti, C&OR, 1989	0.1
	Julian Molina et al., EJOR, 2008	0.1

4.2 Different factor solutions

Table6. Factor Labels and Paper Counts for the 6-Factor Solution

Factor	Factor Label	Paper Counts				
		85-89	90-94	95-99	00-04	05-09
6-1	Analytic network process (ANP)	41	48	54	63	61
6-2	Multicriteria decision support system	37	63	51	49	39
6-3	Multi-Attribute Utility Theory (MAUT)	17	44	62	40	42
6-4	Preference learning	22	27	34	22	55
6-5	ELECTRE methods	38	42	54	59	35
6-6	Heuristics	32	36	49	39	45

Table7. Factor Labels and Paper Counts for the 11-Factor Solution

Factor	Factor Label	Paper Counts				
		85-89	90-94	95-99	00-04	05-09
11-1	Goal programming	21	20	37	27	16
11-2	Multiple criteria sorting problem	37	63	51	49	39
11-3	Interactive fuzzy multiple objective decision making	15	21	13	19	11
11-4	Ranking alternatives	22	14	26	24	7
11-5	Preference representation	21	16	29	34	30
11-6	Heuristic approach	24	22	27	39	29
11-7	Preference structure modeling	19	22	37	26	33
11-8	Machine learning and knowledge discovery	14	16	25	29	16
11-9	Applications	27	29	45	32	26
11-10	Multiattribute utility theory	17	19	19	18	20
11-11	Interactive procedure for MCDM	18	24	37	28	35

Figure 2 suggests that the growth in some research areas, such as preference learning (factor 6-4), heuristics (factor 6-6), and analytic network process (ANP) (factor 6-1) increased considerably from the 1985-1989 period to the 2005-2009 period. In the case of ELECTRE methods (factor 6-5), the number of publications maintained a relatively constant increase from 1985-2008. The research interests in Multi-Attribute Utility Theory (factor 6-3) grew significantly from 1985-1989 to 1995-1999 and dropped during the 2000-2004 period. The number of publications in MAUT remained stable since then. Multicriteria decision support system (factor 6-2) experienced a rapid growth from 1985-1989 to 1990-1995 and declined slightly during 1995-2008.

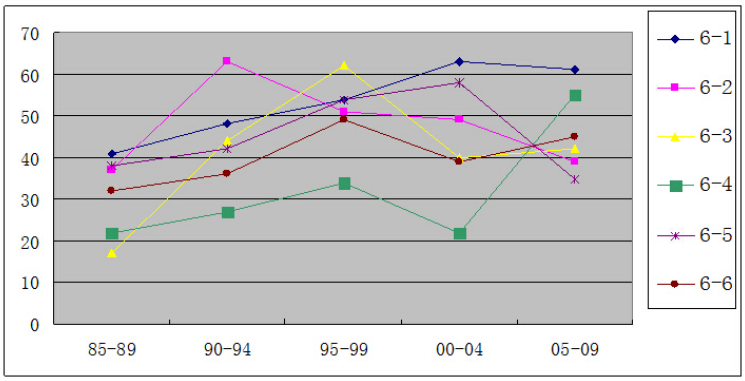


Figure 2: Dynamics of Major Research Areas (six-factor solution)

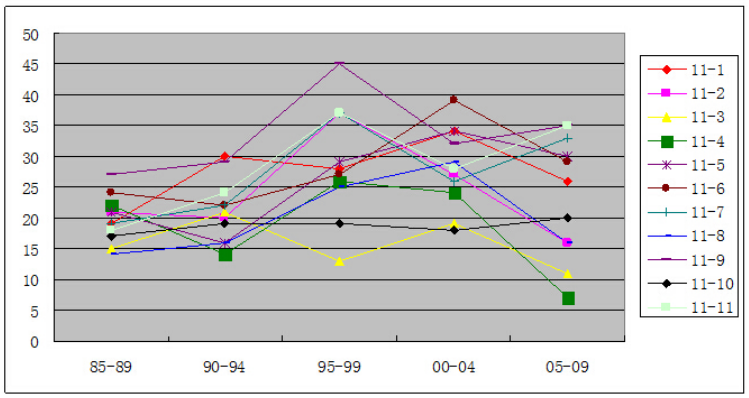


Figure 3: Dynamics of Major Research Areas (eleven-factor solution)

5 Conclusions and limitations

This paper attempted to identify the major research areas and themes of MCDM field by examining a large body of related research papers using latent semantic analysis. In the experimental study, over fifteen hundred abstracts of MCDM/MAUT field were collected and analyzed to obtain thirteen factor solutions. The 6-factor and 11-factor solutions of the analysis reveal key research areas of MCDM/MAUT. MAUT, ELECTRE methods, ANP, multicriteria decision support system (MCDSS), heuristics, preference learning, interactive multiple objective programming, MCDM applications, and goal programming are among the main streams of thought of the field.

The ideas and techniques of MCDM are continuing to integrate into other disciplines. For example, data mining (DM) field used ELECTRE methods to cluster opinions[44] and utilized multiple criteria decision aid process to help users to sort association rules[45]. Artificial neural networks, an artificial intelligence (AI) method, has been used by MCDM researchers to solve discrete MCDM problem[46] and model decision-makers' preference structures[47]. Geographical Information Systems (GIS) and MCDM have been combined to aid spatial decisions[43].

This study has several limitations. First, since the LSA analysis depends on identifying frequent word usage patterns from a collection of text, it is difficult to capture a research area if it is not well established and has not established consistent terminology among its researchers[5] . Second, this study only collected articles published after 1985 because the major areas and the evolution of MCDM and MAUT before 1990s have been investigated in previous studies[10,11,12]. Third, the research abstracts collected in this analysis include only English language journals. Papers published in other languages are not considered.

Acknowledgements

This work was supported in part by grants from the National Natural Science Foundation of China (#71325001, #71222108 and #71173028), Program for New Century Excellent Talents in University (NCET-12-0086).

Bibliography

[1] Deerwester, S.; Dumais, S.; Furnas, G.; et al. (1990). Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science*, 41(6): 391-407.

[2] Landauer, T.; Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge, *Psychological Review*, 104: 211-240.

[3] Landauer, T.; Foltz, P.; Laham, D. (1998). Introduction to Latent Semantic Analysis, *Discourse Processes*, 25: 259-284.

[4] Kou, G.; Lou, C. (2012). Multiple Factor Hierarchical Clustering Algorithm for Large Scale Web Page and Search Engine Clickstream Data, *Annals of Operations Research*, 197(1)25: 123-134.

[5] Sidorova, A.; Evangelopoulos, N.; Valacich, J. S.; et al. (2008). Uncovering the intellectual core of the information systems discipline, *MIS Quarterly*, 32(3): 467-482.

[6] Dumais, S. T.; Furnas, G. W.; Landauer, T. K.;et al (1988). Using latent semantic analysis to improve information retrieval, *Proceedings of CHIŻ88 Conference on Human Factors in Computing Systems*, 281-285.

[7] Dumais, S. T. (2004). Latent Semantic Analysis, *Annual Review of Information Science and Technology*, 38: 189-230.

[8] Gansterer, W.N.; Janecek, A.G.K.; Neumayer, R. (2008). In M. W. Berry and M. Castellanos (eds.), Survey of Text Mining: Clustering, Classification, and Retrieval, Second Edition (pp. 165-183). Springer

[9] Wallenius, J.; Dyer, J. S.; Fisburn, P. C.; et al. (2008). Multiple Criteria Decision Making, Multiattribute Utility Theory: Recent Accomplishments and What Lies Ahead, *Management Science*, 54(7): 1336-1349.

[10] Stewart T. J. (1992). A critical survey on the status of multiple criteria decision making theory and practice, *OMEGA*, 20(5/6): 569-586.

[11] Dyer, J. S.; Fisburn, P. C.; Steuer, R. E.; et al. (1992). Multiple criteria decision making, multiattribute utility theory: the next ten years, *Management Science*, 38(5): 645¨C654.

[12] Urli, B.; Nadeau, R. (1999). Evolution of multi-criteria analysis: a scientometric analysis, *J. Multi-Crit. Decis. Anal.*, 8: 31-43.

[13] International society on MCDM, (2009). http://www.mcdmsociety.org, Accessed 24 Jun 2009

[14] Steuer, R. E.; Gardiner, L. R.; Gray, J. (1996). A bibliographical survey of the activities and international nature of multiple criteria decision making, *J. Multi-Crit. Decis. Anal.*, 5: 195¨C217.

[15] Bragge, J.; Korhonen, P.; Wallenius, J.; et al. (2008). Bibliometric Analysis of Multiple Criteria Decision Making/Multiattribute Utility Theory, *International Society on Multiple Criteria Decision Making*, Accessed 11 June 2009.

[16] Fox, C. (1992). Lexical Analysis and Stoplists. In W. B. Frakes and R. Baeza-Yates (eds.), Information Retrieval: Data Structures and Algorithms (pp. 102-130). Upper Saddle River, NJ: Prentice-Hall.

[17] Han, J.; Kamber, M. (2006). Data Mining: Concepts and Techniques, 2nd edition. San Francisco, CA: Morgan Kaufmann Publishers.

[18] Stopwords. (2008). Webconfs.com, `http://www.webconfs.com/stop-words.php`, Accessed 10 August, 2008.

[19] SQL Sever 2005. Microsoft.com, `http://www.microsoft.com/sqlserver/2005/en/us/overview.aspx`, Accessed 1 Feb 2009.

[20] Porter, M. F. (1980). An algorithm for suffix stripping, *Program*, 14(3): 130-137.

[21] Porter, M. F. (2008). The Porter Stemming Algorithm. http://tartarus.org/ martin/ Porter-Stemmer/. Accessed 22 Feb, 2009.

[22] Baeza-Yates, R.; Ribeiro-Neto, B. (1999). Modern Information Retrieval, *Addison-Wesley, Wokingham*, UK.

[23] LingPipe (2008). `http://alias-i.com/lingpipe/index.html`, Accessed 1 March 2009.

[24] Langen, D. (1989). An (interactive) decision support system for bank asset liability management,*Decision Support Systems*, 5(4): 389-401.

[25] Geiger, M. J. (2007). On operators and search space topology in multi-objective flow shop scheduling, *European Journal of Operational Research*, 181(1): 195-206.

[26] Przybylski, A.; Gandibleux, X.; Ehrgott, M. (2008). Two phase algorithms for the bi-objective assignment problem. *European Journal of Operational Research*, 185(2): 509-533.

[27] Ergu, D.; Kou, G. (2012). Questionnaire Design Improvement and Missing Item Scores Estimation for Rapid and Efficient Decision Making, *Annals of Operations Research*, 197(1):5¨C23, DOI 10.1007/s10479-011-0922-3.

[28] Shi, Y. (2001). Multiple Criteria Multiple Constraint-level (MC2) Linear Programming: Concepts, Techniques and Applications, *World Scientific Publishing*, 539 pages.

[29] Yu, L.; Wang, S.; Lai, K. K. (2009). An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit scoring. *European Journal of Operational Research*, 195(3): 942-959.

[30] Kou, G.; Peng, Y.; Wang, G.X. (2014a). Evaluation of Clustering Algorithms for Financial Risk Analysis using MCDM Methods, *Information Sciences*, 27:1-12, DOI: HTTP://DX.DOI.ORG/10.1016/j.ins.2014.02.137.

[31] Kou, G.; Peng, Y.; Lu, C. (2014b). An MCDM Approach to Evaluate Bank Loan Default Models, *Technological and Economic Development of Economy*, 20(2): 278¨C297, DOI: HTTP://DX.DOI.ORG/10.3846/20294913.2014.913275.

[32] Ergu, D.; Kou, G.; Shi, Y.; et al. (2011). Analytic Network Process in Risk Assessment and Decision Analysis, *Computers & Operations Research*, DOI: 10.1016/j.cor.2011.03.005.

[33] Kou, G.; and Lin, C. (2014) A cosine maximization method for the priority vector derivation in AHP, *European Journal of Operational Research*, 235: 225-232 , DOI: http://DX.DOI.ORG/10.1016/j.ejor.2013.10.019

[34] Montibeller, G.; Belton, V.; Lima, M.V.A. (2007). Supporting factoring transactions in Brazil using reasoning maps: a language-based DSS for evaluating accounts receivable. *Decision Support Systems*, 42(4): 2085-2092.

[35] Yevseyeva, I.; Miettinen, K.; Rasanen, P. (2008). Verbal ordinal classification with multicriteria decision aiding. *European Journal of Operational Research*, 185(3): 964-983.

[36] Hamalainen, R. P. (2003). Decisionarium-aiding decisions, negotiating and collecting opinions on the web. *Journal of Multicriteria Decision Analysis*, 12(2-3): 101-110.

[37] Yu, P.L. (1991). Habitual domains, *Operations Research*, 39(6): 869-876.

[38] Chiu, Y.; Shyu, J. Z.; Tzeng, G. H. (2004). Fuzzy MCDM for evaluating the e-commerce strategy, *International Journal of Computer Applications in Technology*, 19(1): 12-22.

[39] Kameshwaran, S.; Narahari, Y.; Rosa, C. H.; et al. (2007). Multiattribute electronic procurement using goal programming. *European Journal of Operational Research*, 179(2): 518-536.

[40] Moreno-Jimenez, J. M.; Polasek, W. (2003). e-democracy and knowledge. A multicriteria framework for the new democratic era. *Journal of Multi-Criteria Decision Analysis*, 12(2-3): 163-176.

[41] Zeleny, M. (1998). Multiple criteria decision making: eight concepts of optimality, *Human Systems Management*, 17(2): 97-107.

[42] Dong, J.; Zhang, D.; Yan, H.; et al. (2005). Multitiered Supply Chain Networks: Multicriteria Decision Making Under Uncertainty. *Annals of Operations Research*, 135(1): 155-178.

[43] Gomes, E. G.; Lins, M. (2002). Integrating geographical information systems and multicriteria methods: A case study. *Annals of Operations Research*, 116(1-4): 243-269.

[44] Bisdorff, R. (2002); Electre-like clustering from a pairwise fuzzy proximity index, *European Journal of Operational Research*, 138(2): 320-331.

[45] Lenca, P.; Meyer, P.; Vaillant, B.; et al. (2008). On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. *European Journal of Operational Research*, 184(2): 610-626.

[46] Malakooti, B.; Zhou, Y. Q. (1994). Feedforward Artificial Neural Networks for Solving Discrete Multiple Criteria Decision Making Problems, *Management Science*, 40(11): 1542-1561.

[47] Wang, J. (1994). A neural network approach to modeling fuzzy preference relations for multiple criteria decision making. *Computers and Operations Research*, 21(9): 991-1000.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):370-381, June, 2015.

Applications of Ubiquitous Sensor Network: Micro-Scale Air Quality Monitoring

S. Lee, S. Lim

Sangdong Lee
Korea Institute of Science Technology Information (KISTI)
245 Daehangno, Yuseong, Daejeon 305-806, Korea
sdlee@kisti.re.kr

Sang Boem Lim
Department of Advanced Technology Fusion and Division of Interdisciplinary Studies
Konkuk University, Seoul, Korea
1 Hwayang-dong, Kwangjin-gu
Seoul 143-701, Korea
*Corresponding author: sblim@konkuk.ac.kr

Abstract: The rise of densely populated societies in metropolitan areas has led to several forms of pollution, thereby leading to increasing health concerns. Among the various types of pollution, air pollution has become a crucial factor in the quality of life. Currently, the meteorological administration in Korea provides a regional scale of air quality index. However, this broad-based information is not relevant to local urban areas. We believe that air quality information about local environments such as homes, workplaces, and parks would be more relevant to people. We have developed a micro-scale air quality monitoring system called AirScope that covers local environments using a ubiquitous air quality sensor network. In this study, we focus on the applications of AirScope. We have developed AirScope as a general-purpose desktop PC application as well as a mobile application that can increase user mobility and provide real-time air quality information about the local user environment.
Keywords: Air Pollution Monitoring, USN, AirScope, Mobile Application

1 Introduction

The rise of densely populated societies in metropolitan areas has led to several forms of pollution, thereby leading to increasing health concerns. Among the various types of pollution, air pollution has become a crucial factor in the quality of life. However, currently, there is no public access to information about air quality. Efforts that provide air quality information and hence air pollution warnings are restricted to very few areas. A majority of such information is provided as an air quality index for entire cities. Such an index cannot provide information that adequately represents the air quality in local urban areas in a city.

Currently, the meteorological administration in Korea provides a regional scale of air quality index. However, this broad-based information is not relevant to local urban areas. We believe that air quality information about local environments such as home, workplaces, and parks would be more relevant to people. To this end, we have developed a micro-scale air quality monitoring system called AirScope that covers local environments using a ubiquitous air quality sensor network. Our previous studies have already provided a general overview of the AirScope system [1] [2]. A primary feature of AirScope is that it can be run as a smartphone application. Smartphones provide the advantages of mobility and access to information from anywhere at any time. Further, applications for smartphones directly target the end-user, and thus, the smartphone is particularly suited for providing air quality information.

In this study, we focus on the applications of AirScope. We have developed AirScope as a general-purpose desktop PC application as well as a mobile application that can increase user

mobility and provide real-time air quality information about the local user environment. These applications are examples of AirScope usage which can be adopted as a service of company that is interested in real-time air quality service. In order to provide our services to the other research group or companies, we offer open API of AirScope. This paper is mainly demonstration of usage of AirScope using its Open API.

The desktop and smartphone user applications are discussed in section 2, while section 3 provides a brief introduction to AirScope. Two of the most important sub-systems of AirScope Local AirScope and Global AirScope are explained in sections 4 and 5.

2 User Applications

Our desktop application is designed to provide a wide variety of air quality information services. On the other hand, our mobile application is designed to increase user mobility and provide real-time location-based air quality information. Common services for both the desktop and mobile applications include maps, charts, and 3D services. The desktop application provides a rich set of graphics, useful environment information, and high 3D performance, while the strength of the smartphone application is its location-based service. In the study, we have selected iPhone, which provides GPS services, as our choice of smartphone. The application uses GPS information to provide location-based AirScope services. This means that when a user launches the AirScope application using his/her smartphone, the map service automatically detects his/her current location and displays it on the map.

2.1 Desktop Application

The desktop application provides real-time data services using maps, one-day data services using charts, data search and bookmark services, and local environmental information. Fig. 1 shows the AirScope desktop portal interface.

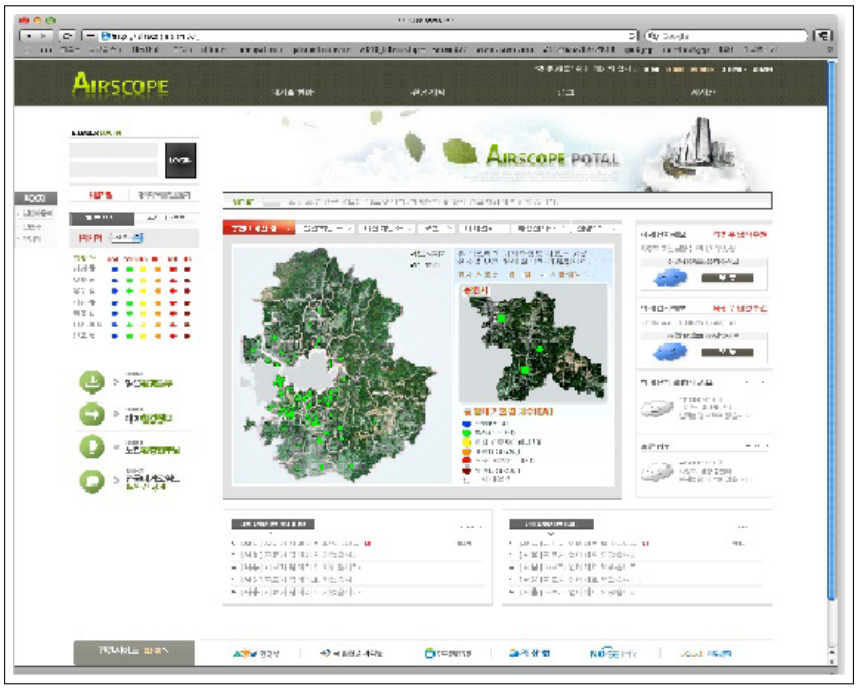


Figure 1: Desktop Portal Interface of AirScope

In the study, we have used Google Maps [3] for real-time data services. A marker that is positioned at the top of the map indicates the location of the air quality sensors. The portal requests data regarding the sensor location from Global AirScope when the real-time data service is launched. Global AirScope replies to the portal with the name of the location, longitude and latitude of sensor, and current sensor data. Upon receiving this information, the AirScope portal displays a marker at the top of the map. When the user clicks on this marker, the current air quality information is displayed by SmartWindow (see Fig. 2).

SmartWindow shows an image of the building in which the sensor is located and the current value of temperature, humidity, CO and CO2 concentrations, and PM-10 pollution levels. Further, SmartWindow provides the local environment data for the previous 24 hours when the user clicks on the average button. Fig. 3 shows the temperature data for the past 24 hours.

Figure 2: Display of real-time sensor data in SmartWindow

The periodic air quality data service (Fig. 4) provides air quality information over certain time periods. The goal of this service is to provide a general idea of the amount of pollutants for specific areas; thus, the user can compare the amount of pollutants between different areas. In order to use this service, the user begins by choosing the relevant start and end data. The user can select one or more sensor locations and the pollutant whose concentration he/she wants to check. The selection provides the user with a bar graph of the concentrations of the selected pollutant for comparison. Further, the application provides a line graph of the pollutant for online access. This graph allows the user to study changes in the amount of pollutants. The iPhone application also provides the same information.

The difficulty in identifying a single location on a map is addressed by the AirScope systems bookmarker service; the bookmarker service allows faster search and easier map navigation. This service works similar to the bookmark feature in any Internet browser. A user can bookmark his/her choices of locations in the AirScope portal by create a location folder. Subsequently, he/she can click on a location in the map and enter the location name. The user can revisit a location by simply clicking on the saved bookmark.

Figure 3: One-hour average data service for local sensor environment

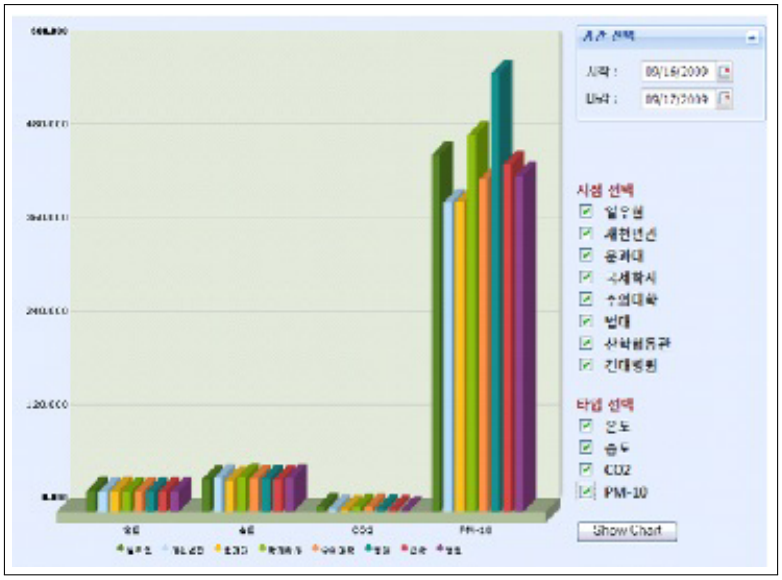


Figure 4: . Real-time monitoring service

2.2 Mobile AirScope

Until 2009, iPhone SDK 2.0 [4] did not support map API development. At that time, we developed the AirScope map service in the server by using Google Maps. In addition, we developed the web-based application with the iPhone browser. In 2010, iPhone SDK 3.0 [5] included map development API libraries called the MapKit framework. Our test results with SDK 3.0 showed a significant increase in process speeds and reliability of our application. Therefore, we redesigned and rebuilt the mobile AirScope application for iPhone SDK 3.0. This has resulted in maximized use of SDK 3.0s enhanced touch features and cursor movement speed when compared with our previous version. The loading speed of text and pictures is also enhanced because the browser component does not require loading. The application interface has been retained with increased reliability of the mobile application.

The smartphone application essentially provides the same functionalities that are available with the desktop application. The application includes real-time data services, sensor data chart services, and computational fluid dynamic (CFD) modeling (Fig. 5). A characteristic mobile service is a location-based user service. The GPS data from the smartphone is used for the map and notification services. The map service automatically displays the users current position and notifies him when the air quality in the environment changes.

Figure 5: Services of mobile AirScope

2.3 AirScope Open API Applications

Recent emergence of Web 2.0 [5], most of services are provided by the form of Open API [6]. Currently, various projects are using ad providing open API to fulfill openness and scalability of system. Project areas are including but not limited to mobile network API [7], enterprise application integration [8], and even in area of biology research [9] and Geographic Information System (GIS) research [10]. Particularly micro-scale environment services are used in variety of applications. To support these services and an API, we are providing Open API of our AirScope system.

A company, called Green Ecos, developed a mobile application using our open API. In this application, they provide real-time sensor data, historical chart, air quality Augmented Reality (AR) and Virtual Reality (VR) application of current air quality, social network service connec-

tion, and tracking of expose to harmful meterals. Other than social network service connection, they are using AirScope open API to provide services. Once they get sensor data using open Api, they will process data for their services. Fig. fig:gre shows UI of Green Ecos application.

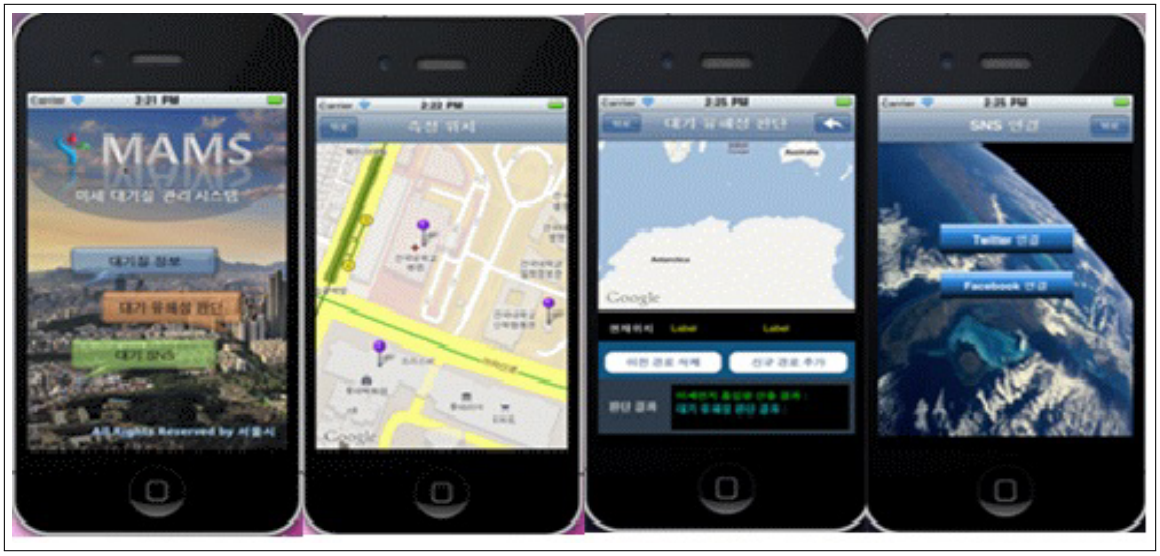


Figure 6: Green Ecos Application

3 AirScope System

In this section, we provide an overview of the AirScope architecture, including the specifications of sensors, communication between the systems involved, and the need and use of CFD modeling data. There are various ubiquitous applications using wireless sensor network researches [11] [12]. Each project has different goals. Former project targeted integrate development of an application between domain experts and network experts without specific knowledge of each others area. Later research uses bicycle to increase mobility. Our project targeted general audience to use our system. Local AirScope, Global AirScope, and their user applications are discussed in the following section.

3.1 General Architecture

Fig. 7 shows the general architecture of AirScope. Our system comprises three important components Local AirScope, Global AirScope, and their end-user applications. A portion of our study includes the installation of several sensors (approximately 1,000 to 10,000 sensors) in one district in Seoul (which comprises 25 districts). The impossibility of handling all the data by one server led to the plan of having one or more servers in a district (based on its size and population). We call this server as Local AirScope. It is expected to handle one area comprising the installed sensors and collect and process the sensor data in its operational area, store the data in a local database (DB), and manage the sensors. Local AirScope can solve the problem of server overloading. However, we require a mechanism to manage and gather information from Local AirScope to provide end-user services. Global AirScope is used to address this issue. This server manages Local AirScopes, distributes queries to obtain air quality information, processes the received data, and sends information to the user application. In this study, we have created the desktop portal and smartphone applications; however, there is no limit on the number and

types of user applications. Our architecture has been designed to accept any user applications regardless of the platforms and programming languages used.

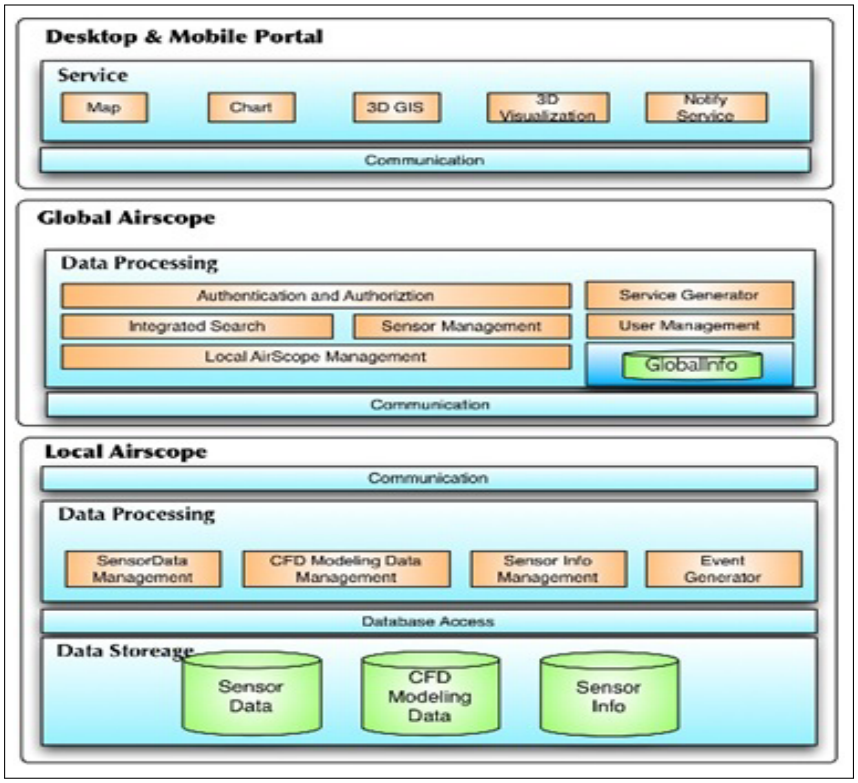


Figure 7: Architecture of AirScope system

4 Local AirScope

Local AirScope is a system that manages the collected data from the sensors and CFD model and provides requested data to the user to provide micro-scale air quality monitoring services. In our study, the purpose of this system was to parse and store data from both the sensors and the CFD model; further, it was required to manage sensor information such as name, type, and location. We assumed the presence of more than one Local AirScope system in the study because the actual implementation of the system will involve the installation of several thousand sensors in one of the districts in Seoul. The large quantities of data generated will require many servers. For the purposes of the study, we assume one Local AirScope system per district but the actual coverage of this system is expected to depend on the number of sensors in the area.

In the following section, we describe the architecture, data packet protocol, data processing, event generation, and DB schema of Local AirScope in detail.

4.1 Architecture of Local AirScope

Local AirScope system (Fig. 8) comprises three layers communication layer, data processing layer, and data storage layer. The communication layer manages communication between Local AirScope and Global AirScope and between Local AirScope and the sensor gateway. By means of this layer, one can send data requests and responses and obtain data from the gateway. The data processing layer is responsible for managing sensor data, CFD data, and sensor information.

All the actual sensor data, CFD modeling data, and sensor information are stored in the data storage layer.

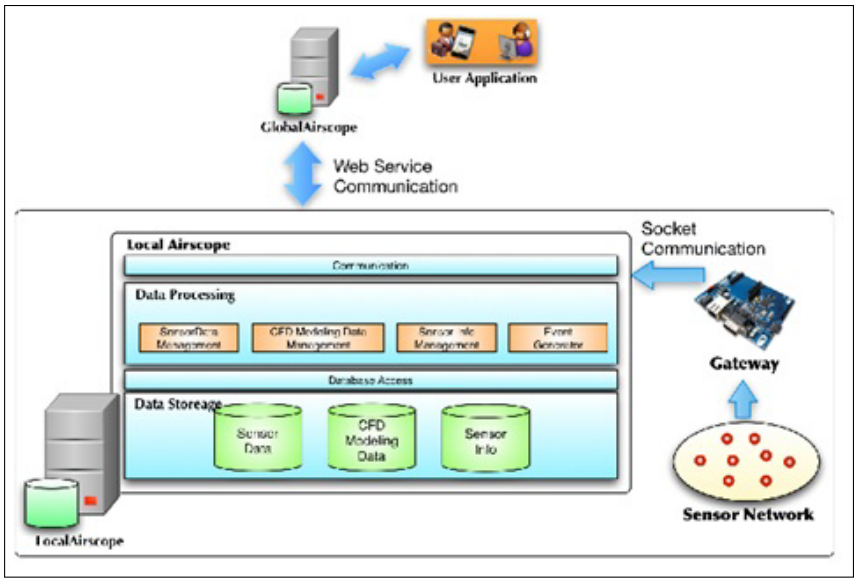


Figure 8: Local AirScope architecture

The communication layer uses Web Service and Socket as components of the communication module. We have explained the Web Service communication module in a previous section. The Web Service (RESTful) is used for communication with Global AirScope. Among the current Web Service technologies, we used a light weight communication package with a light weight data format called REST and JSON, respectively. The Socket communication module is used between Local AirScope and the sensor gateway to collect sensor data. All the sensors send data to the sensor gateway which in turn sends data to Local AirScope. The reason for the use of socket communication is due to the sensor gateway that uses embedded Linux as its operating system; embedded Linux does not provide any communication method other than socket communication.

The data processing layer contains management modules for sensor data, CFD modeling data, sensor information, and event processing. The main task of the sensor data management module is to obtain sensor data from the communication module, store data in the server DB, and respond to data requests from Global AirScope. The sensor data from the communication module is in the form of 39-byte hexadecimal raw data. This data cannot be understood by humans unless it is processed according to the sensor manufacturers protocol. The sensor data management module processes the incoming raw sensor data (Fig. 9) that contains temperature, humidity, CO, CO2, and PM-10 information into meaningful data and stores this processed information in the sensor DB. This raw data also contains a sensor ID to pinpoint the data source. This ID acts as a prime key to the DB.

The CFD modeling data management module manages data from the CFD modeling that can be considered as data from virtual sensors in order to overcome the shortage of sensors in our test-bed. This allows us to provide micro-scale air quality information without the installation of many actual sensors. The virtual sensor data is obtained in double number formats (e.g., 0.495508E-10, 0647687E-11) that need to be processed in order to be understood. The CFD modeling data management module processes this data and stores it in the CFD modeling DB. In addition, this module responds to data requests from Global AirScope for locations that do not have actual sensors.

When a user registers with our system through the Internet, his/her user id, name, age,

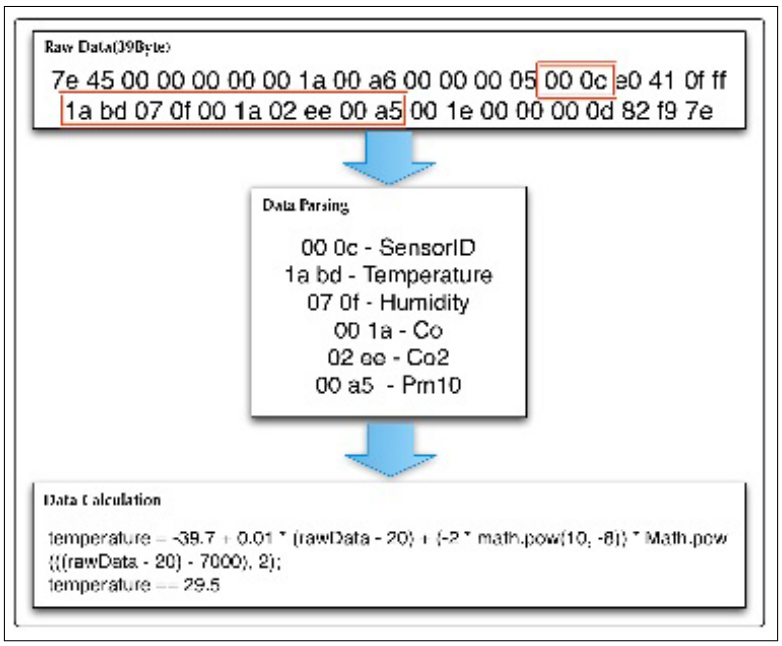


Figure 9: Data parsing

favorite location information, and his/her related medical history are collected. This information is managed by the sensor information management module. The favorite location is used to display a map of the environment when a user logs into the system. Consequently, users do not have to recall their list of most frequented locations or their locations of interest every time. The users age, air-quality related ailments, and his/her favorite location information are used to alert the user about air quality via the smartphone application. Our definition of an event is the occurrence of any unexpected circumstances such as those outside the upper and lower bounds of sensor data or the complete lack of any data from the sensor(s). Such events require to be reported to the system administrator in order to correct the problem. When such an event occurs, the event generator module generates a proper message that contains the event information and sends it to the system administrator. Further, this event generator can be activated when crucial events like fire occur.

The data storage layer comprises three DBs that hold the actual data. Each DB is described in detail in the following section.

5 Global AirScope

The main task of Global AirScope is to provide integrated query and management of the distributed Local AirScope servers and manage sensor and user information. In the early stage of our project, we did not design or accommodate a Global AirScope because our target area (the test-bed) was limited to a small number of sensors (Fig. 10). Therefore, all the system functionalities have been integrated into one Sensor Data Management System. Once our idea is implemented in the Seoul metro area, we plan to divide our system functionalities into two systems Local AirScope to manage sensors and sensor data for small areas and Global AirScope to manage the various Local AirScopes and user information. In this section, we focus on the architecture and the various modules of Global AirScope.

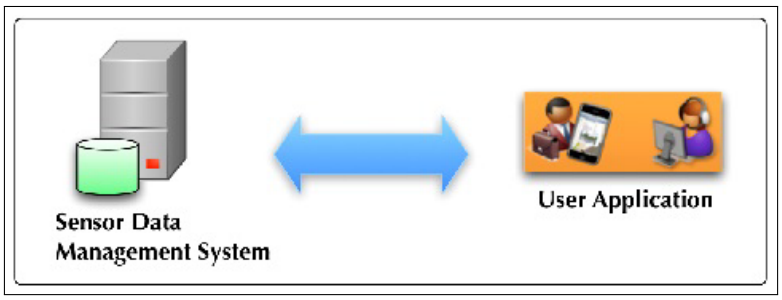


Figure 10: System architecture for existing sensor network

5.1 Architecture of Global AirScope

The main functionalities of Global AirScope include user authentication and authorization, integrated search capability, sensor management, and management of Local AirScopes (Fig. 11). Further, a service creator is added for user notification services. Global AirScope is queried for a service by users. In order to respond to the query, it collects data from Local AirScopes. Global AirScope integrates and processes the collected data and sends it back to the users device.

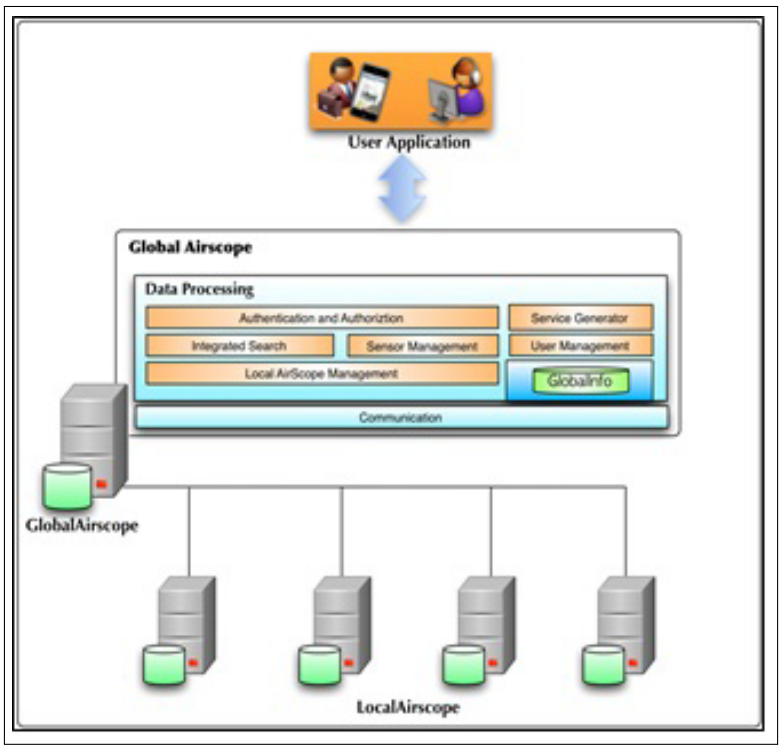


Figure 11: Global AirScope architecture

One of the advantages of the system is the fact that users can use it without registration. However, users are provided with richer personalized services such as user notification and bookmark services upon registration; the registration procedure is simple and easy to use. In addition to general registration, we are collecting any disease information to provide personalized health service. This information can provide health alerts based on the air pollutant rates in his/her current location. Further, users can bookmark their favorite or most frequented locations. When a user logs in to our system, the map automatically displays his/her residence and the surrounding

area.

Because users may request for data from multiple Local AirScopes, we designed Global AirScope to perform and respond to distributed queries. This design provides better user accessibility. When data from multiple Local AirScopes are requested, Global AirScope queries the multiple Local AirScopes and obtains and integrates the relevant data. The integrated data is sent back to the user.

Global AirScope manages Local AirScopes because it receives all the connecting information between Local AirScopes. The main purpose of Local AirScope management service is sensor management [13] [11] (such as automatic sensor failure detection and notification to the system administrator). The manual handling of data from large numbers of sensors is almost impossible, and it may cause critical problems. Under abnormal circumstances such as sensor system failure or environmental disasters, a sensor generates erroneous data.

The event generator module of Local AirScope has been mentioned in a previous section of the paper. If data outside the upper and lower bounds of the sensor are detected, Local AirScope generates an event and sends to it to Global AirScope for user notification. Based on the event, Global AirScope generates a personalized message by utilizing the user information collected during registration; the age, sex, and the relevant medical history of the user are utilized to generate a personal message. This service is designed to send the message to the web portal on a desktop PC and to mobile devices including smartphones, feature phones, and iPads. The user can also receive the message via a windows widget and pop-up windows on the desktop PC without logging into the portal. The message format is generated to suit the capabilities of the users mobile device. For example, we can send a text message to a feature phone, while a combination of text and pictures can be sent to a smartphone.

6 Conclusions

We believe that air quality information about local environments such as homes, workplaces, and parks would be more relevant to people. To this end, we have developed a micro-scale air quality monitoring system called AirScope that covers local environments using a ubiquitous air quality sensor network. A primary feature of AirScope is that it can be run as a smartphone application. Smartphones provide the advantages of mobility and access to information from anywhere at any time. Further, applications for smartphones directly target the end-user, and thus, the smartphone is particularly suited for providing air quality information.

Our desktop application is designed to provide a wide variety of air quality information services. On the other hand, our mobile application is designed to increase user mobility and provide real-time location-based air quality information. Common services for both the desktop and mobile applications include maps, charts, and 3D services. The desktop application provides a rich set of graphics, useful environment information, and high 3D performance, while the strength of the smartphone application is its location-based service.

We provide an overview of the AirScope architecture, including the specifications of sensors, communication between the systems involved, and the need and use of CFD modeling data. Local AirScope is a system that manages the collected data from the sensors and CFD model and provides requested data to the user to provide micro-scale air quality monitoring services. In our study, the purpose of this system was to parse and store data from both the sensors and the CFD mode. The main task of Global AirScope is to provide integrated query and management of the distributed Local AirScope servers and manage sensor and user information.

Bibliography

[1] Woo, J. et al (2011); Constructing u-City of Seoul by future foresight analysis, *Concurrency and Computation: Practice and Experience*, Wiley, ISSN 1532-0626, 23(10):1114-1126.

[2] Woo, J. et al (2011); AirScope: A Micro-scale Urban Air Quality Management System, *Lecture Notes in Computer Science*, Springer, ISBN 978-3-642-24649-4, 520-527.

[3] http://code.google.com.

[4] http://developer.apple.com.

[5] Oreilly, T. (2007); What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *Communications & Strategies*, Digiworld, ISSN 1157-8637, 1:17- 37.

[6] Choi, M. et al(2012); Open API Design for Research Information System, *Lecture Notes in Electrical Engineering*, Springer, ISSN: 1876-1100, 181:187-195.

[7] Liu, K.; Xu, K. (2012); Open Service-Aware Mobile Network API for 3rd Party Control of Network QoS, *Proceedings of the 2012 International Conference on Computer Science and Electronics Engineering (ICCSEE)*, 172-175.

[8] Lu, W.; Liu, Y. (2011); Research of Event-Driven Enterprise Application Integration Base on Service Oriented Architecture, *Applied Informatics and Communication Communications in Computer and Information Science*, Springer, ISSN: 1865-0929, 224:33-37

[9] O'Donovan, D.J.et al (2012); A grid-enabled web service for low-resolution crystal structure refinement, *Acta Crystallographica Section D, Biological Crystallography*, ISSN 0907-4449, 68(3):261-267.

[10] Botts, M. et al (2008); OGC Sensor Web Enablement: Overview and High Level Architecture, *GeoSensor Networks*, Springer, ISBN 978-3-540-79996-2, 4540:175-190.

[11] Rodrigues, T. et al (2012); A Model-Based Approach for Building Ubiquitous Applications Based on Wireless Sensor Network, *Mobile and Ubiquitous Systems*, Springer, ISBN 978-3-642-29154-8, 73:350-352

[12] Nakamura, T. st al (2012); Proposal of Web Framework for Ubiquitous Sensor Network and Its Trial Application Using NO2 Sensor Mounted on Bicycle, *Proceedings of the 2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet (SAINT)*, IEEE, ISBN: 978-1-4673-2001-6, 83-90.

[13] Suganthi, N; Vembu, S.; Energy Efficient Key Management Scheme for Wireless Sensor Networks, *International Journal of Computers Communications & Control (IJCCC)*, ISSN 1841-9844, 9(1):71-78.

[14] Jiang, N. et al (2014); Distributed Compressed Sensing Algorithm for Cluster Architectures of WSNs, *International Journal of Computers Communications & Control (IJCCC)*, ISSN 1841-9844, 9(4):430-438.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):382-389, June, 2015.

CBMIR: Content Based Medical Image Retrieval Using Multilevel Hybrid Approach

B. Ramamurthy, K.R. Chandran

B. Ramamurthy*
Sri Ramakrishna Engineering College, Coimbatore, India
*Corresponding author: hod-ct@srec.ac.in

K.R. Chandran
PSG College of Technology, Coimbatore, India.

Abstract: The problem of retrieval and managing of medical images has become more important due to its scalability and rich information contained in it. In day to day activities, as more as medical images were converted into digital form. Due to its nature, the need for efficient and simple access to this data is essential one. This paper proposed a novel approach namely "content based medical image retrieval using multilevel hybrid approach" to manage and retrieval of this data.This work has been implemented as four levels. In each level, the retrieval performance of the work has been improved. The result of this work has been compared with some of the existing works that are made as literature review on this work.
Keywords: CBMIR, CBIR, euclidean method, precision, recall.

1 Introduction

The rapid expansion of digital image data has led to the need of rich descriptions and efficient retrieval tool. Due to this, many problems were often arises while retrieval of such data in specific applications such as accessing speed, scalability, efficiency etc. To improve this, still more and more research initiative is needed in the field of digital image retrieval system. In medical domain, the amount of image data generation is more and more in day to day activities such as CT, X-ray, and MRI etc. So the scalability of the medical image database has been increased and thereby need of this image retrieval system is essential one. Content based medical image retrieval (CBMIR) is an emerging technique that plays a pivot role in this domain. In this work, shape, texture and intensity contents were used to implementing the system.

2 Related work

CBMIR: shape-based image Retrieval using canny edge Detection and k-means Clustering algorithms for Medical images: In this work, the authors have proposed a shape based image retrieval system for medical images. To implement this, preprocessing, image segmentation, feature extraction, and classification steps were carried out. To retrieval, Euclidian distance calculated between query image and database images. Since the system used shape feature alone, it gives about 50% retrieval performance only [1].

Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms: This system has used a texture based image retrieval system for medical image retrieval. It has capable of retrieving images based on the texture feature of the image. To implement this, the Preprocessing, feature extraction, Classification and retrieval steps are carried out. The main feature of the system is

to utilization of Gray Level Co-occurrence Matrix (GLCM) and k-means clustering algorithm to improve the retrieval efficiency of the system. Since the system used texture feature alone, it gives about 50% retrieval performance only [2].

CBMIR: Content Based Medical Image Retrieval Using Shape and Texture Content: The system proposed a shape and texture based image retrieval system for medical images. It has capable of retrieving images based on the shape and texture feature extraction combined together. The main objective of the system is to improve the retrieval efficiency when compared to single feature method. Since this system has used both shape and texture features combined together,it gives about 50%-90% efficiency [3].

MIRAGE: An E-repository of Medical Images for Learning Biomedical Informatics: This paper proposed an E-repository for medical images that offers great facilities to learn about the biomedical informatics. The facilities of domain-based, atlas-based, and content-based retrieval (CBIR) techniques are implemented to search the images in this developed repository. The uniqueness of the system is CBIR system for 3D is developed and coupled with 3D visualization that has used for educational material and as well as tele-education in future [4].

Computer-aided diagnostics of screening mammography using content-based image retrieval: This paper proposed a computer-aided diagnostics tool for screening the mammography using content based image retrieval technique. In this work, Support Vector Machine (SVM) classification technique was used for doubtful tissue pattern extraction in an image. Based on that, the retrieval of the system implemented using cbir for detecting the mammography of the image [5].

Content-based binary image retrieval using the adaptive hierarchical density histogram: The paper proposed a scheme for binary image retrieval. It has utilized black and white binary represented values as image feature and it named as the adaptive hierarchical density histogram that develops the allocation of the image points on a two-dimensional area. This technique uses the assessment of point density histograms of image regions that are computed by a pyramidal grid that is repeatedly simplified through the calculation of image geometric centroids. This extracted descriptor includes both global and local possessions that can be used in different types of binary image databases for the retrieval of images [6].

Improving the ranking quality of medical image retrieval using a genetic feature selection method: In this paper, the authors proposed a method for improving the ranking quality for medical image retrieval using genetic feature selection method. Here the authors have used single-valued genetic functions for evaluating the rankings to extend a group of feature selection methods based on the genetic algorithm approach to improve the precision of content-based image retrieval systems [7].

3 Proposed Model

The general process method of Content based medical image retrieval system is described as follows. Initially medical images were taken as input to the System and preprocessing of the images carried out in order to improve the flexibility of the images for further processing of the system. Then the output images of this preprocessing step are taken for feature extraction, feature vector construction and feature database construction processes respectively. This Similar process has been applied for query image also (up to feature extraction and feature vector construction process only). While searching the image, the query image feature vector compared with the database image feature vector by using Euclidean distance calculation method. The minimal Euclidean distance value is considered as closest distance image from the database image that ranked as first. The proposed CBMIR framework is shown in Figure 1.

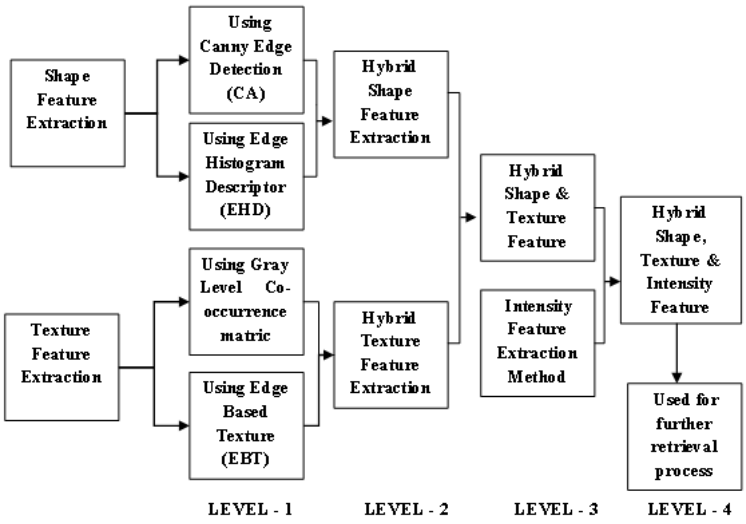


Figure 1: The proposed CBMIR framework

In this proposed approach, there are four levels of results have been derived based on the multilevel hybrid approach method. They are

Level 1. In this level, the content based medical image retrieval system has been implemented based on shape feature extraction method alone and texture feature extraction method alone (ie.single feature extraction method). For shape feature extraction, there are two methods have been implemented. First one is canny edge detection and another one is edge histogram descriptor. Similarly, for texture feature extraction, there are two methods have been implemented. First one is gray level co-occurrence matrix and another one is edge based texture feature extraction. For each method, separate results have been derived for further comparison with the next level in order to measure the performance of the system in each level.

Level 2. In this level, a hybrid feature extraction method implemented that includes a hybrid shape feature extraction method by combining Canny's Edge Detection method (CED) and Edge Histogram Descriptor (EHD) and hybrid texture feature extraction method by combining Gray Level Co-Occurrence Matrix (GLCM) and Edge based texture feature extraction(EBT) in order to extract the best pattern of the images. After obtaining this pattern, the image feature vector value of that pattern has been generated and stored in a database named 'feature database' for further searching and retrieval of the process. While retrieving the images, for each method (hybrid shape, hybrid texture), separate results have been derived for further comparison with the next level in order to measure the performance of the system.

Level 3. In this level, a hybrid shape and texture feature extraction method has been implemented by combining the output result of the hybrid shape feature extraction method and hybrid texture feature extraction method. After combining this, the fine tuned of the both shape and texture pattern of the image has been obtained and then the feature vector value of that pattern also has been generated and stored as a database named 'feature database' for further searching and retrieval of the process. While retrieving the images, the result also has been derived for further comparison with the next level in order to measure the performance of the system.

Level 4. In this level, hybrid intensity feature extraction method has been implemented by combining intensity features with hybrid shape and texture feature extraction method presented in level 3. After combining this, another level of fine tuned image pattern has been obtained

along with combined form of shape, texture and intensity features. The feature vector value for this obtained pattern has been generated and stored as a database named 'feature database' for further searching and retrieval of the process. While retrieving the images, the result also has been derived for further comparison with the next level in order to measure the performance of the system.

Similarly n number of image features can be used to develop n number of levels to measure retrieval accuracy of the system. In this system we have incorporated up to four levels only.

Similarity Comparison

In this approach (in all the levels), for retrieval of the images, similarity comparison technique has been used. For similarity comparison, we have used Euclidean distance, d using the following equation [8, 9].

$$d = \sqrt{\sum_{i=1}^{N}(F_Q[i] - F_{DB}[i])^2} \tag{1}$$

Where $F_Q[i]$ is the i^{th} query image feature and $F_{DB}[i]$ is the corresponding feature in the feature vector database. Here N refers to the number of images in the database.

4 Experimental setup and result analysis

A Intel ® Core 2 Duo CPU Workstation with 2GB RAM computer has been used for conducting the experiments. The MATLAB 7.2.0-Image Processing tool Box was used for developing User Interface components as front end, MATLAB 7.2.0-Image Processing tool Box-Workspace was used as feature database for storage as back end and for image processing work, other MATLAB 7.2.0 utilities were used.

4.1 Retrieval Efficiency

For retrieval efficiency, traditional measures namely precision and recall parameters were computed for several real time medical images [10–12]. Standard formulas that have been computed for determining the precision and recall measures. Precision and Recall parameters can be defined as follows.

Precision (P) is the ratio of the relevant images to the total number of images retrieved

$$P = \frac{r}{n1} \tag{2}$$

where,
r - number of relevant images retrieved
$n1$ - total number of images retrieved.

Recall(R) is the percentage of relevant images among all possible relevant images

$$R = \frac{r}{n2} \tag{3}$$

where,
r - number of relevant images retrieved
$n2$ - total number of relevant images in the database.

By randomly selecting some sample query images from the MATLAB-Image Processing tool Box-Workspace database, the system was tested and the results are shown in the following Tables for all the four levels presented in section 3.

Level 1: In level 1, there are four methods have been implemented as single feature extraction method. The result of all the methods in level 1 has been presented in the following tables. The shape feature extraction method using canny edge detection (CED) method result has been presented in table 1.

Table 1. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	50.00	58.33	68.26	56.95	53.33
Recall	37.50	36.84	51.42	66.66	34.00

The shape feature extraction method using edge histogram descriptor (EHD) method result has been presented in table 2.

Table 2. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	50.00	55.33	65.23	59.84	60.33
Recall	40.50	33.50	45.95	59.00	55.00

The texture feature extraction method using gray level co-occurrence matrix (GLCM) method result has been presented in table 3.

Table 3. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	55.71	65.71	53.33	57.00	56.00
Recall	45.00	48.94	51.42	56.66	60.00

The texture feature extraction method using edge based texture (EBT) method result has been presented in table 4.

Table 4. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	55.71	50.00	65.00	56.33	68.94
Recall	55.66	56.00	65.33	65.00	51.00

In this first level of multi-level hybrid approach, all the approaches such as CED, EHD, GLCM and EBT gives about 50% to 68.94% of the retrieval performance in terms of precision and 33.50% to 66.66% in term s of recall since all the works were in initial stage and also each method gives its own retrieval performance.

Level 2: In level 2, there are two methods have been implemented as hybrid feature extraction method. The result of these methods in level 2 has been presented in the following tables.

The hybrid shape feature extraction using CED and EHD method result has been presented in table 5.

Table 5. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	73.00	70.33	72.26	71.95	73.33
Recall	60.50	68.84	62.42	60.66	61.00

The hybrid texture feature extraction using GLCM and EBT method result has been presented in table 6.

Table 6. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	76.00	70.33	75.26	70.95	78.33
Recall	66.50	60.84	68.42	63.66	66.00

In this, second level of multi-level hybrid approach, the first approach such as the hybrid shape feature extraction using CED and EHD method gives about 70.33% to 73.33% of the retrieval performance in terms of precision and 60.50% to 68.84% in terms of recall. The hybrid texture feature extraction using GLCM and EBT method gives about 70.33% to 78.33% of the retrieval performance in terms of precision and 60.84% to 68.42% in terms of recall. Hence this second level gives better performance than the level 1 performance.

Level 3: In level 3, the hybrid shape & texture feature extraction method has been implemented using the combined form of hybrid shape feature extraction method (HBS) and hybrid texture feature extraction method (HBT).The result of this method in level 3 has been presented in the following table 7.

Table 7. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	94.73	94.44	93.93	92.59	87.50
Recall	90.00	89.47	88.57	83.33	84.00

In this, 3rd level, the performance of the retrieval drastically increased since both the shape and texture features combined together used for retrieval process. When compared to the level 2, this level gives its performance much better and also in terms of precision it gives about 87.50% to 94.73% and in terms recalls about 83.33% to 90.00% of the retrieval performance.

Level 4: In level 4, the hybrid shape, texture and intensity feature extraction method has been implemented by combining the hybrid shape and texture feature extraction method with intensity feature extraction method in order to extend the next level of the work. The result of this method has been presented in the following table 8.

Table 8. Precision and Recall values in %

Query Image	Image 1	Image 2	Image 3	Image 4	Image 5
Precision	95.75	95.00	94.93	93.58	94.65
Recall	91.66	90.10	90.57	92.00	87.00

In this, 4^{th} level, the performance of the retrieval highly increased since all the shape, texture, and intensity features combined together used for retrieval process. When compared to the level 3, this level outperforms well and also in terms of precision it gives about 93.58% to 95.75% and in terms of recall about 87.00% to 92.00% of the retrieval performance. Similarly n-number of image features can be integrated in order to improve the retrieval performance of the system since the image pattern has played a vital role in an image retrieval process.

4.2 Result Comparison and Discussion

The result comparison also made for the above described levels of the work. The following table 9 shows comparison results of the work.

Table 9. The Comparison of the result (Precision and Recall values in %)

Level 1								Level 2				Level 3		Level 4	
Shape (CED)		Shape (EHD)		Texture (GLCM)		Texture EBT		Hybrid Shape (CED+EHD)		Hybrid Texture (GLCM+EBT)		Hybrid Shape & Texture (HBS+HBT)		Hybrid Shape, Texture & Intensity (HBS+HBT+HBI)	
P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R
50.00	37.50	50.00	40.50	55.71	45.00	55.71	55.66	73.00	60.50	76.00	66.50	94.73	90.00	95.75	91.66
58.33	36.84	55.33	33.50	65.71	48.94	50.00	56.00	70.33	68.84	70.33	60.84	94.44	89.47	95.00	90.10
68.26	51.42	65.23	45.95	53.33	51.42	65.00	65.33	72.26	62.42	75.26	68.42	93.93	88.57	94.93	90.57
56.95	66.66	59.84	59.00	57.00	56.66	56.33	65.00	71.95	60.66	70.95	63.66	92.59	83.33	93.58	92.00
53.33	34.00	60.33	55.00	56.00	60.00	68.94	51.00	73.33	61.00	78.33	66.00	87.50	84.00	94.65	87.00

P - Precision, R - Recall

From the above table, the proposed framework has improved results in significantly for content based medical image retrieval system using a new approach called "multilevel hybrid approach". It has been shown that results are in level wise improvement. In level 1, experiments provide about 50% result performance. In level 2, it has been increased to in some extent that is about 25% higher than the levelb1. In level 3, the retrieval performance of the system drastically increased since both the shape and texture features combined together used in retrieval. In level 4, a new feature called 'intensity' has been used along with shape and texture method. It gives a better result than previous level 3. This work also provides an entire framework for inserting and deleting the features as user preferred.

5 Conclusion

On the whole, this proposed system gives its own performance and as well as flexibility for content based medical image retrieval system since the system supports n number of features for pattern recognition of the image. In each level, the retrieval performance of the system also significantly improved. The experimental results show that the medical images have been retrieved in an efficient manner using multilevel hybrid approach method. The system is designed in a flexible and consistent flow for easy understanding. Some of the goals that have been achieved by the developed system are accuracy of retrieved images by designing an efficient framework and incorporating it with the feature extraction techniques.

Bibliography

[1] B. Ramamurthy, K.R.Chandran (2011); CBMIR: Shape-based image Retrieval using canny edge Detection and k-means Clustering algorithms for Medical images, *International Journal of Engineering Science and Technology (IJEST)*, ISSN: 0975-5462, 3(2): 1864-1872.

[2] B. Ramamurthy, K.R.Chandran (2012); Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms, *Journal of Computer Science*, ISSN 1549-3636, 8(7): 1070-1076.

[3] B. Ramamurthy, K.R.Chandran (2013); CBMIR: Content Based Medical Image Retrieval Using Shape and Texture Content, *Advances in Modeling B:* AMSE Press, France, 56(2): 84-95.

[4] Xiaohong Gao, Yu Qian (2012); MIRAGE: An E-repository of Medical Images for Learning Biomedical Informatics, eTELEMED 2012: *The Fourth International Conference on eHealth, Telemedicine, and Social Medicine, January 30, 2012 to February 4*, 2012, ISBN: 978-1-61208-179-3, Valencia, Spain, 209 - 214

[5] Thomas M. Deserno, Michael Soiron, Júlia E. E. de Oliveira, and Arnaldo de A. Araújo(2012); Computer-aided diagnostics of screening mammography using content-based image retrieval, *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, Vol 8315, Medical Imaging 2012: Computer-Aided Diagnosis,(February23,2012), San Diego, California, pp.831527-831527-9.

[6] Panagiotis Sidiropoulos, Stefanos Vrochidis, Ioannis Kompatsiaris (2011); Content-based binary image retrieval using the adaptive hierarchical density histogram, *Pattern Recognition*, 44(4): 739 - 750.

[7] Sérgio Francisco da Silva, Marcela Xavier Ribeiro,João do E.S. Batista Neto, Caetano Traina-Jr.,Agma J.M. Traina(2011); Improving the ranking quality of medical image retrieval using a genetic feature selection method, *Decision Support Systems*, 51(4): 810 - 820.

[8] Mohammed Eisa,Ibrahim Elhenawy, A. E. Elalfi and Hans Burkhardt(2006); Image Retrieval based on Invariant Features and Histogram Refinement, *International Journal on Graphics, Vision and Image Processing (ICGST)*, pp. 6: 7-11.

[9] M. Flickner, et al (1995); Query by Image and Video Content: The QBIC System, *IEEE Computer*, 28(9): 23-32.

[10] Xiaohui Yang, Lijun Cai (2014); Adaptive Region Matching for Region- based Image Retrieval by Constructing Region Importance Index, *IET Comput. Vis*, 8(2): 141-151.

[11] Bor-Chun Chen,Yan-Ying Chen et al (2013); Scalable Face Image Retrieval Using Attribute-Enhanced Sparse Codewords, *IEEE Transactions On Multimedia*, 15(5): 1163-1173.

[12] Subrahmanyam Murala et al (2014); Local Mesh Patterns Versus Local Binary Patterns: Biomedical Image Indexing and Retrieval, *IEEE Journal of Biomedical and Health Informatics*, 18(3):929-938.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):390-402, June, 2015.

A Reference Dataset for Network Traffic Activity Based Intrusion Detection System

R. Singh, H. Kumar, R.K. Singla

Raman Singh*, Harish Kumar and R.K. Singla
University Institute of Engineering and Technology
Panjab University, Chandigarh, India
raman.singh@ieee.org, harishk@pu.ac.in, rksingla@pu.ac.in
*Corresponding author: raman.singh@ieee.org

Abstract: The network traffic dataset is a crucial part of anomaly based intrusion detection systems (IDSs). These IDSs train themselves to learn normal and anomalous activities. Properly labeled dataset is used for the training purpose. For the activities based IDSs, proper network traffic activity labeled dataset is the first requirement, however non-availability of such datasets is bottlenecked in the field of IDS research. In this experiment, a synthetic dataset "Panjab University - Intrusion Dataset (PU-IDS)" is created. The purpose of this study is to provide the researchers a reference dataset for the performance evaluation of network traffic activity based IDSs. University of New Brunswick Network Security Laboratory - Knowledge Disscovery in Databases (NSL-KDD) is a benchmark dataset for anomaly detection but it does not contain activity based labeling. So basic characteristics of this dataset are taken for the generation of the new synthetic dataset with various activities based labels. The dataset is first categorized as per *protocol* and *service*. Thereafter, as per minimum & maximum values of attributes, activity profiles are synthetically generated. This paper also discusses various statistical characteristics of PU-IDS. The total number of 198533 instances along with 273 of activity profiles are created. This dataset also contain different 98 *protocol_ service* profiles.
Keywords: Intrusion Detection System, Network Traffic Dataset, Network Traffic Profiling, Behavioral Profiling, Traffic Activity profiling

1 Introduction

Anomaly based Intrusion Detection System (IDS) is an emerging research interest for network security researchers and professionals. In order to detect intrusions, IDSs learn the network traffic activities/ behaviors of malware rather than rely on virus definitions and security updates. Well labeled network traffic dataset is used for this training. This dataset is crucial and only properly labeled dataset can serve this purpose. Profiling based anomaly detection is a new and emerging field of network security research, but, no standard network traffic activities based labeled dataset is available for the training and performance testing.
Some available datasets are internet traces and un-labeled while others are only single level labeled (normal or anomalous). Despite the wide availability of these datasets, non-availability of network traffic activities based datasets is hindrance in the intrusion detection research. In the anomaly based IDSs, normal behavior of networks may differ for different kind of organizations. For example, network game packets (online or intranet games) may be malicious for consultancy companies while they are absolutely normal for game development and testing companies. In the same manner within an organization, different departments may have different network traffic patterns. There is need of development of such anomaly detection techniques which can identify such network traffic activities/behaviors and create normal (and/or anomalous) profiles for intrusions detection. No such dataset exists for this purpose. The synthetic dataset should be generated so as to be used as a reference dataset for testing and performance evaluation. In this paper, University of New Brunswick Network Security Laboratory - Knowledge Disscovery

in Databases (NSL KDD) network traffic dataset is taken as a base and "Panjab University - Intrusion Dataset (PU-IDS)" network traffic dataset is generated synthetically. NSL KDD is benchmark dataset but does not contains network traffic activities based labels. These labels are introduces in PU-IDS.
This paper is divided into five Sections; Section 1 introduces the topic, Section 2 discusses the various available network traffic datasets. This Section also discusses available network traffic profiling based threat detection techniques. Section 3 describes the dataset used. This section also explain the methodology of synthetic dataset generation. In Section 4, statistical analysis of synthetic dataset is carried out. Finally, Section 5 discusses the conclusions and future works.

2 Network Traffic Dataset and Profiling : Related Work

2.1 Network Traffic Dataset

In the last few years, some datasets are either generated or collected for research purpose. These datasets are used by researchers worldwide. The Cyber Systems and Technology Group (formerly the DARPA Intrusion Detection Evaluation Group) of MIT Lincoln Laboratory has collected network traffic dataset in 1998 and 1999. This dataset is collected by simulating various attacks like denial of services (DOS), remote to local (R2L), user to remote (U2R) etc. on different platforms like Windows, Unix etc. This dataset become benchmark KDD dataset for research in IDS [1]. Later, some of the problems like redundant and duplicate records present in this dataset are removed and more efficient dataset become available for researcher which is known as NSL KDD dataset [2]. This dataset is widely used for performance analysis of various intrusion detection techniques. In this dataset each instances are labeled as normal or anomalous. In another version of this dataset, instances are classified as par various attacks like dos,U2R, L2R, probe and others. The drawback of this dataset is that, no further network traffic activities/behaviors based labels are available. The Stanford Network Analysis Project created 50 network traffic datasets by using various nodes of social networks, internet work, web graphs, etc. Some of this collection of datasets is labeled while others are un-labeled. Profiling based classification is missing in this dataset [3].
The dataset is created by the Center for Applied Internet Data Analysis (CAIDA) at various topologically and geographically separated locations. Anonymized internet traces along with other worms related dataset is created by ensuring privacy preservation from the year 2008 to the year 2013. This is a collection of various datasets like anonymized internet traces and attack specific dataset. The internet traces dataset is un-labeled and so may not be useful directly for performance analysis of IDS. It needs to be the first pre-processed for labeling [4]. CERT synthetic *sendmail* system dataset is created by the University of New Mexico using Sun SPARC stations. Some *system call* instances are live while others are synthetic. Normal instances and intrusion traces are provided separately. Further level of labeling is not provided in this dataset [5]. Real traffic of HTTP, SMTP, SSH, IMAP, POP3, and FTP are used to create a network traffic dataset in the University of New Brunswick Information Security ? Centre of Excellence (UNB - ISCX) lab. This dataset which contains seven days of normal and malicious instances is known as UNB ISCX dataset [6]. This dataset is available on request for university researchers. This dataset includes normal and malicious instances. Profiling level labeling is not available for this dataset. Internet traces of the sub-network of Panjab University network (PU-CAN) is captured from July 2011 to August 2011 in order to create a network traffic dataset. This dataset consists of un-labeled internet traces which needs further processing for labeling [7]. Internet traffic traces like LAN/WAN Ethernet traffic, TCP traffic, HTTP traffic, HTTP logs etc., are available in other dataset. The collection of datasets includes LAN/WAN traces, various

logs and web client traces. A further level of classification and labeling is not present in these datasets [8].
From the literature studied it has been found that most of the datasets are not labeled on the basis of network traffic activities/behaviors. These datasets are labeled on single level profiling. Therefore, there is a need for the development of a dataset which is based on the behavior profiling that consider network traffic activities and various user's behaviors.

2.2 Network Traffic Profiling

Researchers are using clustering and profiling in order to detect intrusions. Network traffic activities based IDSs are the future of network security. In this sub-section various techniques which are used to detect intrusions are discussed. Clustering is used in the network traffic dataset in order to cluster various instances into different activity profiles. These profiles are used to determine normal and malicious instances [9]. Behavioral distance based anomaly detection for real time traffic analysis is also proposed. Horizontal and vertical distance metrics are used for different attributes of network traffic datasets [10]. Behavioral foot printing method along with content based signature is used to profile self-propagating worms. The worm's dynamic infection sequence is learned to detect it [11].
Clustering based anomaly detection technique is proposed for modeling user's normal behavior in [12]. Traffic causality graphs (TCGs) are used to analyze and visualize temporal and spatial causality of flows to profile network traffic. The advantage of this technique is that there is no need of payload inspection [13]. Personal and application profiles are created by tagging network traffic. Traffic from known source is profiled with role tag and application tag [14]. IP to IP communication graph and information is used to profile internet backbone traffic. This technique is known as profiling by association [15]. Data mining and entropy based techniques are used to profile the behavior of internet end to end hosts [16]. Researchers also proposed behavior based tracking to create long term profiles of user?s interest. This methodology is tested on DNS traffic [17].
These techniques can be used on network traffic activity based datasets in order to detect various normal and anomalous network traffic activities/behaviors. In the state of the art IDSs, the normal behavior of user's groups needs to be learned to effectively detect intrusions.

3 Materials and Methods

Unavailability of traffic activity/ behavior level network traffic dataset motivates this experiment of generation of synthetic dataset. The network traffic dataset used for generation of the synthetic dataset is discussed as below:

3.1 Network Traffic Dataset Used

Benchmarked NSL-KDD network traffic dataset is used as a base and its various characteristics of different attributes are used to synthetically generate instances. The various characteristics used are minimum and maximum values of different attributes. Table 1 shows the list of attributes of this dataset. The same attributes are taken in PU-IDS. These attributes are either continuous or categorical type. Continuous attributes are those in which the value belongs to indefinite set while in categorical attribute values are assigned from a definite set. Attribute number 2, 3, 4, 7, 12, 14, 15, 21, 22 and 42 are categorical attributes while all others are continuous attributes.

Table 1: List of Attributes of NSL KDD and Synthetically generated dataset

Sr. No.	Feature	Sr. No.	Feature	Sr. No.	Feature
1	Duration	15	Su attempted	29	Same srv rate
2	Protocol type	16	Num root	30	Diff srv rate
3	Service	17	Num file creations	31	Srv diff host rate
4	Flag	18	Num shells	32	Dst host count
5	Source bytes	19	Num access files	33	Dst host srv count
6	Destination bytes	20	Num outbound cmds	34	Dst host same srv rate
7	Land	21	Is host login	35	Dst host diff srv rate
8	Wrong fragment	22	Is guest login	36	Dst host same src port rate
9	Urgent	23	count	37	Dst host srv diff host rate
10	Hot	24	Srv count	38	Dst host serror rate
11	Number failed logins	25	Serror rate	39	Dst host srvserror rate
12	Logged in	26	Srvserror rate	40	Dst host rerror rate
13	Num compromised	27	Rerror rate	41	Dst host srvrerror rate
14	Root shell	28	Srvrerror rate	42	Class label

3.2 Synthetic Dataset Generation Methodology

Figure 1 shows the methodology of synthetic dataset generation. It describes the procedures followed in the experiment to generate instances of synthetic dataset. This methodology is implemented using Matlab [18] as a tool. The various steps are described as below:

Separation of dataset

In the first step, NSL KDD training and testing dataset is clubbed into a single set. This dataset is then separated into two categories of normal and anomalous sub-datasets.

Protocol_ service profile creation

In the second step, *protocol_ service* profiles are created by integrating protocol and service attributes for both categories of sub-datasets (normal and anomalous sub-datasets).

All *protocol_ service* profiles created in the first level of profiling are shown in table 2. Thereafter, all instances are separated as per *protocol_ service* profiles for both normal and anomalous sub-datasets.

Extraction of basic characteristics

In the third step, all continuous attributes of all *protocol_ service* profiles of both sub-datasets are considered and basic characteristics like minimum and maximum values are extracted. Unique values for each categorical attributes are also calculated.

Calculation of *cluster_ gap* and *number_ of_ instances_ per_ cluster*

In the fourth step, for each *protocol_ service* profile, number of clusters and the number of instances to be created are taken as input from the user.

As per equation 1 *cluster_ gap* and as per equation 2 *number_ of_ instances_ per_ cluster* is calculated for each *protocol_ service* profile.

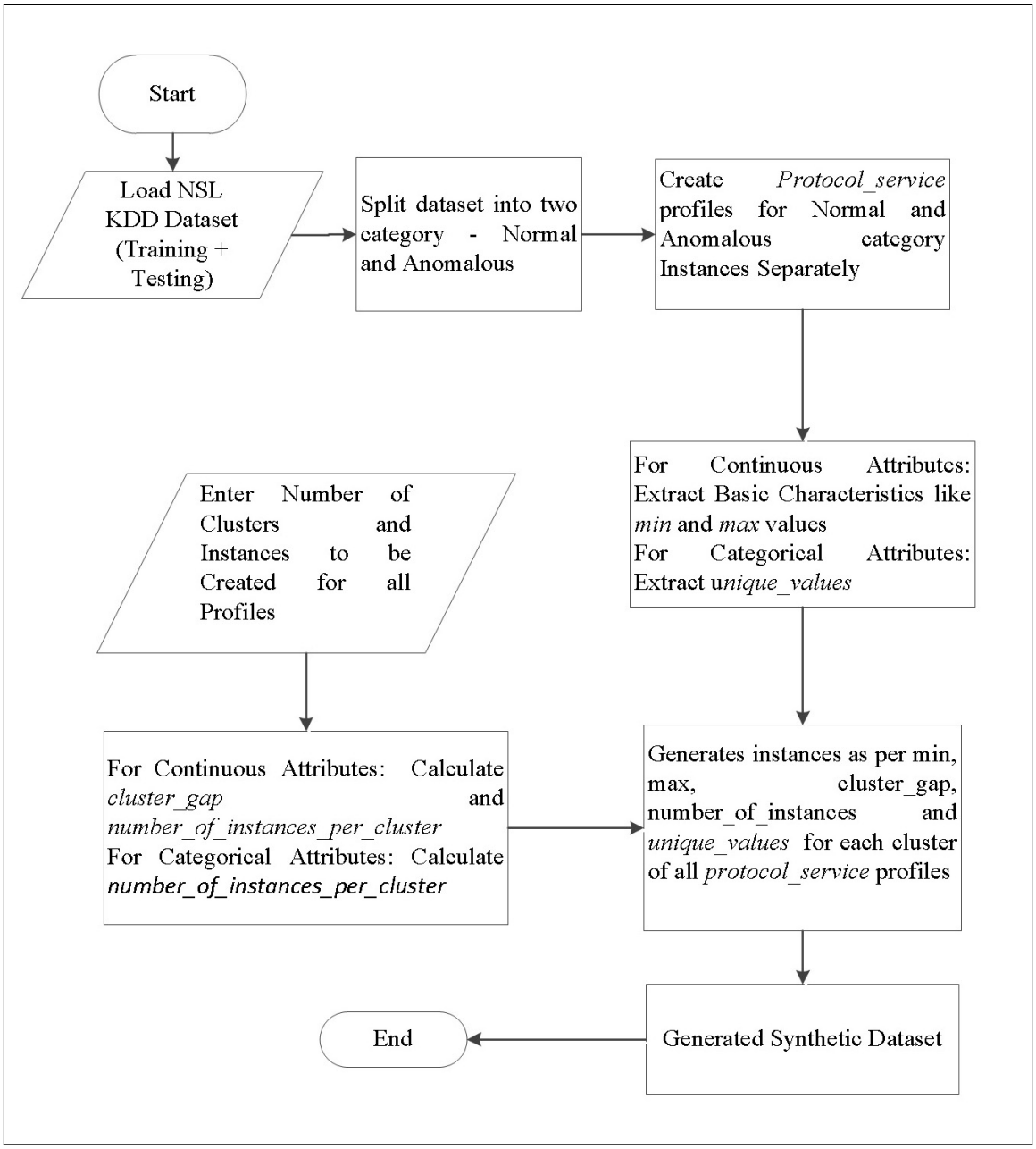


Figure 1: Synthetic dataset generation methodology

Table 2: List of *protocol_ service* Profiles

icmp_eco_i	tcp_courier	tcp_gopher	tcp_link	tcp_pop_2	tcp_systat
icmp_ecr_i	tcp_csnet_ns	tcp_harvest	tcp_login	tcp_pop_3	tcp_telnet
icmp_red_i	tcp_ctf	tcp_hostnames	tcp_mtp	tcp_printer	tcp_time
icmp_tim_i	tcp_daytime	tcp_http	tcp_name	tcp_private	tcp_uucp
icmp_urh_i	tcp_discard	tcp_http_2784	tcp_netbios_dgm	tcp_remote_job	tcp_uucp_path
icmp_urp_i	tcp_domain	tcp_http_443	tcp_netbios_ns	tcp_rje	tcp_vmnet
tcp_IRC	tcp_echo	tcp_http_8001	tcp_netbios_ssn	tcp_shell	tcp_whois
tcp_X11	tcp_efs	tcp_imap4	tcp_netstat	tcp_smtp	udp_domain_u
tcp_Z39_50	tcp_exec	tcp_iso_tsap	tcp_nnsp	tcp_sql_net	udp_ntp_u
tcp_aol	tcp_finger	tcp_klogin	tcp_nntp	tcp_ssh	udp_other
tcp_auth	tcp_ftp	tcp_kshell	tcp_other	tcp_sunrpc	udp_private
tcp_bgp	tcp_ftp_data	tcp_ldap	tcp_pm_dump	tcp_supdup	udp_tftp_u

In these equations x_{max} is the maximum and x_{min} is the minimum value extracted in step three for a particular *protocol_ service* profile. $k_{cluster}$ is the number of clusters required and *number_ of_ instances_ protocol_ service* is the number of instances required for particular *protocol_ service* profile which are taken from the user.

$$cluster\_gap = \frac{x_{max} - x_{min}}{k_{cluster}} \tag{1}$$

$$number\_of\_instances\_per\_cluster = round\{\frac{number\_of\_instances\_protocol\_service}{k_{cluster}}\} \tag{2}$$

Table 3 shows various parameters like $k_{cluster}$ and *number_ of_ instances_ protocol_ service* used to generate synthetic dataset.

In table 3, *Cluster_ N* ($k_{cluster}$) and *Insta_ N* (*number_ of_ instances_ protocol_ service*) corresponds to the number of clusters and instances to be created for normal dataset respectively.

Cluster_ A and *Insta_ A* is the number of clusters and instances to be created for anomalous dataset respectively.

Table 3: Synthetic Dataset Generation Parameters

Protocol Service Profiles	Cluster _N	Insta _N	Cluster _A	Insta _A	Protocol Service Profiles	Cluster _N	Insta _N	Cluster _A	Insta _A
icmp_eco_i	3	1215	6	4200	tcp_link	1	5	4	1536
icmp_ecr_i	2	456	4	3120	tcp_login	0	0	3	855
icmp_red_i	1	125	0	0	tcp_mtp	0	0	3	858
icmp_tim_i	1	220	1	42	tcp_name	0	0	4	780
icmp_urh_i	1	150	0	0	tcp_netbios_dgm	0	0	4	2920
icmp_urp_i	4	1200	1	45	tcp_netbios_ns	0	0	3	960
tcp_IRC	3	900	1	50	tcp_netbios_ssn	0	0	2	1048
tcp_X11	1	175	1	124	tcp_netstat	0	0	3	1305
tcp_Z39_50	0	0	5	2800	tcp_nnsp	0	0	4	2600
tcp_aol	0	0	1	50	tcp_nntp	0	0	2	708
tcp_auth	2	450	4	952	tcp_other	2	460	5	1850
tcp_bgp	0	0	6	2100	tcp_pm_dump	0	0	1	135
tcp_courier	0	0	7	2100	tcp_pop_2	0	0	1	235
tcp_csnet_ns	0	0	5	3270	tcp_pop_3	1	350	1	115
tcp_ctf	0	0	4	800	tcp_printer	0	0	1	354
tcp_daytime	0	0	6	1800	tcp_private	1	22	6	28000
tcp_discard	0	0	4	1680	tcp_remote_job	1	5	1	154
tcp_domain	1	255	3	405	tcp_rje	0	0	1	204
tcp_echo	0	0	4	1040	tcp_shell	1	16	1	165
tcp_efs	0	0	5	1850	tcp_smtp	5	9600	2	386
tcp_exec	0	0	3	1395	tcp_sql_net	0	0	2	840
tcp_finger	3	375	5	1750	tcp_ssh	1	20	2	400
tcp_ftp	4	1616	4	1800	tcp_sunrpc	0	0	2	856
tcp_ftp_data	6	3006	5	2000	tcp_supdup	0	0	3	1404
tcp_gopher	0	0	3	570	tcp_systat	0	0	2	1206
tcp_harvest	0	0	1	26	tcp_telnet	5	5750	3	1836
tcp_hostnames	0	0	2	500	tcp_time	1	170	2	680
tcp_http	6	50010	4	2800	tcp_uucp	0	0	3	2085
tcp_http_2784	0	0	1	12	tcp_uucp_path	0	0	3	2346
tcp_http_443	0	0	3	999	tcp_vmnet	0	0	3	1875
tcp_http_8001	0	0	1	14	tcp_whois	0	0	3	2256
tcp_imap4	1	160	3	888	udp_domain_u	5	11825	1	14
tcp_iso_tsap	0	0	4	820	udp_ntp_u	1	320	0	0
tcp_klogin	0	0	3	924	udp_other	4	2624	2	304
tcp_kshell	0	0	2	650	udp_private	3	1212	4	2600
tcp_ldap	0	0	2	360	udp_tftp_u	1	35	0	0

Synthetic instances generation

In the fifth step, for each continuous attributes of various *protocol_service* profiles, $k_{cluster}$ numbers of traffic activity/ behavior profiles are created by using x_{min}, x_{max} and $k_{cluster}$.

For categorical attributes of various *protocol_service* profiles $k_{cluster}$ number of traffic activity/ behaviors are created by considering *cate_unique_values*.

For each traffic activity profile, *number_of_instances_per_cluster* numbers of instances are

generated by using same characteristics used for generation of various traffic activity/ behavior profiles.

Values of continuous attributes for each cluster is calculated as given in equation (3)

$$value\_conti\_Attribute = max\_value\_previous\_cluster + cluster\_gap \tag{3}$$

Subject to:
$max\_value\_previous\_cluster = x_{min}$, for first cluster
and
$x_{min} \geq value\_conti\_attribute \leq x_{max}$ for other clusters

Where *value_conti_attribute* represents value of a particular continuous attribute and *max_value_previous_cluster* represents the maximum value of previous cluster of that particular attribute.

For categorical attributes *cate_unique_values* are extracted by taking each unique value for each categorical attribute separately as shown in equation (4).

$$value\_cate\_Attribute = div\{cate\_unique\_values(k_{\text{cluster}})\} \tag{4}$$

Function *div* means values of particular categorical attributes are obtained by dividing *cate_unique_value* into cluster groups and then for each group/ cluster *number_of_instances_per_cluster* instances are created.

3.3 Illustrative Example

let?s say for *tcp_http protocol_service profile*, 3 clusters (k_{cluster}) and 1000 instances (*number_of_instances_protocol_service*) are required to be generated.

So the values of x_{min} and x_{max} for all continuous attributes are extracted. Also assume the values of x_{min} and x_{max} for attribute *source_bytes* are 30 and 255. So *cluster_gap* and *number_of_instances_per_cluster* are calculated as per equation (1) and (2) respectively as shown in example 1.

Example 1. *Cluster_gap = (255-30) /3 = 75*
number_of_instances_per_cluster = round (1000/3) = 333
Range of First Cluster value_conti_attribute1 = (from 30 up to 30+75=105)
Range of Second Cluster value_conti_attribute2 = (from 105 up to 105+75=180)
Range of Third Cluster value_conti_attribute3 = (from 180 up to x_{max} = 255)

For each continuous attribute , 333 instances per cluster are created and the same procedure is followed for all continuous attributes (Last cluster may have one additional instance to sum up total instances to 1000). Example 2 explain the assignment of values of different clusters for categorical attributes (like *flag*).

Example 2. *cate_unique_values = S0, S1 ,S2, S3, S4, S5, S6, S7, S8 (assumed value)*
For First Cluster, value_cate_attribute1 = S0, S1, S2
For Second Cluster, value_cate_attribute2= S3, S4, S5
For Third Cluster, value_cate_attribute3= S6, S7, S8

Now for each categorical attribute, 333 instances per cluster are created by taking value from *value_cate_attribute* for all categorical attributes (Last cluster may have one additional instance to sum up total instances to 1000). These procedures are followed for each continuous

and categorical attributes of all *protocol_ service* profiles and for both categories of sub-datasets. Both categories of normal and anomalous synthetic sub-datasets are integrated to obtain one synthetically generated dataset. This dataset has two levels of profile labeling. One level is *protocol_ service* while other is *traffic activity/ behavior* level.

4 Results And Analysis

The NSL KDD (Training + Testing) dataset taken for experimentation has 148517 instances and 72 *protocol_ service* profiles. The same numbers of protocol *protocol_ service* are created in synthetic dataset. Table 4 shows the comparative statistical analysis of NSL KDD and PU-IDS dataset.

Table 4: Basic statistics of NSL KDD and PU-IDS dataset

Statistics	NSL KDD Dataset		PU-IDS	
	Normal	Anomaly	Normal	Anomaly
Numbers of instances	77054	71463	92727	105806
Numbers of *protocol_ service* profiles	30	68	30	68
Numbers of network traffic activity profiles	Not present	Not present	72	201

In NSL KDD dataset, 30 normal and 68 anomalous *protocol_ service* profiles are present. Some of these profiles are common in both normal and anomalous sub-datasets. Total 198533 instances are generated synthetically out of which 92727 are normal while other are anomalous. In PU-IDS, the generated instances are more than the base dataset. These excess instances provide an opportunity of better and effective training of IDS as dimensions of normal and anomalous behavior is increased. These instances also helps in performance testing of IDS in broad spectrum. Equal numbers of *protocol_ service* profiles as of base dataset are created in PU-IDS dataset. Generated dataset has equal first level of profiles. In NSL KDD dataset, traffic activity labeling are not present and hence this second level of labeling is synthetically generated. Different 273 traffic activities in datasets are synthetically generated as shown in table 4. 72 normal and 201 anomalous behavior profiles are generated. This second level of labeling will help the researcher to train and test anomaly detection techniques, where NSL KDD provides limited opportunity. Figure 2 shows the number of normal instances generated for each *protocol_ service* profile.

53.93% of the generated normal instances are of *tcp_ http protocol_ service*. This profile is the biggest contributor in normal instances. The different significant normal *protocol_ service* profiles are *udp_ domain_ u* (12.75%), *tcp_ smtp* (10.35%), *tcp_ telnet* (6.2%), *tcp_ ftp_ data* (3.24%), *udp_ other* (2.83%), *tcp_ ftp* (1.74%), *icmp_ eco_ i* (1.31%), *udp_ private* (1.31%) and *icmp_ urp_ i* (1.29%). Other *protocol_ service* (*tcp_ IRC, tcp_ other, icmp_ ecr_ i, tcp_ auth, tcp_ finger, tcp_ pop_ 3, udp_ ntp_ u, tcp_ domain, icmp_ tim_ i, tcp_ X11, tcp_ time, tcp_ imap4, icmp_ urh_ i, icmp_ red_ i, udp_ tftp_ u, tcp_ private, tcp_ ssh, tcp_ shell, tcp_ link, and tcp_ remote_ job*) collectively contribute 5.04% of total normal instances. Only 0.0053% of *tcp_ remote_ job* instances are generated which is lowest in normal instances.

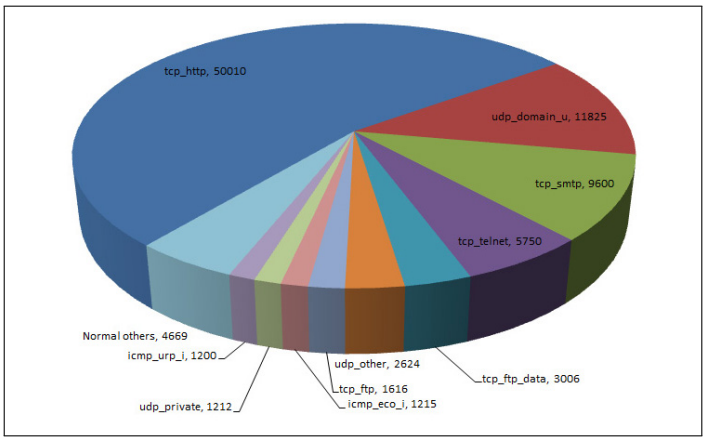


Figure 2: Distribution of normal instances as per *protocol_ service* profiles

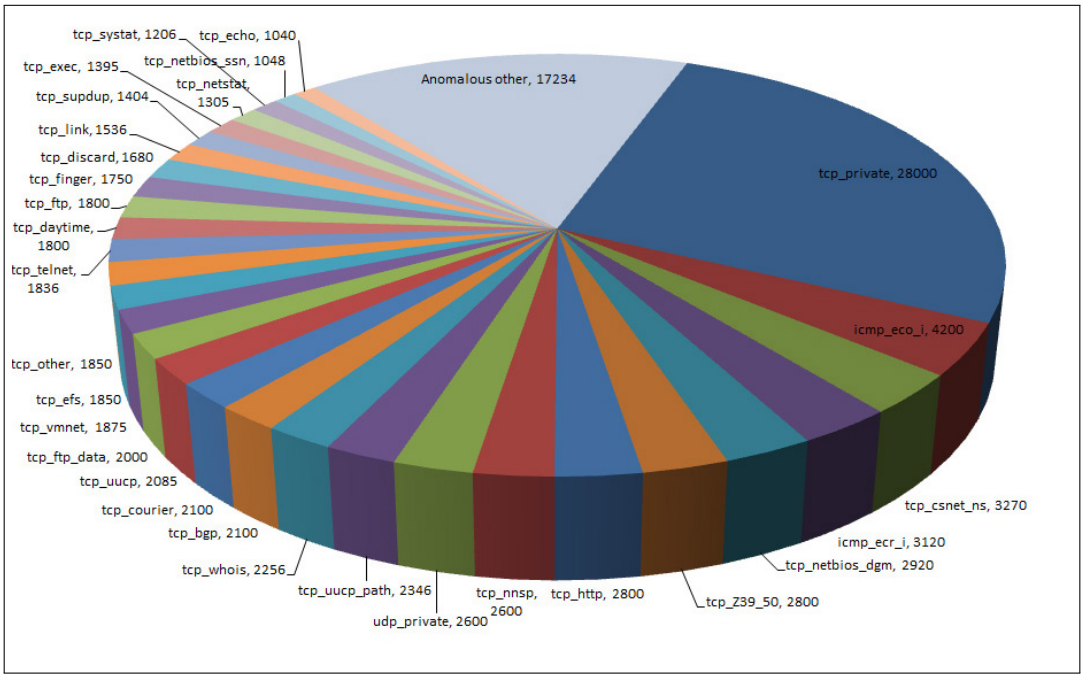


Figure 3: Distribution of anomalous instances as per *protocol_ service* profiles

Figure 3 shows the number of anomalous instances generated for each *protocol_service* profile. In the anomalous instances the *tcp_private* profile has largest share, which is 26.46%. The various significant anomalous *protocol_service* profiles are *icmp_eco_i* (3.97%), *tcp_csnet_ns* (3.09%), *icmp_ecr_i* (2.95%) and *tcp_netbios_dgm* (2.76%). Other anomalous *protocol_service* profiles (like *tcp_http_443, tcp_netbios_ns, tcp_auth, tcp_klogin, tcp_imap4, tcp_mtp, tcp_sunrpc, tcp_login, tcp_sql_net, tcp_iso_tsap, tcp_ctf, tcp_name, tcp_nntp, tcp_time, tcp_kshell, tcp_gopher, tcp_hostnames, tcp_domain, tcp_ssh, tcp_smtp, tcp_ldap, tcp_printer, udp_other, tcp_pop_2, tcp_rje, tcp_shell, tcp_remote_job, tcp_pm_dump, tcp_X11, tcp_pop_3, tcp_IRC, tcp_aol, icmp_urp_i, icmp_tim_i, tcp_harvest, tcp_http_8001, udp_domain_u, and tcp_http_2784*) contribute 16.29% of total anomalous instances. Least number of anomalous instances of *tcp_http_2784 protocol_service* (only 0.011%) are generated.
Figure 4 shows the comparative study of various protocols like *tcp*, *udp* and *icmp*, present in NSL-KDD and synthetically generated PU-IDS dataset. In the base dataset 121569 of *tcp*, 17614

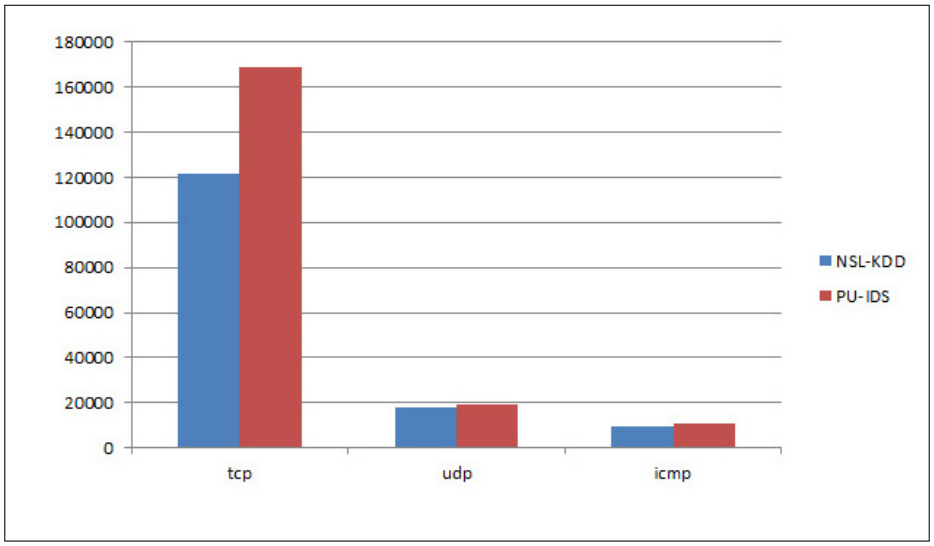


Figure 4: Numbers of instances of *tcp*, *udp* and *icmp* protocols

of *udp* and 9334 of *icmp* instances are present. In PU-IDS, the numbers of synthetically generated instances of *tcp*, *udp* and *icmp* are 168826, 18934 and 10773 respectively.

5 Conclusions and Future Works

The network traffic dataset is an important part of Intrusion Detection System as it learns normal and anomalous behavior of computer networks. Perfect training of IDSs depends on the proper labeled dataset. The well-known benchmark NSL KDD dataset is widely used in IDS research, but this dataset is very old and only has one level of labeling. The other available unlabeled datasets has little importance in performance evaluation of malware detection techniques. Traffic activity labeled datasets are not present for research purpose. Activity/ behavioral based network traffic datasets are needed for state of the art IDSs. A new synthetic network traffic dataset (PU-IDS) with two levels of labeling is generated by taking basic characteristics of the NSL KDD dataset. "Protocol service" level and "traffic activity/ behavior" level labeling is created so as this dataset can be used for performance evaluation. This will overcome limited utility of NSL KDD network traffic dataset. PU-IDS consists of total numbers of 198533 instances along

with 72 protocol profiles. It also consists 273 synthetically created traffic activity/ behavioral profiles which is the novelty of this dataset. In the future, an organizational network with different departments and various user groups should be set up to create the simulated dataset. Users within one organization should exhibit similar network activities/ behaviors as they are working on the same set of software. Behavior based IDSs should identify these similar network activities and create normal or anomalous model.

Acknowledgment

This research work is supported by World Bank funded Technical Education Quality Improvement Program ? Phase II (TEQIP - II) in the form of research assistantship. This research is revised and extended version of paper presented at International Conference on Computers, Communications and Control (ICCCC 2014) held at Oradea, Romania on May 7-9, 2014.

Bibliography

[1] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[2] http://nsl.cs.unb.ca/NSL-KDD

[3] http://snap.stanford.edu/data

[4] http://www.caida.org/data/overview

[5] http://www.cs.unm.edu/ immsec/data

[6] http://www.iscx.ca/datasets

[7] Singh, R., Kumar H., Singla R.K (2012); Traffic Analysis of Campus Network for Classification of Broadcast Data. *47th Annual National Convention of Computer Society of India. Int. Conf. on Intelligent Infrastructure*, MacGraw Hill Professional: 163-166.

[8] http://ita.ee.lbl.gov/html/traces.html

[9] Marchette, D. (1999); A Statistical Method for Profiling Network Traffic, *Workshop on Intrusion Detection and Network Monitoring* : 119-128.

[10] Sengar, H.; Wang, X.; Wang, H.; Wijesekera, D.; Jajodia, S. (2009); Online detection of network traffic anomalies using behavioral distance, *17th Int. Workshop on Quality of Service* : 1-9.

[11] Jiang, X.; Zhu X. vEye (2009); Behavioral footprinting for self-propagating worm detection and profiling, *Knowledge and information systems* ; 18(2): 231-262

[12] Oh, H.S.; Lee, W.S. (2003); An anomaly intrusion detection method by clustering normal user behavior, *Computers & Security*, 22(7): 596-612.

[13] Asai, H.; Fukuda, K. ; Esaki, H. (2011); Traffic causality graphs: profiling network applications through temporal and spatial causality of flows, *Proc. of the 23rd Int. Teletraffic Congress* : 95-102.

[14] Zoquete, A.; Correia, P.; Shamalizadeh, H. (2011); Packet tagging system for enhanced traffic profiling. *IEEE 5th Int. Conf. on Internet Multimedia Systems Architecture and Application (IMSAA)* : 1-6.

[15] Iliofotou, M.; Gallagher, B.; Eliassi-Rad, T.; Xie, G.; Faloutsos, M.(2010); Profiling-by-association: a resilient traffic profiling solution for the internet backbone. *Proc. of the 6th Int. Conference Co-NEXT'10* : DOI: 10.1145/1921168.1921171.

[16] Xu, K.; Zhang, Z.L.; Bhattacharyya S.(2008); Internet traffic behavior profiling for network security monitoring. *IEEE/ACM Trans. on Networking*, 16(6): 1241-1252.

[17] Herrmann, D.; Banse, C.; Federrath, H.(2013); Behavior-based tracking: Exploiting characteristic patterns in DNS traffic. *Computers & Security*, 39 (Part A): 17-33.

[18] http://www.mathworks.in/products/matlab

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):403-419, June, 2015.

Performance Comparison of TCP Spoofing and End to End Approach to Enable Partial QoS on IP Based Network

Y. Suryanto, R.R. Nasser, R.F. Sari

Y. Suryanto*, **R.R. Nasser**, **R.F. Sari**
1. Department Electrical Engineering Faculty of Engineering,
University of Indonesia, Depok, 16424, INDONESIA.
yohan.suryanto@ui.ac.id, rizki.reynaldo@ui.ac.id, riri@ui.ac.id
*Corresponding author: yohan.suryanto@ui.ac.id

Abstract: Implementation has a purpose to give adequate guarantee for multimedia application to be able delivered according to the priority and the class of services. But basically, end to end QoS guarantee is very difficult to be realized, especially when it involves a lot of operators with a variety of interconnection networks such as Internet. To overcome that difficulty of implementing end-to-end QoS in IP-based networks, we propose a partial QoS approach through TCP spoofing technique. Partial QoS is implementation of QoS subset like bandwidth parameter in certain ip based network segment. TCP spoofing is a technique to intercept TCP connection between user and content server to be further manipulated. Each TCP connection will be intercepted by spoofing gateway and will be adjusted to the appropriate window size parameter that has been approved for each user. Spoofing Gateway will forward the request from the client to the content server and change the original TCP protocol to TCP Linux Highspeed. Our Simulation results, using NS-2.35, show that for some cases of Partial QoS through spoofing technique produced better performance, such as completion time and average throughput for each user class priority, compared to the end-to-end QoS approach. Even in cases where the Internet network delay product characteristics are relatively high, partial QoS spoofing technique has a much better performance than end-to-end QoS.
Keywords: TCP Spoofing, partial QoS, End to End Approach, Performance Comparison, Wireless Network, Internet, IP based network.

1 Introduction

Rapid growth of Internet users today, bringing a significant impact to the network load including access network. A lot of users compete to get access to the Internet network through ISP. They are willing to pay higher to get a higher class of service. For that, ISPs need to implement QoS [1] parameters, such as maximum bandwidth, which is in accordance with the level of service provided.

Setting different priority level needs QoS implementation for full or partial QoS parameters. [2] It requires appropriate bandwidth settings for the level of service provided by the operator. Unfortunately the implementation of the end-to-end QoS for the Internet network is very difficult, because many networks interconnected without common QoS standard and there is no single operator which controls the end-to-end Internet network.

The technology used, especially in the access network, does not always support the QoS capabilities. For example in the case of WiFi 802.11 which is widely used today as an access network, it does not allow the operator to provide the service class based on adequate bandwidth. Through user requires a priority based on the level of service they have paid to the operator. Although setting QoS in 802.11 networks can also be done by implementing 802.11e standard, but these techniques require modification of the media access layer (MAC layer) of WiFi 802.11. [3]

Priority is generally given to the user in the form of different access speeds for each class of service or total bytes accessed by users each month. The 802.11 WiFi network can not

distinguish between the speed per user because there is no mechanism in layer 2, which could limit the throughput per given time frame. While the Remote Access Server (RAS) which is commonly used as a WiFi gateway, can only limit the bandwidth capacity that is used for a relatively long period of time, such as for a period of a day or a month.

We propose a bandwidth setting at TCP level because it is independent to hardware of the access network being used by the customer. Due to the fact that the TCP connection is end to end from the client to the server, it is not easy to implementing bandwidth control through TCP manipulation. Therefore, we propose the use of spoofing techniques to overcome these problems. TCP connection will be intercepted on the gateway, which is spoofing server, so that a subset of the QoS parameters can be added. This connection is then converted into a highspeed TCP in order to be utilize as quickly as possible the available capacity backbone networks that has been designed using traffic engineering process. [4]

Partial spoofing techniques to enable QoS that we propose, can be applied not only on the wireless network, but also can be applied to the wireline network. This is in contrast to the application of the 802.11e EDCA that can only be applied to the WiFi network by changing the 802.11 MAC layer. [3]

We simulate the application of the partial QoS wired or wireless network by implementing spoofing techniques using NS-2.35, a tool commonly used to simulate TCP performance. [5,6] The performance of partial QoS through spoofing approach will be compared with the performance of end-to-end approach. Although partial QoS can be applied also to the UDP protocol, in this simulation we only apply to TCP protocol. Therefore spoofing technique used is TCP spoofing.

Further in this paper, we will discuss the theoretical background in section II, including the QoS and needs [4], Internet traffic flow model, TCP, Highspeed TCP and throughput analysis. [7] The proposed solution is discussed in section III. Meanwhile the partial QoS simulation scenarios through TCP spoofing approach will be presented in section IV. Performance analysis of partial QoS using TCP spoofing techniques will be presented in section V, and then the conclusion in section VI.

2 Supporting theory

2.1 QoS Overview

In communication using the Internet network, network bandwidth generally has a maximum limit for distributing multimedia content over time. Specific content in order to be delivered according to the priority level required for the implementation of QoS. [8] QoS implementation is possible, because even if the user wants all the applications to arrive at the destination as quickly as possible, however:

1. Not all applications must be delivered within limited time duration.
2. Only certain applications that require limited time.
3. Only certain applications that sensitive to jitter.
4. Not all content is very sensitive to packet lost.

On IP-based networks, QoS services provided by the operator can be measured according to the following performance criteria:

a. Throughput
b. Service availability
c. Packet loss

d. Delay
e. Variation of Delay or jitter

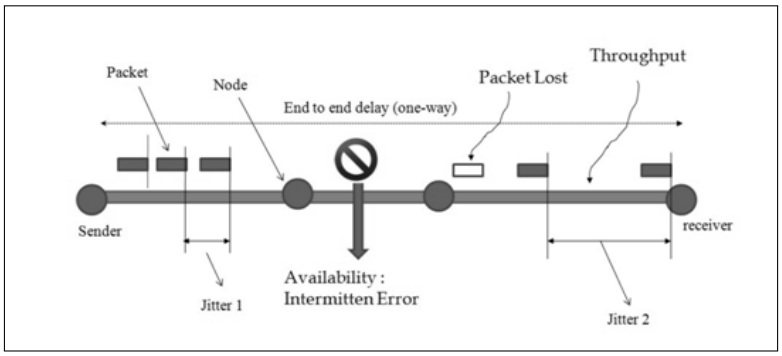


Figure 1: QoS Parameter

Quality of service (QoS) means to guarantee the commitment of those criteria according to the awarded contract. The given contract may be based on something that is explicitly mentioned, such as network access at the speed of 2 Mbps. But it can also be something that is not explicitly presented, such as the multimedia package, 2 Mbps office package, 2 Mbps of easy browsing and so on.

Each group applications such as voice, video streaming and browsing has a different sensitivity to the QoS parameters. Some approaches have been proposed to improve QoS.([1]- [4], [9]- [12]) In practice, users are also difficult to determine how many time in a month they will use the voice applications, video streaming, browsing and download. Therefore in practice, only a subset of QoS parameters that often become criterion, generally only a maximum throughput that can be obtained by the customer when connected to the network.

2.2 Internet Application Traffic Flow Model

Commonly most popular Internet activities are based on client-server traffic model. List of the most popular activity on the internet, can be categorized in the followfng activities:

1. Email : webmail like gmail or yahoo, and also offline email like outlook.
2. Browsing : common website server located at Data Center.
3. Search Engine : google, yahoo, etc
4. Social Netwoking : facebook
5. Video streaming : youtube
6. Audio streaming : mp3 server
7. Audio Communication : skype
8. Chatting : Text message conversation such as YM and whatsapp.

When explored further, such activity has a traffic pattern from the client to the server, not the pattern of the user to the user directly. That means most of the Internet traffic in the network is a communication from the client to the server which usually located in a server farm. Traffic patterns just as hub-spoke or star topology is depicted in Figure 2 below.

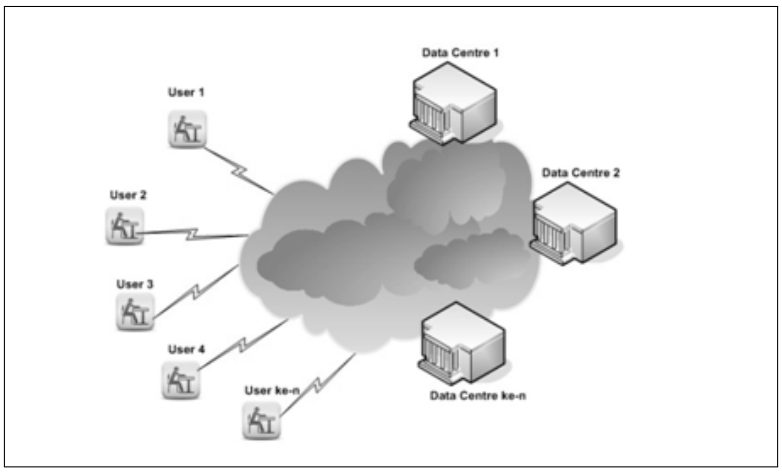


Figure 2: Traffic Flow Client-Server Model

With data flow pattern from the server farm to the customer or vice versa, the traffic flow can optimally be intercepted on the access gateway. Connections from the access gateway to the server farm usually connected through the Internet network backhaul or backbone. The backbone capacity can be adjusted to follow design of the maximum traffic demand from customers to a server farm. With this pattern, partial QoS can be applied in access network.

2.3 Transmission Control Protocol (TCP)

Transmission Control Protocol (TCP) is a connection-oriented protocol which can be relied upon. It provides a transport service for applications that run in a network based on IP protocol.

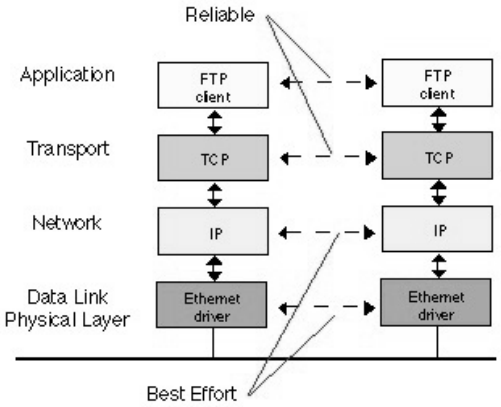


Figure 3: TCP providing reliable data transfer to FTP over an IP network using ethernet

TCP is a stream-oriented protocol, which means that the TCP protocol entities exchange data streams. Every byte of data from the application layer protocol or session layer is placed in the memory buffer and then transmitted by the TCP transport protocol in the data unit. Reliability of service TCP flow control is far more complex than UDP, since UDP only provides

best effort service. To implement such a reliable service, TCP protocol uses a timer to ensure reliability and to synchronize between the two ends of the TCP connection.

For most networks approximately 90% of current traffic uses this transport service. It is used by such applications as telnet, World Wide Web (WWW), FTP, electronic mail. The transport header contains a Service Access Point which indicates the protocol which is being used (23=telnet, 25=mail, 69=TFTP, 80=WWW).

There are many kind of TCP:

- TCP Selective Acknowledgments (SACK)
- TCP Reno
- TeP Vegas
- HSTCP (HighSpeed TCP)
- XCP (eXplicit Control Protocol)

Some paper suggest the way to improve TCP performance ([5], [13]- [15]), in this paper we propose a TCP spoofing technique that uses TCP normal in access network and forwards packets received by the spoofing gateway to the server destination with highspeed connections TCP (HSTCP). [15] HSTCP is an update of TCP that react better when using large windows on high bandwidth, high-latency networks. HSTCP alters how the window is opened each round trip and closed on congestion events as a function of the absolute size of the window. [15]

2.4 Average Throughput and Completion Time

Performance comparison between partial QoS through spoofing techniques and QoS through end-to-end approach is done by looking at the average throughput and completion time for each user using both methods. Completion time or transfer time is the time the last byte is received by the receiver subtracted by start time packet sent by the sender. Meanwhile, in our NS2 simulation, we capture the throughput of Th(i) on each TCPSink node for the unit time of Δt(i). The example graph of Th(i) for each Δt(i) can be seen in Figure 4.

To make it easier to compare the throughput performance, we need the average throughput at time t(i) during the transfer process. This can be obtained using equation (1):

$$Avg\_Th\,(i) = \frac{\Sigma Th\,(i)\,x\Delta t(i)}{\Sigma \Delta t(i)} \tag{1}$$

Where Avg_Th(i) is average throughput from the starting transfer time to t(i). Th(i) is throughput at Δt(i).

for the example For Δt of 0.02 second, the average throughput for each time t(i) based on the equation (1), the average throughput can be depicted in Figure 5:

To determine the average throughput as a whole, from the start transfer time until the transfer is complete, also can use the following equation:

$$Avg\_Th = \frac{Packet\ Size\ x\ 8}{time\ completion} \tag{2}$$

Where Avg_Th is the average throughput since the transmitter starts sending packets until the receiever received the entire packet.

Meanwhile, the maximum capacity of TCP link can be calculated from the equation (3) below:

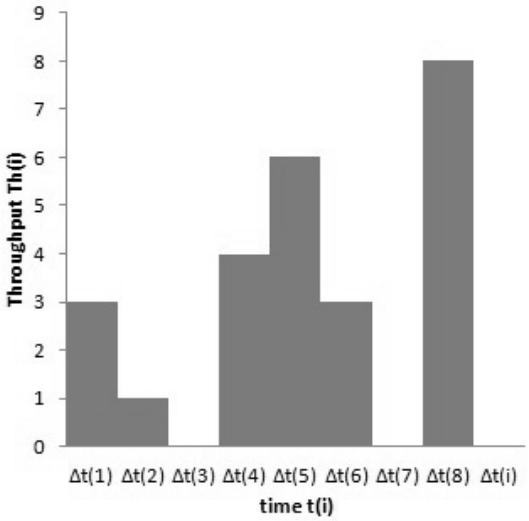


Figure 4: Throughput for each Δt captured at TCPSink

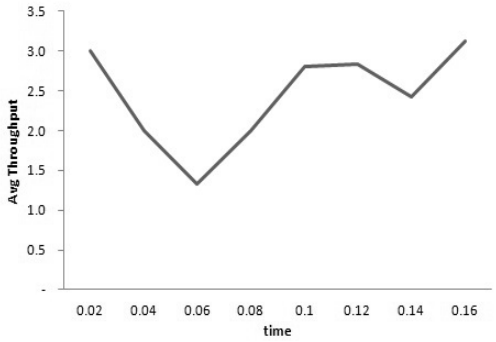


Figure 5: The normalized average Throughput

$$Maxs\_Th = \frac{Window\_Size\,(i)\,x\ 8\ x\ Packet\_Size(i)}{2xDelay(i)} \tag{3}$$

Where the Maxs_Th is the maximum throughput, Window_size(i) is TCP window size at frame i, Packet_size(i) is the TCP packet size at i and Delay(i) is the link delay at frame i.

3 Proposed solution

In this paper we propose the use of TCP spoofing techniques to enable the partial QoS. TCP Spoofing is a method to intercept TCP. TCP connection which has been intercepted can be converted into another TCP protocol type and can also be added to a subset of the QoS parameters. By utilizing the TCP spoofing, the partial QoS can be implemented independent to the type of network used. We can implement the partial QoS on wired or wireless networks.

Partial QoS is an implementation technique of the subset QoS parameters in a particular network segment. In this paper, a partial QoS implementation proposed is done at the network access. With a partial QoS through TCP spoofing approach, access technology currently installed can support the bandwidth control without the need to change the MAC layer. This is different from the approach of the 802.11e solution [3, 12]. Thus the TCP spoofing approach has better flexibility than the implementation of QoS by changing the access network hardware. Partial QoS configuration using spoofing techniques can be seen in Figure 6.

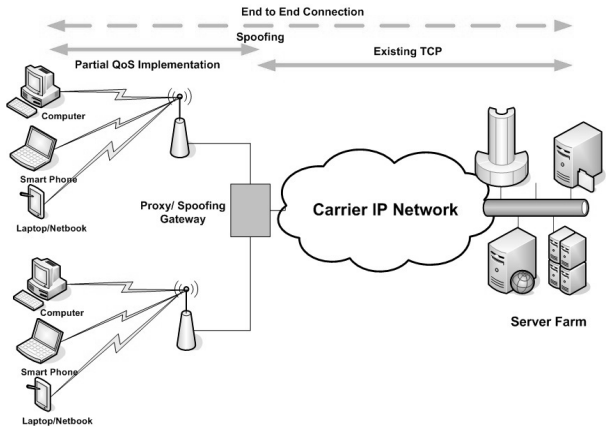


Figure 6: Partial QoS through spoofing technique

Spoofing gateway charged to intercept TCP connection from the user to the content server. In access network side from the user to the gateway spoofing, partial QoS will be applied. This is done by adding parameters window size and congestion window to the corresponding TCP connection. Then by spoofing gateway, packets will be forwarded to the destination server by first changing into highspeed TCP connection. [15]

4 Simulation scenario

This paper analyzed the completion time performance and average throughput between partial QoS using TCP spoofing and end-to-end approach. Simulations performed on wireless and

wireline networks. In the wireline network simulation, we use a dumb bell topology with 3 user configurations, each with priority classes low, medium and high. Node 3 is a spoofing gateway, while node 5, node 6 and node 7 are acted as content server as shown in Figure 7 below.

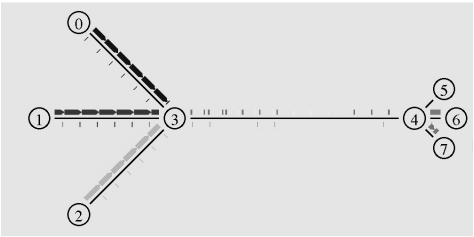


Figure 7: Topology of wireline simulation

For simulations in wireless networks, we used a single 802.11 base station with three users, each configured with priority low, medium and high. Node 1 is a spoofing gateway and node 0 act as a content server, as shown in Figure 8 below.

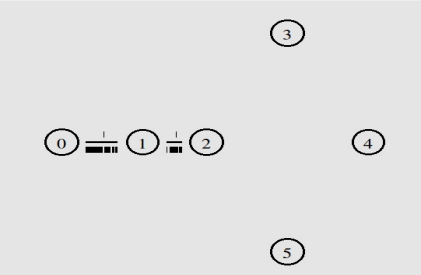


Figure 8: Topology of wireless simulation

Chosen to measure the performance is average throughput and completion time. This is selected because the QoS subset used is bandwidth parameter. Users will transfer the FTP file with file sizes M bytes to the server. Comparative performance begins by comparing the average throughput and the transfer time for spoofing techniques and end-to-end approach for single user without setting the window size. This is done to determine the speedup ratio of spoofing techniques for various total delay conditions. Validation is made by measuring the size of packets received by the server in TCP spoofing approach should be the same as the end-to-end approach.

The next comparison is done for a single user with a window size limitation. This simulation is run to determine the effect of window size restrictions that will be used as a QoS subset. The next simulation is the performance comparison of 3 users without the window size, which aims to measure the average throughput without additional subset of QoS. The next scenario is to make a comparison of simulated single user and multiuser with window size under varied backbone delay conditions. This is done to measure the QoS performance with partial spoofing approach compared with end-to-end approach.

In the NS2 simulations we set a subset of QoS parameters by adjusting the window size using tcl commands:

```
set TCP0 [new Agent/TCP]
$TCP0 set packetSize_ $pktsize
$TCP0 set window_ $window1
$TCP0 set cwnd_ $cwnd1
$ns attach-agent $n0 $TCP0
```

TCP connection from the spoofing gateway to the server is converted into TCP highspeed linux with the command:

```
set TCP4 [new Agent/TCP/Linux]
$ns at 0 "$TCP4 select_ca highspeed"
$ns attach-agent $n3 $TCP4
```

Packets received by the spoofing gateway will be forwarded to the server with the command:

```
set relay1 [new Application/FTP]
$relay1 attach-agent $TCP3
set bw0 [$sink0 set bytes_]
$relay1 producemore [expr $bw0/$pktsize]
```

5 Result analysis

When we do a file transfer from the user to the server through a wireline network with dumb bell topology as shown in Figure 8, we can estimate the maximum transfer speed which can be achieved. The maximum throughput between the user and the server on end-to-end approach is the minimum value of TCP connection capacity and a media capacity of a segment which can be written as:

$$Th\_maxs = Min(MaxsTCP, minB\_segment) \tag{4}$$

Where are:
Th_maxs : Maximum Throughput
Maxs TCP : Maximum capacity of TCP connection according to equation (3)
Min B_segment : Minimum bandwidth of each segment of transmition media between user and server.

Meanwhile, the maximum throughput between the user and the server on TCP spoofing approach is the minimum throughput of the access and backbone section that can be written as:

$$Th\_maxs = Min(Th\_maxs\_access, Th\_maxs\_backbone) \tag{5}$$

Where are:
Th_maxs access : maximum throughput of access network according to equation (4)
Th_maxs backbone : maximum thorughput of backbone network according to equation (4)

The network in TCP spoofing approach if broken down into two segments, which means that each segment delay $\leq$ TCP connection end-to-end delay. Based on equation (3), equation (4) and equation (5), in general it can be concluded that the Th_maxs spoofing $\geq$ Th_maxs end

to end. This can be seen for example in the first simulation, where we compare the average throughput between TCP spoofing approach and end-to-end approach.

For the case of access bandwidth set to 1 Mbps, access delay 10 ms and backbone bandwidth 2 Mbps, 1 Mbps bandwidth server, backbone and server delay 95 ms, we get:

$$\begin{aligned} \text{Th\_maxs end to end} &= \min(0.76 \text{ Mbps}, 1 \text{ Mbps}) \\ &= 0.76 \text{ Mbps} \end{aligned}$$

$$\begin{aligned} \text{Th\_maxs spoofing} &= \min(\min(8 \text{ Mbps}, 1 \text{ Mbps}), \min(0.84 \text{ Mbps}, 1 \text{ Mbps})) \\ &= 0.84 \text{ Mbps} \end{aligned}$$

The result show that average speed df the end-to-end approach $\leq$ Th_maxs end to end, and average speed of spoofing $\leq$ Th_maxs spoofing approach. The comparison result of the average throughput of the two approaches for the case is consistent with the above analysis. This can be seen in Figure 9.

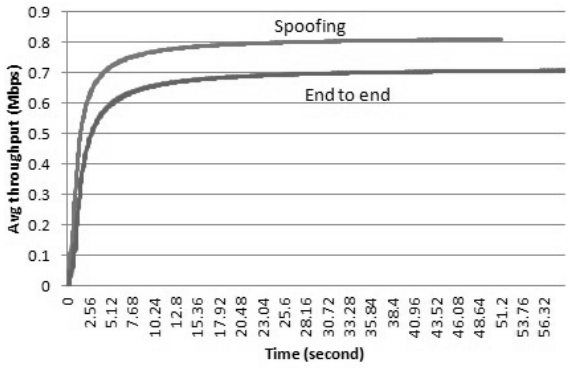


Figure 9: The average throughput comparison between end to end approach and TCP spoofing approach

In the simulation of TCP spoofing approach in wireline network topology as depicted in Figure 7, thg backbone delay is varied, comparative relationshhp between the average throughput and total delay in the end-to-end approach and the spoofing approach can be seen in Figure 10. When the capacity of TCP $\geq$ capacity of a media segment, according to equation (4) and (5), this shows that

the average throughput is limited by the available bandwidth limit of transmission media. The average throughput between TCP spoofing approach and the end-to-end approach is not very different. But when TCP connection capacity $\leq$ segment capacity of transmission media, according to equation (4) and (5), this shows that the performance of spoofing approach is much better than the end-to-end approach.

Figuse 11 confirm that the analysis of the completion time of the end-to-end approach $\geq$ completion time of TCP spoofing approach. The completion time difference beeween the end-to-end approach and TCP spoofing approach for 5 MB FTP packets to tht total delay is clearly according to equation (4) and (5).

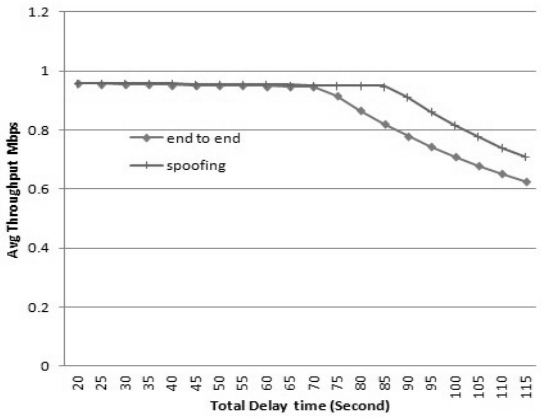


Figure 10: Average throughput to total delay

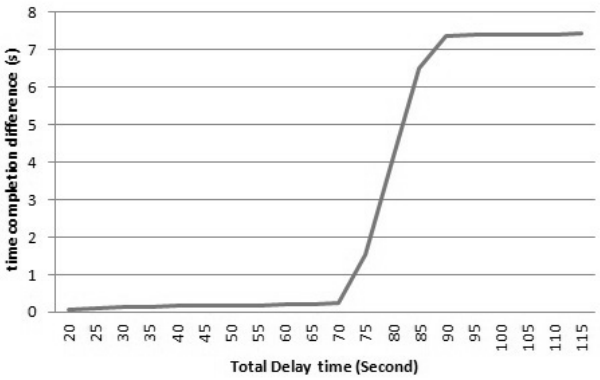


Figure 11: Completion time difference to total delay

As expected, the result shows that the average throughput of the end-to-end approach $\leq$ TCP spoofing approach. The difference in the completion time becomes quite large when TCP capacity $\leq$ available bandwidth of media segment. In this case, the total delay if 70 ms in which a TCP connection capacity approaching media segment bandwidth limitations, the completion time of the TCP spoofing is much faster than the end-to-end approach.

When we choose to run the simulation in condition where TCP connection capacity $<$ bandwidth of media segment, the completion time is more influenced by the capacity of the TCP connections. The comparison of the completion time to the FTP packet sizes, under the condition where bandwidth of media segment is still sufficient such as in the case when we select the delay of backbone and the server 105 ms and the access delay 50 ms. It is clear that completion time by spoofing approach is better than the completion time by the end approaches. This can be seen in Figure 12 below.

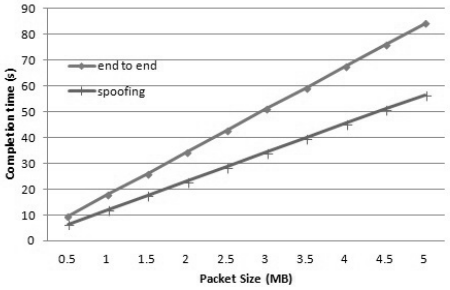


Figure 12: Comparison of completion time to the size of FTP file

In the next simulation, the access and backbone segments bandwidth is set to 10 Mbps so that the media connection does not become a bottleneck, because in a multi-user scenario with a QoS subset we want to show a completion time for each user priority based on the TCP connection capacity to the delay. The implementation of QoS subset for the low user priority is represented by the window size 5, the medium user priority with window size 10 and the high user priority with window size 15. For the condition of delay access 50 ms, delay backbone plus delay server is 55 ms, the average throughput graphs for each user priority using end-to-end approach can be seen in Figure 13.

It appears that the user with lower priority has the lowest average throughput and the user with high priority has the highest throughput. The duration of 5 MB file transfer time for each scenario with multi-user network delay for each priority users with end-to-end approach can be seen in Table 1. We can see that the completion time of user with high priority $<$ completion time of medium user $<$ completion time of low priority user.

The average throughput of each user priority with a media connection of 10 Mbps using TCP spoofing approach can be seen in Figure 14. The average throughput for each priority user using TCP spoofing approach, either for low priority, medium priority and high priority has a similar pattern with the end-to-end approach. However it can be seen from the graph, that the average throughput of TCP spoofing approach is higher than the end-to-end approach.

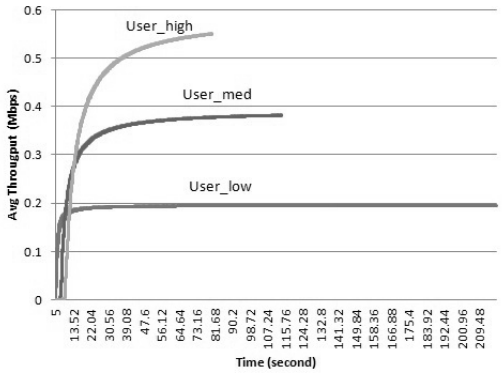


Figure 13: The average throughput for each user priority for end to end approach

Table 1: Completion time each User Priority for the End to End Approach

No	Description	Delay		completion time end to end		
		Access	Backbone +server	user 1	user 2	user 3
1	Scenario 1	10	15	52.68	26.46	17.72
2	Scenario 2	10	55	132.8	66.64	44.64
3	Scenario 3	50	55	212.92	106.84	71.56

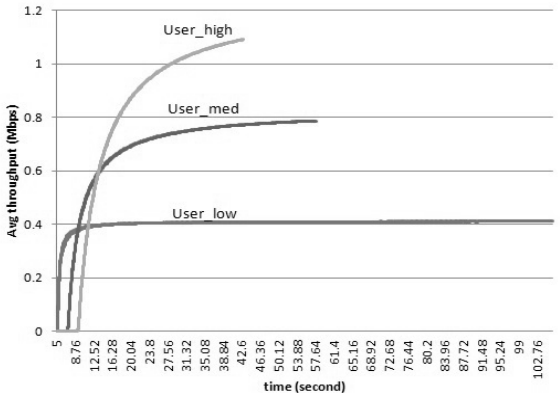


Figure 14: The average throughput for each user priority for TCP spoofing approach

The maximum throughput of spoofing approach can be obtained from equation (5), where in the access network using TCP with window size parameter and in backbone side using TCP highspeed. In the access side, for access delay of 10 ms, based on equation (5) we obtained that:

Th_maxs access side user low = min (2 Mbps, 10 Mbps)
Th_maxs access side user med= min (4 Mbps, 10 Mbps)
Th_maxs access side user high= min (6 Mhps, 10 Mbps)

Meanwhile, for the backbone delay of 10 ms and the server delay of 5 ms, where highspeed TCP protocol with window size 20 as the default is used, we obtained:

Th_maxs backbone = min (5.33 Mbps, 10 Mbps)

It can be seen from Table 2, that for the backbone delay of 10 ms and access delay of 10 ms, the transfer rate is determined by the settings window size of access side. This is because the backbone throughput > access throughput for each user priority. When the backbone and server delay of 55 ms, Th_maxs backbone is 1.45 Mbps. Under these conditions, the backbone becomes a bottleneck so that the completion time is determined by the speed of transfer over the backbone side. As soon as the access delay changed to 50 ms, which means Th_maxs access side of the low priority user = 0.4 Mbps, Th_maxs user side access med = 0.8 Mbps and the user side access Th_maxs high = 1.2 Mbps, the speed of the transfer process is more determined by the network access. Further completion time for all three spoofing scenarios can be seen in Table 2.

Table 2: Completion Time Each User Priority for The Spoofing Approach

No	Description	Delay		completion spoofing-qos		
		Access	Backbone +server	user 1	user 2	user 3
1	Scenario 1	10	15	20.92	10.54	10.08
2	Scenario 2	10	55	28.28	28.28	28.26
3	Scenario 3	50	55	101.08	50.78	34.04

Overall, based on Table 1 and Table 2, the completion time of each level of user priority using spoofing approach is faster than the completion time of end-to-end approach. The comparison of the completion time between the spoofing approach and the end-to-end approach, i.e. for high priority users, to the total delay can be seen in Figure 15.

The simulation of the partial QoS implementation in wireless networks is done using a topology as seen in Figure 9, with 54 Mbps wireless bandwidth share, access delay of about 10 ms, backbone bandwidth of 10 Mbps and backbone delay 30 ms. The average throughput of each user priority with end-to-end approach can be seen in Figure 16 below.

Using the same wireless network topology, partial QoS implementation through TCP spoofing approach for each class of user priority, can be seen in Figure 18. The interesting result for the user with the lowest priority is the average throughput flat when there is a user with a higher priority. But when users with higher priority reaches its optimum speed, the average user with low priority throughput also rise if the bandwidth is still available. Thus the overall average throughput of partial QoS through spoofing approach has a higher average throughput compared with end-to-end approach.

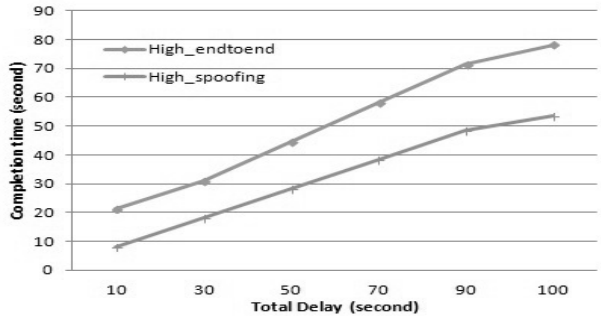


Figure 15: Comparison of Completion time for high priority user to total delay

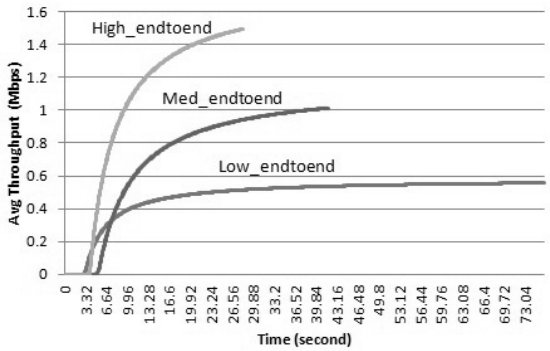


Figure 16: The average throughput of each priority user for end to end approach on wireless network

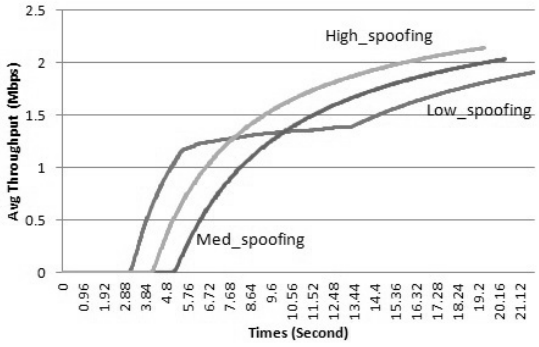


Figure 17: Average throughput each priority user for spoofing approach on wireless network

To get an idea of how much faster the FTP transfer 5 MB in wireless networks for all priority users by spoofing approach compared with end-to-end approach, we created wireless simulation network with two scenarios. Scenario 1 is the backbone delay of 10 ms and scenario 2 with the backbone delay of 30 ms delay. In Table 3 it appears that, for the both scenario, spoofing approach is faster than end-to-end approach to factor > 1. Speed up user with lower priority is higher than speed up user with a higher priority. It occurs because the forwarding mechanism by highspeed TCP has higher ratio for lower throughput of the access side. Speed up can reach a significant number in certain cases. For example the speed up of low priority user for scenario 2 achieves 3.49 times.

Table 3: Speedup of The Spoofing and The End to End Approach on Wireless Network

No	User Priority	Completion Time				Speed Up	
		End to End		Spoofing			
		Scr 1	Scr 2	Scr 1	Scr 2	Scr 1	Scr 2
1	Low	36.26	76.16	20.56	21.8	1.76	3.49
2	Medium	19.88	39.82	14.6	18.46	1.36	2.16
3	High	11.26	24.34	9.34	15.5	1.21	1.57

6 Conclusion

In general the average throughput of TCP spoofing approach is higher than the average throughput of end-to-end approach both in wireline network and the wireless network. This is in accordance with equation (4) and (5) of section analysis. TCP spoofing approach allows the implementation of a subset of the QoS parameters applied partially in access networks. Thus TCP spoofing approach is more flexible than the end-to-end approach to enable partial QoS. TCP spoofing approach is also more flexible when compared to QoS implementations that require MAC layer changes such as 802.11e EDCA solution on WiFi network. Simulation results show that the TCP spoofing techniques allow implementation of partial QoS in the access network with better performance under to various delay conditions compared to the implementation of end-to-end QoS. The completion time speedup of the spoofing approach varies which value $\geq$ 1 compared with end-to-end approach. For certain cases, completion time speed up value is significant, more than 3 times. However, in the spoofing approach we need to consider the buffer availability of the spoofing gateway. This is especially the case where the maximum throughput of access network side much higher than the throughput of the backbone side. In this condition, the spoofing gateway needs to buffer the received data from the access side because the forwarding speed is less then the input rate from access network.

Bibliography

[1] A. Tuoriniemi, G. Eriksson, N. Karlsson, and A. Mahkonen (2002); QoS concepts for ip-based wireless systems, *3G Mobile Communication Technologies, 2002. Third Int. Conf. on (Conf. Publ. No. 489)*, 229-233.

[2] G. Rong and G. Lei (2010); A Partial Bandwidth Reservation Scheme for QoS Support in Ad Hoc Networks, *Computer Engineering and Applications (ICCEA), 2010 Second International Conference on*, 380-384.

[3] L. Jae Young, L. Hwang Soo, and M. Joong Soo (2009); Model-based QoS parameter control for IEEE 802.11e EDCA, *Communications, IEEE Trans. on*, 57: 1914-1918.

[4] D. Miras (2002); Network QoS needs of advanced internet applications: A survey, *Internet2 QoS Working Group.*

[5] D. X. Wei and P. Cao (2006); NS-2 TCP-Linux: an NS-2 TCP implementation with congestion control algorithms from Linux, Proc. from the 2006 workshop on ns-2: the IP network simulator, Pisa, Italy, 2006.

[6] S. Xu and Y. Yang (2011); Protocols simulation and performance analysis in wireless network based on NS2, *Multimedia Technology (ICMT), Int. Conference on*, 638-641.

[7] S. Floyd (2003); HighSpeed TCP for Large Congestion Windows, IETF, Network Working Group RFC 3649, December 2003.

[8] J. W. Evans and C. Filsfils (2010); *Deploying IP and MPLS QoS for Multiservice Networks: Theory & Practice*: Morgan Kaufmann.

[9] K. L. Dias, D. F. Sadok, and J. Kelner (2003); A hierarchical resource management approach for QoS provisioning in IP-based wireless and mobile networks, *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, 3498-3502.

[10] S. H. Bouk, I. Sasase, S. H. Ahmed, and N. Javaid (2012); Gateway discovery algorithm based on multiple QoS path parameters between mobile node and gateway node, *Communications and Networks, Journal of*, 14: 434-442.

[11] V. B. Iversen (2012); Systems with selective overflow and change of bandwidth, *Communications in China (ICCC), 2012 1st IEEE Int. Conference on*, 694-697.

[12] R. Verschae, M. Koeppen, and K. Yoshida (2009); User Modeling for Improving QoS Using Partial User-Supplied Information," in *Intelligent Networking and Collaborative Systems, 2009. INCOS '09. International Conference on*, 387-390.

[13] P. Kulkarni, M. Sooriyabandara, and L. Li (2009); Improving TCP performance in wireless networks by classifying causes of packet losses, *Wireless Communications and Networking Conference, WCNC 2009. IEEE*, 1-6.

[14] Y. Li, R. Zhang, Z. Liu, and Q. Liu (2009); An improved TCP congestion control algorithm over mixed wired/wireless networks, *Broadband Network & Multimedia Technology, IC-BNMT '09. 2nd IEEE Int. Conf. on*, 786-790.

[15] D. Lopez-Pacheco and C. Pham (2004); Performance comparison of TCP, HSTCP and XCP in high-speed, highly variable-bandwidth environments, *Proc. of the IEEE 3rd Int. Conf. on Network Protocols (ICNP 2004). Berlin*, 2004.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):420-427, June, 2015.

Engineering Human Stigmergy

I. Susnea

Ioan Susnea
Dunarea de Jos University of Galati
Romania, 800008 Galati, Domneasca, 47
ioan.susnea@ugal.ro

Abstract:
Discovered in the context of a research about insects, stigmergy – the indirect coordination mechanism that allows ant colonies to achieve intelligent behavior – has been extensively studied with the aim to create artificial, ant-like agents. Although stigmergic behavior has been also identified in human collectivities, there are relatively few reports about technological solutions that facilitate the emergence of such interactions between people. This paper proposes the concept of virtual pheromones, defined as engrams created by the agents not in the environment, but in a representation thereof – a map, and outlines several use cases, wherein pheromones embedded in maps are the key element for inducing stigmergic behavior in human multi-agent systems. Without proposing a theoretic generalization, this paper aims to emphasize the broad range of possible technological applications of human stigmergy, and, maybe, to mark a new starting point for a more in-depth study of this topic.
Keywords: Human stigmergy, virtual pheromones, pheromone maps, intelligent transportation systems, stigmergic shopping, stigmergic learning.

1 Introduction

Stigmergy is a process of indirect coordination in multi-agent systems, by means of traces that agents create in the environment. Upon sensing these traces, other agents are stimulated to perform similar actions, thus reinforcing the traces, in a self-catalytic process.

As a result,simple, local and unplanned actions of the agents emerge in a complex, and apparently intelligent behavior of the system as a whole.

Discovered in the context of a biological research [6], the stigmergy in biological systems relies on the capacity of the agents to deploy in the environment small amounts of chemicals, called pheromones [8].

Pheromones diffuse in space, so that a local source can be detected from within a certain range, and at the same time they evaporate, which makes unreinforced traces to decay and eventually disappear.

A typical example of stigmergic interaction is the ant foraging behavior. When a searcher ant finds a food source, it starts leaving a pheromone trail on its way back to the nest. Other ants tend to follow the path created by the searcher, reinforcing the initial trail. Due to evaporation, longer trails that require more time for the ants to walk along become less attractive and, in the absence of reinforcement, they eventually disappear. In the end, the majority of ants will choose the shortest route between the food source and the nest.

This processof finding the optimal route was called "ant colony optimization" (ACO – [4]), and attracted a great deal of interest for the emergent intelligent behavior in swarms.

Numerous researchers have studied possible uses of stigmergy in robotics [19], for military applications [12], for routing data packets in computer networks [18], web mining [1], mobile sensor networks [7], or in cognitive sciences [9].

A variety of solutions have been proposed for the implementation of the artificial pheromones, including automatic deployment of chemicals in the environment [16], storing special data structures in RFID tags [19], exchanging short range messages between agents [15], or by creating dedicated software entities [10]. All the implementations of stigmergic systems with artificial agents are based on a set of concepts and principles that have been formulated in [13]. According to Parunak, a multi-agent system (MAS) is a three-tuple:

MAS = {Agents, Environment, Coupling} ,

wherein the agents are characterized by a *set of states*, a number of *inputs and outputs*, and *a program* that governs the transitions between internal states. The agent's program runs autonomously (i.e. without being invoked by an external entity).

The environment also has a *set of states* and a "*dynamics*" (program) that dictates the transitions between states. By means of their inputs and outputs, the agents interact with the environment by spreading pheromones, in certain conditions, according to their own program (*coupling*).

In this approach, a MAS presents the following features:

- agents are simple, identical entities;

- the environment is seen as a shared memory for all the agents;

- agents communicate with each other indirectly, by creating and sensing changes in the state of the environment.

Obviously, this model of interaction, which describes well multi-agent systems with simple, "lightweight" agents [13] is not suitable to describe interactions between so-called BDI agents (BDI stands for Beliefs, Desires, Intentions [17]). The decision making, and the final behavior of such agents are governed by mechanisms far more complex than the simple, reactive response to environmental stimuli.

However, there are many situations when human behavior is stigmergic [5], and it seems that the technological progress, which facilitates indirect communication between people, is likely to create more opportunities for human stigmergy. Notable examples in this direction are: the market system, intelligent transportation systems, web page ranking (Google), recommender systems (e.g. Amazon ranking, IMDB rating) online auctions (eBay), and many others.

The present paper aims to facilitate the understanding of human stigmergy by presenting several examples of technological solutions to induce stigmergic behavior in human collectivities. Without having the intention to propose theoretical generalizations, this may be the starting point for other interesting research in this direction.

Beyond this introduction, this work is structured as follows:

Section 2 discusses the concept of "virtual pheromones", and proposes a simple model for it.

Section 3 presents several use cases wherein virtual pheromones are used to create stigmergic interactions between human agents, and Section 4 is reserved for discussion and conclusions.

2 Virtual Pheromones as Pheromones Embedded in Maps

In an application of robotics [20], we proposed a solution for the implementation of artificial pheromones, wherein the agents leave traces not in the environment, but in a representation thereof – a map. In this approach, a number of agents, equipped with their own localization system and wireless communication, periodically communicate with a "pheromone server". The "pheromone server" is a computer, running a special application software that creates and maintains a data structure called "pheromone map" (see figure 1). Basically, the pheromone map is a 2D grid array, wherein each cell is associated with the following data:

- the coordinates of the corresponding geographic space,

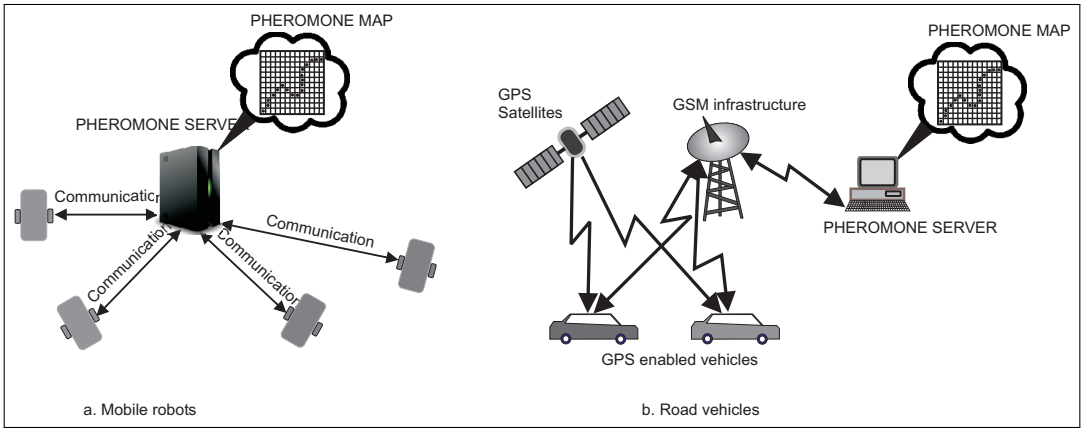


Figure 1: An illustration of the implementation of virtual pheromones

- an integer representing the resultant pheromone intensity of all the pheromone sources located in the respective cell

Periodically, each agent sends queries to the pheromone server, and includes in the query packet position information, as reported by its own localization system. Upon reception of a query packet, the server locates the agent on the internal map, and computes for the respective position the resultant pheromone intensity, by means of a potential field model. With the notations in figure 2:

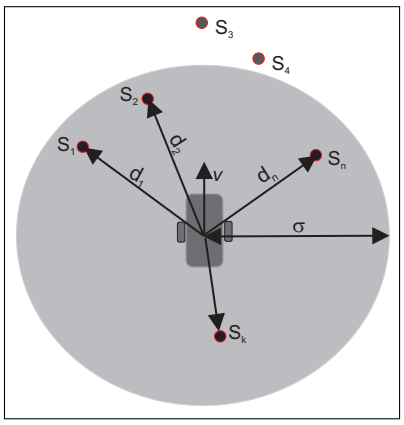


Figure 2: Notations used to describe virtual pheromones aggregation and diffusion

$$p(x) = \begin{cases} p_k\left(1 - \frac{x}{\sigma}\right) & 0 < x < \sigma \\ 0 & x >= \sigma \end{cases} \tag{1}$$

where $p(x)$ is the intensity of the pheromones source S_k, sensed at the distance d_k, due to diffusion. Due to the superposition of the effects (aggregation) of all N pheromone sources located within the sensitivity range σ, the resulting pheromone intensity, sensed in an arbitrary location is:

$$P_R = \sum_{k=1}^{N} p_k \left(1 - \frac{d_k}{\sigma}\right) \tag{2}$$

and, assuming that the evaporation produces a linear decrease of the pheromone intensity, it is possible to write:

$$P_R = \sum_{k=1}^{N} p_k \left(1 - \frac{d_k}{\sigma}\right)\left(1 - \frac{t - t_k}{\tau}\right) \tag{3}$$

where t_k is the time of creation for the source S_k, and τ is an evaporation constant.

This simple, linear model was selected in order to reduce the computational load of the server, because the pheromone intensity (3) must be computed repeatedly, at regular time intervals, for every cell of the grid map, and this can be cumbersome in applications operating with very large maps. The value P_R, computed by the server is returned to the querying agent in the response packet, and thereafter the agent acts as if it had its own pheromone sensing system.

3 Applications of Virtual Pheromones

The concept of virtual pheromones as described in Section 3 of this article was initially developed in a research aiming to reduce cost of some service robots. By manipulating the pheromone maps it is possible to define easy-to-follow robot paths, by means of virtual pheromone gradients, as shown in figure 3.

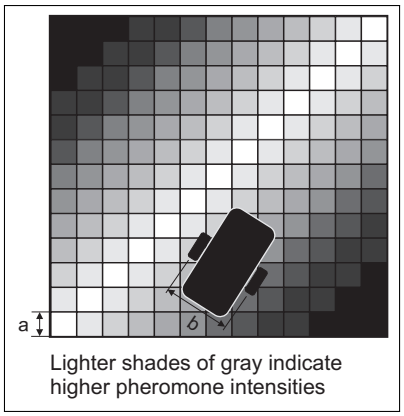


Figure 3: Defining Robot paths by means of pheromone gradients

The idea was later used in an experiment for creating mixed robot formations with physical and virtual agents [21]. In this experiment, a real robot was programmed to follow a software simulated leader robot.

Though interesting, the experiments described in [20] and [21] are limited to controlling individual agents, and do not make use of the capability of pheromones to mediate stigmergic interactions. The following use cases focus on human stigmergy.

3.1 Use case 1: Real-Time Ant Colony Optimization for Road vehicles

Common GPS navigation systems for vehicles comprise a GPS locator, which provides the information about the current position to a microcontroller that extracts from a local memory a predefined map of the geographic area, corresponding to the actual location of the vehicle. This map is displayed on a local screen, and serves for orientation.

The main drawback of this solution derives from the fact that it uses static maps. No matter how often these maps are updated, they still do not include information about how fluent is the traffic at present time on a particular road segment, or whether there is a traffic congestion or a traffic jam.

There are numerous attempts to overcome this limitation, (e.g. [14]) mainly based on providing an additional communication channel between the on-vehicle navigation assistant and an external device, which can be an Internet server, or a roadside equipment designed to store an updated knowledge base about traffic conditions, actual practicability of the roads, etc.

Some web mapping services (e.g. Google maps) include –for limited geographical areas - color coded traffic information in maps, but these solutions usually offer just an estimate of the traffic conditions, based on previously collected data, or they rely on measuring the average speed of individual vehicles by means of GPS data collected via smartphones.

Even so, the problem of updating the additional knowledge base still remains open, and can only be solved with substantial operation costs.

A possible implementation of an improved GPS navigation system for vehicles based on stigmergy would have the general structure presented in figure 1b. A fragment of the pheromone map, is transmitted by the server in response to queries sent by individual agents and superposed over the existing GPS navigation maps as a transparent layer. Since the pheromone map is created and maintained by the agents themselves, no additional operation costs are required for this system.

3.2 Use case 2: In-Store Recommender Systems

Most department stores and supermarkets use powerful inventory management software, which uses data from the POS equipment to keep a detailed database of the goods offered for sale. Every item in this database has an unique barcode identifier, which is associated with substantial additional information (price, lot number, expiry date etc.) including the geographic position on the shelf where the respective item is located.

Having this information, it is relatively easy to set up a network like the one presented in figure 4, wherein a pheromone server maintains a map of the store and updates the pheromone distribution every time the POS reports the sale of an item. The resulting map is displayed for public on TV screens deployed throughout the store.

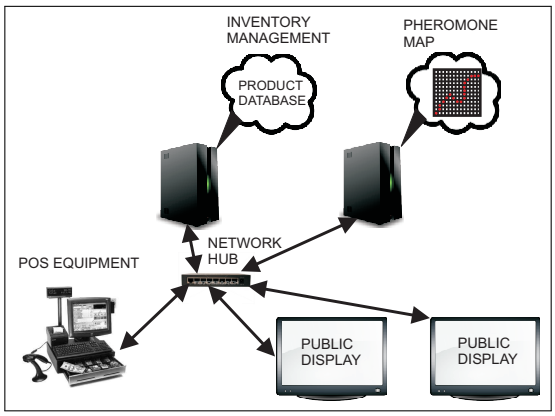


Figure 4: The structure of a stigmergic in-store recommender system

The shelves with frequently sold products are displayed on the map as “hot spots”, thus creating on the customers an effect similar – to a certain degree – with the Amazon sales rank.

The underlying idea of “value” associated with a high sales rank is likely to produce a stigmergic effect on the shopper’s behavior: some shoppers will visit the places indicated as holding highly ranked products and eventually will buy those products, contributing to a further increase of their sales (at least, this happens in the case of e-commerce - see amazon.com)

3.3 Use case 3: Self-Organization of the Educational Objects

In certain conditions, the emergence does not affect the agent's behavior, but the environment shared by the agents. Consider, for example, a database of "educational objects" [23] used by a (large) community of students, or researchers. Each object is described by a set of metadata, including:

- an unique identifier,
- a set of tags (keywords) describing the educational content of the object,
- a means to detect that the object has been accessed by some user,
- a means to store virtual pheromone information

Every time a user accesses an object in the database, the pheromone intensity associated with that object increases. Besides that, the pheromone intensity of all objects that share the same keywords with the object directly accessed is also increased, but with a lesser ratio (diffusion).

On the other hand, the pheromone intensity of all objects in the database is automatically decreased by software at regular time intervals, so that, if an object is not accessed for a long time, its pheromone intensity eventually drops to zero (through evaporation).

This simple scheme, automatically creates some metadata describing the "educational value" of the objects in the database, so that any user can immediately select educational objects (filtered by tags) with the highest value.

Though very simple, this approach, combined with a P2P environment, may be the starting point for a serious attempt to overcome the "teacher-student" learning paradigm, and replace it with a new "teach while you learn" paradigm. Further research in this direction might be beneficial.

4 Discussion and Future Work

There are several aspects of the concept of virtual pheromones, as defined in this paper, worth to discuss:

a. Built this way, the pheromone maps can be seen as a special type of cognitive maps [24], having a very interesting feature: they create a link between the behavior implicit communication messages, generated by the agents [11] and certain geographical places.

If somebody buys a product, chooses to watch an item in an exhibition, or visits a certain web page, he involuntarily sends a message carrying information about a personal decision. Fusing the information from a plurality of such messages in a unique map, and making this map available for all the agents is the key element that can trigger stigmergic interactions.

The idea that there is a connection between what we call swarm intelligence and the concept of cognitive maps has been suggested back in 1995 by Chialvo and Millonas in [2].

A closer look at the examples outlined in the above paragraphs (see also [22]), suggests that the concept of virtual pheromones as defined here offers the means to create such cognitive maps in just a few simple steps:

- Automatic identification of Behavior Implicit Communication (BIC) signals. It is worth to note that the problem of automatic detection of certain behaviors has been extensively studied under the name activity recognition in various contexts (e.g. human - robot interaction, ambient intelligence, pervasive healthcare etc.), but there are no signs of capitalization of the results of the research on activity recognition from the perspective of capturing BIC messages;

- Localization - Linking the BIC messages to specific places within the environment. This can be done either by enabling agents with a certain degree of location awareness, as in the above examples, or by creating a number of place agents [3] (i.e. active places in the environment), capable to recognize and keep record of certain behaviors of the agents;

- Creating a dynamic cognitive map of the environment based on the place-activity information provided by the virtual pheromones;

- Sharing the resulting map (or parts of it) with all agents.

b. The model of the virtual pheromones described by (3) does not require the absolute position of the agents to compute the pheromone intensity associated with a certain place. It only needs the relative distances between the specific place considered and the pheromone sources. This means that the proposed model can be used - in principle - in any metric space, whenever we can define a distance function. This leaves room for some interesting applications, where the intelligent agents operate in more abstract spaces, like databases, or social networks (see also [23]).

c. Designing systems that treat humans as simple "agents", certainly may raise some ethical issues. For example, a stigmergic evacuation system designed to save lives in case of fire, involves RFID badges and readers deployed throughout the building, which can easily be used to monitor the movements of individuals. And this is not always ethical, and sometimes not even legal.

To conclude, we will note that, although stigmergy was intensely studied over the past decades, the above mentioned use cases suggest that there is still plenty of room for interesting research and applications, mainly in what concerns the coordination mechanisms between complex, intelligent agents. In this context, the centralized approach on implementing virtual pheromones, presented here, may be useful.

Bibliography

[1] Abraham, A., & Ramos, V. (2003). Web usage mining using artificial ant colony clustering and linear genetic programming. In E*volutionary Computation, 2003. CEC'03. The 2003 Congress on (Vol. 2, pp. 1384-1391). IEEE.*

[2] Chialvo, D. R., Millonas, M. M. (1995). How swarms build cognitive maps. In *The biology and technology of intelligent autonomous agents*, 439-450.

[3] Conradt, J. A. (2008). A distributed cognitive map for spatial navigation based on graphically organized place agents *Doctoral dissertation, Swiss Federal Institute of Technology, Zurich.*

[4] Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: optimization by a colony of cooperating agents. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 26(1): 29-41.

[5] Doyle, M. J., Marsh, L. (2013). Stigmergy 3.0: From ants to economies. *Cognitive Systems Research*, 21: 1-6.

[6] Grasse, P. P. (1959) La Reconstruction du nid et les Coordinations Inter-Individuelles chez Bellicositermes Natalensis et Cubitermes sp: La theorie de la Stigmergie: Essai d'interpretation du Comportement des Termites Constructeurs. *Insectes Sociaux*, 6: 41-82.

[7] Howard, A., Matarić, M. J., & Sukhatme, G. S. (2002). An incremental self-deployment algorithm for mobile sensor networks. A*utonomous Robots*, 13(2): 113-126.

[8] Karlson, P., Luscher, M. (1959); Pheromones: a new term for a class of biologically active substances. *Nature*, 183: 55-56

[9] Marsh, L., & Onof, C. (2008). Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research*, 9(1):136-149.

[10] Negulescu, S. C., Dzitac, I., Lascu, A. E. (2010). Synthetic genes for artificial ants. Diversity in ant colony optimization algorithms. International Journal of Computers Communications & Control, 5(2): 216-223.

[11] Omicini, A., Ricci, A., Viroli, M., Castelfranchi, C., Tummolini, L. "A conceptual framework for self-organising MAS". In: *AI*IA/TABOO Joint Workshop Dagli oggetti agli agenti: sistemi complessi e agenti razionali (WOA 2004), Turin, Italy, Pitagora Editrice Bologna* (30 November- 1 December 2004), 100-109.

[12] Parunak, H. V. D., Brueckner, S., Sauter, J., & Posdamer, J. (2001). Mechanisms and military applications for synthetic pheromones. *Ann Arbor*, 1001: 48113-4001.

[13] Parunak, H. V. D. (2006). A survey of environments and mechanisms for human-human stigmergy. In *Environments for Multi-Agent Systems II*, Springer Berlin Heidelberg

[14] Patent application US2006/0094443

[15] Payton, D. W., Daily, M. J., Hoff, B., Howard, M. D., & Lee, C. L. (2001, March). Pheromone robotics. In *Intelligent Systems and Smart Manufacturing* , International Society for Optics and Photonics, 67-75.

[16] Purnamadjaja, A. H., & Russell, R. A. (2007). Guiding robots' behaviors using pheromone communication. *Autonomous Robots*, 23(2): 113-130.

[17] Rao, A. S., Georgeff, M. P. (1995). BDI agents: From theory to practice. *ICMAS*, 95: 312-319.

[18] Sim, K. M., & Sun, W. H. (2003); Ant colony optimization for routing and load-balancing: survey and new directions. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 33(5):560-572.

[19] Susnea, I., Vasiliu, G., & Filipescu, A. (2008). RFID digital pheromones for generating stigmergic behaviour to autonomous mobile robots. In N. E. Mastorakis, et al. (Eds.), *International Conference. Proceedings. Mathematics and Computers in Science and Engineering (No. 4).*

[20] Susnea, I., Vasiliu, G., Filipescu, A., & Radaschin, A. (2009). Virtual pheromones for real-time control of autonomous mobile robots. *Studies in Informatics and Control*, 18(3): 233-240.

[21] Susnea I. Vasiliu G (2011); A framework for creating mixed robot formations with physical and virtual agents *Proc. of the International Conference of the Air Force Academy "Henri Coanda", AFASES2011*

[22] Susnea I. (2012). Applications of the Emergence in Cognitive MAS, *The Annals of the University Dunarea de Jos of Galati, Fascicle III, 2012*, 35(2): 41-48.

[23] Susnea, I., Vasiliu, G., & Mitu, D. E. (2013). Enabling Self-Organization of the Educational Content in Ad Hoc Learning Networks, *Studies in Informatics and Control*, 22(2): 143-152.

[24] Tolman, E.C. (1948). Cognitive Maps in Rats and Men, *The Psychological Review*, 55(4): 189-208.

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):428-440, June, 2015.

Group Pattern Mining Algorithm of Moving Objects' Uncertain Trajectories

S. Wang, L. Wu, F. Zhou, C. Zheng, H. Wang

Shuang Wang, Lina Wu, Fuchai Zhou, Cuicui Zheng
Software College
Northeastern University
Shenyang, China
wangsh@mail.neu.edu.cn

Haibo Wang
H. John Heinz III College
Carnegie Mellon University
Pittsburgh, USA
haibowang@cmu.edu

Abstract: Uncertain is inherent in moving object trajectories due to measurement errors or time-discretized sampling. Unfortunately, most previous research on trajectory pattern mining did not consider the uncertainty of trajectory data. This paper focuses on the uncertain group pattern mining, which is to find the moving objects that travel together. A novel concept, uncertain group pattern, is proposed, and then a two-step approach is introduced to deal with it. In the first step, the uncertain objects' similarities are computed according to their expected distances at each timestamp, and then the objects are clustered according to their spatial proximity. In the second step, a new algorithm to efficiently mining the uncertain group patterns is designed which captures the moving objects that move within the same clusters for certain timestamps that are possibly nonconsecutive. However the search space of group pattern is huge. In order to improve the mining efficiency, some pruning strategies are proposed to greatly reduce the search space. Finally, the effectiveness of the proposed concepts and the efficiency of the approaches are validated by extensive experiments based on both real and synthetic trajectory datasets.
Keywords: probabilistic frequent group pattern, uncertain data, trajectory pattern mining, moving object.

1 Introduction

In recent years, with the development of various location-aware devices, such as RFID tags, cell phones, GPS navigation systems, and point of sale terminals, trajectory data has become ubiquitous in various domains [1]. Such data provides the opportunity of discovering usable knowledge about movement behavior, which fosters ranges of novel applications in human mobility understanding [2], smart transportation, urban planning [3], biological studies [4], environmental and sustain ability studies [5]. Ideally, a trajectory is represented as a sequence of locations, and each is being associated with a corresponding time stamp. However, this is a simplification that does not take into account the inherent uncertainties in such trajectories. For example, a position reported in a GPS signal usually implies a location point with an error range rather than an exact location. Moreover, as location privacy has become increasingly a concern, many locations are blurred when they are made public. Though the importance of mining uncertain data has been recognized [6–8], to the best of our knowledge there is little work in the trajectory pattern mining literature that studies its effect in the knowledge discovery process.

A useful data analysis task in movement is to find a group of moving objects that are traveling together for a certain time period. This concept, what we refer to as group patterns, can

be defined in both spatial and temporal dimensions: (1) a group of moving objects should be geometrically close to each other; (2) they should be together for at least some minimum time duration. Although mining group patterns has been extensively studied on certain trajectory database, such as flock [9, 10], convoy [11, 12], swarm [1], traveling companion [13, 14], and gathering [15], none of these work deal with the inherent uncertainty in trajectory database.

In this paper, we consider the problem of mining group patterns in the context of uncertain trajectory database. In contrast to previous work that did not model the trajectory with uncertainty, we represent a trajectory of a moving object as a sequence of reported locations at corresponding time stamps and a probability distribution function. The probability distribution function represents the probability distribution of possible locations of the trajectory at a given time instant. In particular, the probability distribution function can be discrete or continuous.

However, discovering the group patterns from uncertain trajectories is not an easy task because of adding a new dimension-probability. The challenges are two-fold: (1) The first challenge is how to design an appropriate distance function to measure the (dis)similarity between two uncertain trajectories. Particularly, an effective similarity metric should be able to conduct measurements in terms of different probability distributions, taking into account spatial distances, temporal intervals, as well as relevant probabilities. (2) Another challenge is how to find group patterns efficiently. Due to adding the probability in uncertain trajectory database, judging frequent group patterns requires to compute the frequent probabilities, and this is not a trivial task. Additionally, group patterns mining solutions may lead to an exponential number of results due to the downward closure property. So it is important to propose some effective pruning rules to reduce the massive search space and the number of probability computations.

Given the afore mentioned challenges, existing works for group patterns mining on exact trajectory data do not tackle the data uncertainty problem well. For instance, due to adding the probability, the concept and similarity distance metric function are not suitable for the uncertain environment. And also the traditional mining algorithms cannot be used directly to solve the uncertain group patterns mining. Our research makes the following contributions: (1) We propose a new problem, mining group patterns over uncertain trajectory data. (2) We introduce a novel and adaptive metric to measure the dissimilarity between two uncertain trajectories. (3) We design an efficient algorithm to discover all uncertain group patterns. In addition, we propose several pruning techniques to reduce search space and avoid redundant computation. (4) Extensive experiments demonstrate that the effectiveness and efficiency of our algorithm.

The remaining of the paper is organized as follows. Section 2 discusses the related work. We introduce the distance similarity function of two uncertain trajectories in Sections 3. In Section 4, we give the definitions of uncertain group patterns, and then introduce our efficient algorithm based on breath-first search strategy. Experiments testing effectiveness and efficiency are shown in Section 5. Finally, our study is concluded in Section 6.

2 Related work

Group pattern mining over exact trajectory data. Group pattern mining, which is to discover a group of objects that move together for a certain time period, is an important data analysis task for moving object trajectories. The research mainly included flock, convoy, and swarm pattern mining. The concepts of group patterns can be distinguished based on how the group is defined and whether they require the time periods to be consecutive. Specifically, a flock [9, 10] is a group of objects that travel together within a disc of some user-specified size for at least k consecutive time stamps. A major drawback is that a circular shape may not reflect the natural group in reality, which may result in the so-called lossy-flock problem [11]. To avoid rigid restrictions on the sizes and shapes of the group patterns, the convoy is proposed to capture

generic trajectory pattern of any shape and extent by employing the density-based clustering. Instead of using a disc, a convoy requires a group of objects to be density-connected to each other during k consecutive time points. While both flock and convoy have strict requirement on consecutive time period, the rigid definition of flock and convoy sometimes makes it not practical to find potentially interesting patterns. In contrast to flock, convoy, Li et al [1] proposed a more general type of trajectory pattern, called swarm, which is a cluster of objects lasting for at least k (possibly non-consecutive) timestamps. Because this is more realistic as different people may temporarily leave the cluster at some snapshots, in this paper our uncertain group pattern definition is based on swarm and is extended to uncertain data model. Although there are a lot of works on mining group patterns, all these algorithms are designed for exact trajectory data and cannot be extended to uncertain trajectory data directly.

Pattern mining on uncertain data. Another set of researches related with our work are pattern mining over uncertain data. Existing work on mining frequent itemsets from uncertain databases falls into two categories based on the definition of a frequent itemset: expected support-based frequent itemset and probabilistic frequent itemset. Both definitions consider the frequency (support) of an itemset as a discrete random variable. The former employs the expectation of the support as the measurement. That is, an itemset is frequent only if the expected support of the itemset is no less than a specified minimum expected support. The latter uses the frequentness probability as the measurement, which is the probability that an itemset appears no less than a specified minimum support times. Then, an itemset is frequent only if its frequentness probability is no less than a specified minimum probability threshold. However, the use of expected support may lead to the loss of important patterns. Thus, the use of a probabilistic frequentness measure has been more popular recently. A recent survey for comparing these two measures and analyzing their relationships is given in [16].For the problem of uncertain sequence pattern mining, some initial research has been undertaken. For example, the expected support-based frequent sequential pattern mining has been studied in [17]. In contrast, Zhao et al. [18] proposed to mine probabilistic frequent sequential patterns according to the frequentness probability. However, all of these works only considered the simple value-based data type, and are not suitable for the complex data type-trajectory data, which contains both spatial and temporal information. Our paper is the first work to solve the group patterns mining problem on uncertain trajectory data.

3 Similarity metric of two uncertain locations

In this section, we present our method of computing the spatial proximity of objects with uncertain locations. We formalize the model of uncertain trajectories in Section 3.1 and give the similarity distance function in Section 3.2. Frequently used symbols throughout this paper are summarized in Table 1.

3.1 Uncertain trajectory model

Let $T_{DB} = \{t_1, t_2, \cdots t_n\}$ be a linearly ordered list of n timestamps. Let O_{DB}=$\{o_1, o_2, \cdots, o_m\}$ be a collection of m moving objects that appear in T_{DB}. The locations of objects o observed at timestamps T_{DB} are uncertain. We first give the definition of a certain trajectory.

Definition 1. (Certain Trajectory). A certain trajectory T_r is represented as a sequence of points $\{(x_1, y_1, t_1), (x_2, y_2, t_2), \cdots, (x_n, y_n, t_n)\}$ $(t_1 < t_2 < \cdots < t_n)$, where n is the number of points in the trajectory and (x_i, y_i) are the coordinates of the ith point at timestamp t_i.

In some uncertain trajectories' studies [6,7], they commonly used probability density function (*pdf*) to represent the uncertainty. Unfortunately, the exact probability distribution is not easily computed. So in this paper, we adopt the expectation and variance to model the uncertain trajectory. We can use the Evolving Density Estimator [8] to compute the mean value u and standard deviation v of an object's position at each time. Based on the definition of certain trajectories, we have the definition of an uncertain trajectory as follows:

Definition 2. (Uncertain Trajectory). An uncertain trajectory UTr is a sequence of random variables, and all the random variables at different timestamps are assumed to be independent. An Utr is represented as $\{(x_1, y_1, u_1, v_1, t_1), (x_2, y_2, u_2, v_2, t_2), ..., (x_n, y_n, u_n, v_n, t_n)\}$ $(t_1 < t_2 < ... < t_n)$, where u_i and v_i are the expectation and variance of Utr at timestamp t_i.

Table 1: Summary of the use of notations

Symbol	Illustration
O_{DB}	Moving object set, $O_{DB} = \{o_1, o_2, o_3, ..., o_m\}$
o_j	The jth object
O	Objects subset,$O \subseteq O_{DB}$
T_{DB}	Timestamp set, $T_{DB} = \{t_1, t_2, ..., t_n\}$
t_i	Timestamp t_i
T	Time subset,$T \subseteq T_{DB}$
UTr_i	The ith object's uncertain trajectory
$UTr_i(j)$	The jth sample point of the ith uncertain trajectory
$dist_i(UTr_a, UTr_b)$	The distance of the two objects at time t_i
C_{DB}	The database of clusters using FCM algorithm
C_{t_i}	The clusters at time t_i
c_{t_ij}	The jth cluster of C_{t_i}
$C_{t_i}(o_j)$	The set of clusters that object o_j is in at timestamp t_i
$p(o_j \in c_{t_ij})$	The probability of object o_j belonging to an cluster c_{t_ij}
$p(O \in c_{t_ij})$	The probability of objects O belonging to an cluster c_{t_ij}
$p(O \in C_{t_i})$	The probability of objects O belonging to the clusters C_{t_i}
min_o	The minimum objects threshold
min_t	The minimum timestamps threshold
$minprob$	The minimum probability threshold
$\Pr(\text{support}(O,T) \geq min_t)$	The probability that objects of O are in the same cluster for at least min_t timestamps

3.2 Expected distance function

Before giving the definition of the uncertain group pattern, we introduce a novel and adaptive metric EE-distance (expected Euclidean distance) for measuring the similarity between two uncertain trajectories.

Trajectory similarity is commonly estimated using trajectory distance measures, such as the Euclidean distance, the dynamic time warping (DTW) distance, the principal component analysis (PCA) distance, the edit distance with real penalty (ERP), and the longest common subsequence (LCSS) distance. However, there is no trajectory similarity measure that can beat all the others in every circumstance as introduced in [19]. In this paper we adopt the Euclidean distance due to its simplicity in implementation and low computation complexity. Next, we will show how to

compute the uncertain instant distance between two trajectories based on the expected Euclidean distance.

Definition 3. (Uncertain Instant Distance). We can treat the distance between two uncertain trajectories at timestamp t_i as a square sum of the sample points, as shown in equation(1), $UTr_a(i)$ is the sample point of uncertain trajectory UTr_a at timestamp t_i, the same as $UTr_b(i)$.

$$dist_i(UTr_a, UTr_b) = (UTr_a(i) - UTr_b(i))^2 \tag{1}$$

$UTr_a(i)$ and $UTr_b(i)$ are the independent random variables, so $dist_i(UTr_a, UTr_b)$ is also the random variable. The expectation of the random variable $dist_i(UTr_a, UTr_b)$ can be computed in equation (2).

$$\begin{aligned} E((UTr_a(i) - UTr_b(i))^2) &= E(UTr_a(i))^2 + var(UTr_a(i)) \\ &+E(UTr_b(i))^2 + var(UTr_b(i)) - 2E(UTr_a(i)) \cdot E(UTr_b(i)) \end{aligned} \tag{2}$$

According to the equation (2), we can easily compute the expected distance of two uncertain locations in $O(1)$ time complexity. Unlike other works requiring exact probability distribution function, our distance formulation is statistically sound and only requires knowledge of the general characteristics of the data distribution, namely, its mean and variance.

4 Mining Probabilistic Frequent Group Patterns

In this section, we first provide definitions of our probabilistic frequent group patterns, and then show how to find all probabilistic frequent group patterns in an uncertain trajectory database.

4.1 Probabilistic frequent group patterns definition

Definition 4. (frequent group pattern)[1]. A pair (O, T) is said to be a group pattern if all objects in O are in the same cluster at any timestamp in T. Specifically, given two minimum thresholds min_o and min_t, for (O, T) to be a frequent group pattern, where $O = \{o_{i_1}, o_{i_2}, ..., o_{i_l}\} \subseteq O_{DB}$ and $T \subseteq T_{DB}$, it needs to satisfy two requirements: (1) there should be at least min_o objects; (2) objects in O are in the same cluster for at least min_t timestamps.

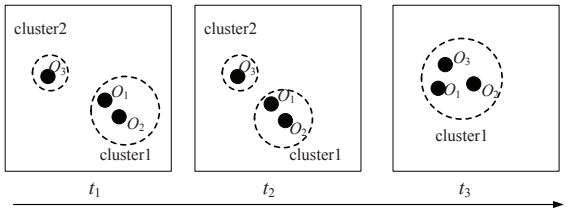


Figure 1: Object clusters at each timestamp in certain database

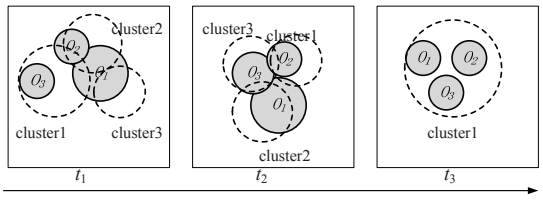


Figure 2: Object clusters at each timestamp in uncertain database

Fig.1 shows an example. There are 3 objects and 3 timestamps, $O_{DB} = \{o_1, o_2, o_3\}$, $T_{DB} = \{t_1, t_2, t_3\}$. Each sub-figure is a snapshot of object clusters at each timestamp. It is easy to see that o_1 and o_2 travel together for most of the time. Given min_o=2 and $min_t = 2$, there are only one frequent group pattern:$(\{o_1, o_2\}, \{t_1, t_2, t_3\})$.

As shown in Fig.2, in the uncertain scenario, the object's location is not an exact position, but an uncertain range. Thus, it is more realistic to cluster the object in different clusters than to a single cluster. Based on this point, we can apply a fuzzy clustering algorithm (e.g.FCM [20,21]) to create the cluster at timestamp t_i. Fuzzy clustering would assign each object a Ąºdegree of belongingnessĄą(belongingness probability) for each cluster. A simple example of clusters is given in table 2.

Table 2: An example of uncertain clusters.

Time	Uncertain clusters
1	$c_{11} = \{o_1 : 0.3; o_2 : 0.5; o_3 : 1.0\}, c_{12} = \{o_1 : 0.5; o_2 : 0.5\}, c_{13} = \{o_1 : 0.2\}$
2	$c_{21} = \{o_1 : 0.3, o_2 : 0.8\}, c_{22} = \{o_1 : 0.7, o_3 : 0.2\}, c_{23} = \{o_2 : 0.2, o_3 : 0.8\}$
3	$c_{31} = \{o_1 : 1.0, o_2 : 1.0, o_3 : 1.0\}$

An object $o_j \in O_{DB}$ with an uncertain location at timestamp $t \in T_{DB}$ has a belongingness probability of cluster $c \in C_{DB}$, the probability is denoted as $p(o_j \in c)$, where $p(o_j \in c) \in [0, 1]$. At timestamp t_i, an object could belong to more than one cluster, we use $C_{t_i}(o_j)$ to denote the set of clusters that object o_j is in, the total belongingness probability of object o_j in different clusters is one $p(C_{t_i}(o_j)) = \sum_{c \in C_{t_i}} p(o_j \in c) = 1$. In addition, for a given objectset O, we write $C_{t_i}(O) = \bigcap_{o_j \in O} C_{t_i}(o_j)$, which means O occurs at timestamp t_i in the same cluster. We assume that different objects and different timestamps are mutually independent, i.e., the belongingness probability of an object has no effect on those of other objects. So, the probability of the object set O belonging to a cluster c could be computed in equation (3), the probability of object set O in the same clusters at timestamp t_i could be computed in equation (4). To make our framework more general, we take clustering as a preprocessing step. The details are given in example 1.

$$p(O \in c_{t_ij}) = \prod\nolimits_{o_j \in O} p(o_j \in c_{t_ij}) \tag{3}$$

$$p(C_{t_i}(O)) = \sum\nolimits_{c_{t_ij} \in C_{t_i}} p(c_{t_ij} \in C_{t_i}) \tag{4}$$

Example 1: For $T_{DB} = \{t_1, t_2\}, O_{DB} = \{o_1, o_2, o_3\}$ in table 2, at timestamp $t_1 : c_{11} = \{o_1 : 0.3; o_2 : 0.5; o_3 : 0.8\}$; $c_{12} = \{o_1 : 0.5; o_2 : 0.5; o_3 : 0.2\}$; $c_{13} = \{o_1 : 0.2\}$, $O = \{o_1; o_2\}$, then $p(o_1 \in c_{11}) = 0.3$, $p(o_2 \in c_{11}) = 0.5, p(O \in c_{11}) = 0.3 * 0.5 = 0.15$,and $p(C_{t_1}(O)) = p(O \in c_{11}) + p(O \in c_{12})$=$0.15 + 0.25 = 0.4$. In the same way, at timestamp t_2 and t_3, $p(C_{t_2}(O))$=$p(O \in c_{21})$=0.24, $p(C_{t_3}(O)) = 1.0$.

In the uncertain scenario, the number of timestamps that objects O in the same cluster at the timestamps T, denoted as $support(O, T)$, is no longer certain. Instead co-occurrence is described by a discrete probability distribution function. As shown in example 1, the probability of the objects $O = \{o_1; o_2\}$ occurring in the same cluster at timestamp t_1 , t_2 and t_3 is 0.4, 0.24 and 1.0 respectively. The frequency distribution is described in table 3. For example, the probability that O occurring in the same cluster at least two timestamps is 0.054. The definition of probabilistic frequent group pattern should consider this uncertainty.

Table 3: Frequency distribution of $O = \{o_1; o_2\}$ occurring at timestamps $T = \{t_1, t_2, t_3\}$

Frequency of timestamp	≥ 0	≥ 1	≥ 2	≥ 3
Probability	1.0	1.0	0.054	0.096

Definition 5. (Probabilistic frequent group pattern). A pair (O, T) is an probabilistic frequent group pattern $iff(O, T)$ is a frequent group pattern and $\Pr(\text{support}(O, T) \geq min_t) \geq minprob$, $minprob$ is the probability threshold.

For a given group pattern (O,T), the frequentness probability, $\Pr(\text{support}(O,T) \geq min_t)$, which is interpreted as the probability that objects of O are in the same cluster for at least min_t timestamps. Under the definition of the probabilistic frequent group pattern, it is critical to compute the frequent probability of a group pattern efficiently. The frequentness probability $\Pr(\text{support}(O,T) \geq \min_t)$, could be computed by means of the paradigm of dynamic programming shown in equation(5) [22].

$$P_{\geq i,j}^{(O,T)} = \begin{cases} p_{\geq i-1,j-1}^{(O,T)} * p(O \in C_{t_j}) + p_{\geq i,j-1}^{(O,T)} * (1 - p(O \in C_{t_j})) & O \in C_{t_j} \\ p_{\geq i,j-1}^{(O,T)} & O \notin C_{t_j} \end{cases} \tag{5}$$

For the sake of the following discussion, we define that $P_{\geq i,j}^{(O,T)}$ denotes the probability that objects O appears at least i timestamps among the first j timestamps in the given time set T. The dynamic programming approach is to split the problem of computing $P_{\geq i,j}^{(O,T)}$ at the first j timestamps into sub-problems of computing the frequentness probabilities at the first $j-1$ timestamps. Our goal is to find all probabilistic frequent group patterns. Note that even though our problem is defined in the similar form of uncertain frequent pattern mining [22], none of previous work in uncertain frequent pattern mining area can solve exactly our problem. Because FP mining problem takes transactions as input, group pattern discovery takes clusters at each timestamp as input. If we treat each timestamp as one transaction, each "transaction" is a collection of "itemsets"rather than just one itemset.Therefore, there is no trivial transformation of FP mining problem to group pattern mining problem. The difference demands new techniques to specifically solve our problem.

4.2 Uncertain trajectory data preprocessing

When mining probabilistic frequent group patterns, we assume that each moving object has a reported location at each timestamp. However, in most real cases, the raw data collected is not as ideal as we expected. The sampling timestamps for different moving objects are usually not synchronized. Even though many complicated interpolation methods could be used to fill in the missing data with higher precision, any interpolation is only a guessing of real positions. In this paper, we use linear interpolation to obtain possible position and statistical value at an arbitrary time between two consecutive sample times. We define the $o(t) = \{x_t, y_t, u_t, v_t\}$ as an uncertain object between two consecutive sample timestamp t_i and t_{i+1} , the value of $o(t)$ can be computed in equation(6).

$$\begin{aligned} x_t = x_i + (x_{i+1} - x_i) \cdot \frac{t - t_i}{t_{i+1} - t_i}, y_t = y_i + (y_{i+1} - y_i) \cdot \frac{t - t_i}{t_{i+1} - t_i} \\ u_t = u_i + (u_{i+1} - u_i) \cdot \frac{t - t_i}{t_{i+1} - t_i}, v_t = v_i + (v_{i+1} - v_i) \cdot \frac{t - t_i}{t_{i+1} - t_i} \end{aligned} \tag{6}$$

In order to make our approach has extensibility, we treat the computation of the spatial proximity and construction the clusters of objects with uncertain locations as a preprocessing step. In this way, spatial proximity of objects and clustering methods can be flexible and application-dependent. We only require a set of clusters as input for each timestamp, where each object is associated with a belongingness probability that specifies the confidence the object is in a cluster at a given timestamp. The preprocessing algorithm is as follows.

Algorithm Preprocessing. Preprocessing algorithm

input: uncertain trajectory database UTD
output: uncertain clusters database C_{DB}
1.For each object o_i
2. For each timestamp t_j
3. if (no sample value)
4. interpolate the value;
5. compute the expected distance $dist_j(UTr_i, UTr_k)$ for each object o_k in O_{DB};
6.use FCM cluster algorithm based on the expected similarity;
7.output uncertain clusters database C_{DB};

For each object o_i, we compute the expected distance of every object o_k at time t_j (line 1-5), then we uses the FCM mining algorithm to cluster the uncertain objects at each timestamps(line 6). Finally, we get the uncertain clusters database C_{DB}(line 7).

4.3 Pruning Techniques

Although we use the dynamic programming to compute the frequent probability of a group pattern, the cost of computation is still high, the time complexity is $O(n^2)$, n is the number of timestamps in T. Fortunately, we find two efficient pruning rules based on properties of the frequent probability.

Pruning rule 1(count prune): Given a group pattern (O,T), if $cnt(O,T) < \min_t$, then (O,T) is not a probabilistic frequent group pattern,$cnt(O,T)$ is the numbers of timestamps in T that O in the same clusters.

Pruning rule 2 (expected prune): Given a group pattern (O,T), $e\sup(O,T)$ is the expected support of a group pattern (O,T), is defined as the sum of the probabilities that objects O occurring in the same cluster in each of the timestamps in T, $e\sup(O,T) = \sum_{t_j \in T} p(O \in C_{t_i})$. If $\lambda^- < minprob$, then (O,T) is not a probabilistic frequent group pattern.

$$\lambda^- = \begin{cases} \dfrac{-(\min_t - 1 - e\sup(O,T)^2)}{4e\sup(O,T)} & 0 < \dfrac{\min_t - 1 - e\sup(O,T)}{e\sup(O,T)} < 2e-1 \\ 2^{1+e\sup(O,T)-\min_t} & \dfrac{\min_t - 1 - e\sup(O,T)}{e\sup(O,T)} \geq 2e-1 \\ \dfrac{e\sup(O,T)}{\min_t} & other \end{cases} \tag{7}$$

Pruning rules could be used to prune infrequent group pattern before computed the exact frequent group probability. The running time of computing the $cnt(O,T)$ and $e\sup(O,T)$ is $O(n)$, but $O(n^2)$ for the exact probability $\Pr(\text{support}(O,T) \geq min_t)$. So pruning rules can significantly improve the running time of mining algorithm.

4.4 Mining algorithm of probabilistic frequent group pattern

Traditional frequent itemset mining is based on support pruning by exploiting the anti-monotonic property of support. In uncertain databases, recall that support is defined by a probability distribution and that we mine group patterns according to their frequentness probability. It turns out that the frequentness probability is also anti-monotonic.

Theorem 6. $\forall O \subseteq O(prime), \Pr(\text{support}(O,T) \geq min_t) \geq \Pr(\text{support}(O(prime),T) \geq min_t)$, *all subsets of a probabilistic frequent group patterns are also probabilistic frequent group patterns.*

We can use the Theorem 1 to prune the search space for probabilistic frequent group pattern. That is, if a group pattern (O,T) is not a probabilistic frequent group pattern,i.e.Pr(support$(O,T) \geq min_t) < minprob$, then all patterns $O(prime) \supset O$ cannot be probabilistic frequent group patterns.

We propose a probabilistic frequent group patterns mining approach based on the Apriori algorithm. Like Apriori, our method iteratively generates the probabilistic frequent group patterns using a bottom-up strategy. Each iteration is performed in two steps, a join step for generating new candidates and a pruning step for calculating the frequentness probabilities and extracting the probabilistic frequent group patterns from the candidates. The pruned candidates are, in turn, used to generate candidates in the next iteration. Theorem 1 is exploited in the join step to limit the candidates generated and in the pruning step to remove group patterns that need not be expanded. In the join step, our algorithm adopts the breadth-first implementation. Because the depth-first strategy does not fully use the downward closure of the probabilistic support,this is due to the fact that the depth-first implementation does not know all frequent k-objectset before considering the $(k+1)$-objectset. This may lead to a bigger search space. The detailed steps of our algorithm to compute the probabilistic frequent group pattern is listed in Algorithm PFGPM.

We first select frequent 1-object sets (line 1), and then recursively generate candidate $(k+1)$-objectset from k-objectset (line 2-9). At each iteration, only the frequent k-objectsets are extended (Apriori property, line 3-4). We first scan the database to calculate the expected support of each candidate, and use the pruning rules to prune candidate (line 5), if not be pruned, compute the frequentness probability(line 6-8). Next, we output those patterns satisfying $|O| \geq min_o$ and probabilistic support (line 9).

Algorithm PFGPM. Probabilistic frequent group pattern mining algorithm

input: uncertain clusters database C_{DB}
output: probabilistic frequent group patterns
1.Apply the pruning rules first, then calculate the frequentness probability for each 1-objectset,find all the frequent 1-objectset called $Cand$, and sort them in alphabetic order;
2.$K = 2$
3.For each O in $Cand$
4. extend O using a breadth-first search like strategy to its supersets with O as prefix, denoted $O(prime)$;
5. Using pruning rule 1 and 2 for $O(prime)$;
6. if($O(prime)$ not pruned)
7. Compute the frequent group probability(fgp) as shown in equation(5);
8. If($fgp \geq$ $minprob$)
9. OUTPUT $(O(prime), fgp)$, if$|O(prime)| \geq min_o$
10.$K = K + 1$;
11.go to 3;

5 Experiments

5.1 Datasets

Real data. A truck dataset (http://www.chorochronos.org/) recording 50 trucks delivering concrete to construction sites around Athens over 33 days and consisting of 276 trajectories. To increase the size of moving objects, we considered each distinct trajectory as the ID of an object, yielding 276 trucks with 2449 timestamps. In our experiments, we consider only the first 128

positions of each trajectory. We normalize the positions of trajectories into a unit space. The probabilities of each trajectory were assigned according to two different distributions: (1) Each certain position p was assigned a probability according to a uniform distribution in the range of (0.5, 1.0]. (2) Each position was assigned a normal distribution N(0.5,0.2) in the range of [0, 1.0].

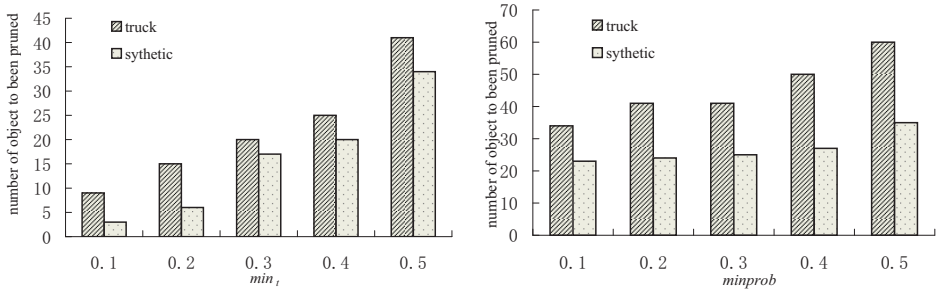
Synthetic data. We first generate two certain trajectories based on a custom GSTD data generator [23]. Each dataset contains 10,000 uncertain trajectories with the same length 128. We then convert these certain trajectories to uncertain trajectories in the way described above.

The fuzzy clusters of teach timestamp are obtained by the fuzzy c-means clustering algorithm [20] with $m = 2$ and $EPS = 0.01$, where m is the weighting exponent and EPS is the termination criterion. Each object is assigned a belongingness probability by the fuzzy clustering algorithm. The default values of min_o=10, min_t=0.5,(half of the overlapping time span), $minprob$=0.3.

5.2 Performance evaluation

No previous technique addresses probabilistic frequent group patterns mining for uncertain trajectories. We compare our PFGPM approach with an alternative, PFGPM-NP, that does not use any pruning rules.We present the experimental results in this section. All the experiments are run on a desktop PC with a 2.66GHz CPU and 4GB RAM.

Effect of punning rules: We use two dataset with varying parameter settings to test the performance of the pruning rules. (1) Synthetic dataset. We extract the first 274 trajectories of synthetic dataset; (2) Truck datasets. Each position p in two datasets was assigned a probability according to a uniform distribution approach.As shown in Fig.3, the pruning works well for skewed dataset (truck dataset). The reason is straightforward: the more skewed the data, the higher the number of objects is infrequent and, thus, cannot be pruned. Regarding the effect of parameters, the larger min_t and $minprob$ are, the more objects will be pruned. However,min_t has a more significant influence than $minprob$, that is, pruning rule 1 has more strong power than pruning rule 2.



(a) varying min_t (b) varying $minprob$

Figure 3: Pruning power

Effect of $|O_{DB}|$ and $|T_{DB}|$. We ran out our algorithm in synthetic data set. Fig.4 and Fig.5 depict the running time when varying $|O_{DB}|$ and $|T_{DB}|$ respectively. In both figures, PFGPM-NP is much slower than PFGPM. Furthermore, PFGPM-NP is usually 5 times slower than PFGPM. Comparing Fig.4 and Fig.5, we can see that PFGPM is more sensitive to the change of $|O_{DB}|$. This is because its search space is enlarged with larger $|O_{DB}|$, whereas the change of $|T_{DB}|$ increases the computing time of frequent probability, which does not directly affect the running time of PFGPM.

Effect of min_o. Fig.6 shows the running time w.r.t. min_o. With the increasing of min_o, the running time will decrease because less group patterns would meet the requirement. Besides, it is

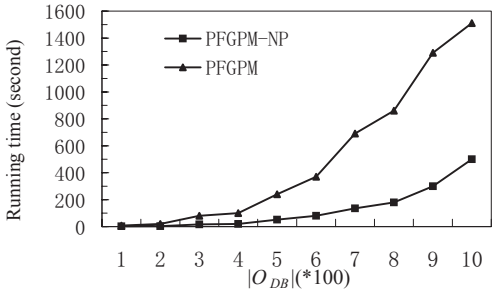


Figure 4: Running time with varying $|O_{DB}|$

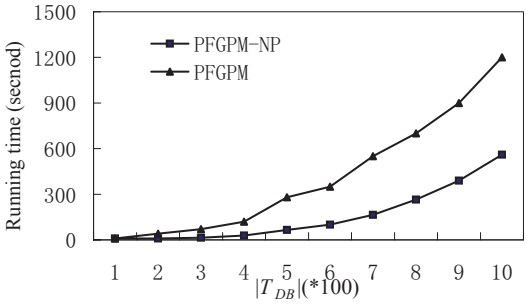
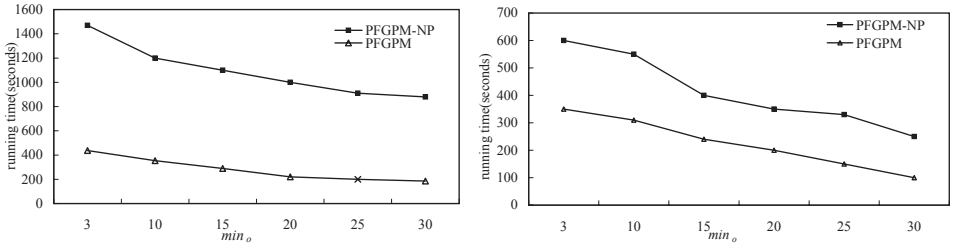


Figure 5: Running time with varying $|T_{DB}|$

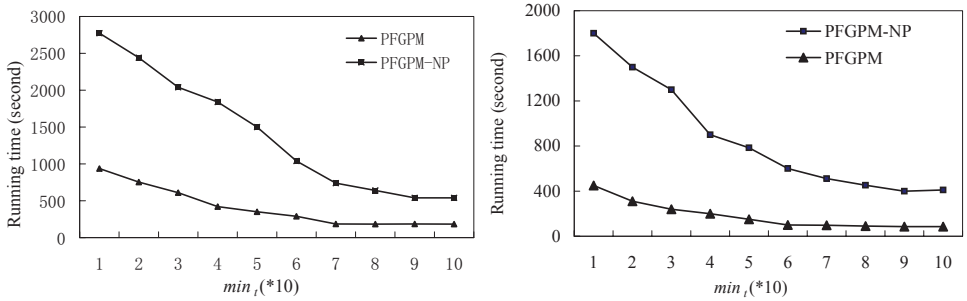
obvious that PFGPM-NP takes much longer time than PFGPM. The reason is that PFGPM-NP does not use any pruning rules to find the frequent group patterns, this leads to more redundant computation for frequent probability and larger search space.



(a) synthetic data set (b) truck data set

Figure 6: Object cluster at each timestamp in certain trajectory database

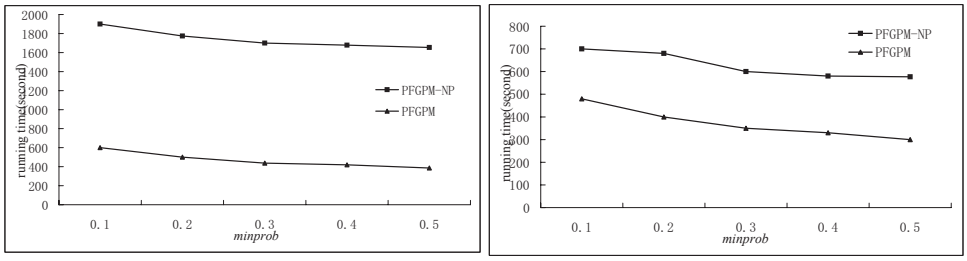
Effect of min_t. Fig.7 shows the influence of min_t on the runtime when using different data sets. When min_t decreases, we observe that the running time of all the algorithms goes up due to the number of probabilistic frequent group patterns increases. However, we find that the growth speed of PFGPM is low due to all pruning methods it employed.



(a) synthetic data set (b) truck data set

Figure 7: Running time with varying min_t

Effect of *minprob*. Finally, we test the running time of two compared algorithms with varying the probabilistic frequent threshold, *minprob* in two different datasets. In Fig.8, we can obverse that PFGPM is always faster than PFGPM-NP algorithm. With regards to the change of *minprob*, the running time of all algorithms remains approximately the same. Thus, we can discover that the influence of probabilistic frequent threshold will be smaller than that of min_t to the total running time.



(a) Synthetic data set (b) truck data set

Figure 8: running time with varying *minprob*

6 Conclusion

In this paper, we have formulated and studied the problem of mining probabilistic frequent group patterns in uncertain trajectory database. We introduce a novel notion expected distance to measure the dissimilarity between uncertain locations. In order to mining such group patterns efficiently, we proposed several pruning techniques to reduce search space and to avoid many complicated computations. We further designed a Apriori-based algorithms using breadth-first implementations for efficient enumeration of all probabilistic frequent group patterns from uncertain data. Extensive experimental results show the effectiveness and efficiency of the mining algorithm. In our further study, we aim to extend our current approach to be able to handle more complex patterns for the trajectory data.

Acknowledgement

This research is partially supported by Fundamental Research Funds for the Central Universities (Grant No. 130317003), the National Natural Science Foundation of China (Grant No. 61440014, Grant No. 61300196), the Natural Science and Technology Major Project (Grant No. 2103ZX03002006), and the Liaoning Province Science and Technology Project (Grant No. 2013217004).

Bibliography

[1] Z. Li, B. Ding, et al, Swarm: Mining relaxed temporal moving object clusters, *the VLDB Endowment*, 3(1-2):723-734.

[2] D. Wegener, D. Hecker, et al, Parallelization of R-programs with GridR in a GPS-trajectory mining application, *1st Ubiquitous Knowledge Discovery Workshop on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, (2008).

[3] X. Li, J. Han, et al, Traffic density-based discovery of hot routes in road networks, *the 10th International Symposium on Spatial and Temporal Databases*, 441-459, (2007).

[4] Z.Li, J.G. Lee, et al, Incremental Clustering for Trajectories, *the 15th Database Systems for Advanced Applications*, 32-46, (2010).

[5] X. Li, J. Han, et al, Motion-alert: automatic anomaly detection in massive moving objects, *the 4th IEEE International Conference on Intelligence and Security Informatics*, 166-177.

[6] N. Pelekis, I. Kopanakis, et al, Clustering uncertain trajectories, *Knowledge and Information Systems*, 28(1): 117-147.

[7] M. Chunyang, L. Hua , et al, KSQ: Top-k Similarity Query on Uncertain Trajectories,*Knowledge and Data Engineering, IEEE Transactions*, 25(9): 2049-2062.

[8] J. Hoyoung, Managing Evolving Uncertainty in Trajectory Databases, *IEEE Transactions on Knowledge and Data Engineering*, 26(7): 1692-1705.

[9] J. Gudmundsson, M. V. Kreveld, Computing longest duration flocks in trajectory data, *the 14th annual ACM international symposium on Advances in geographic information systems*, 35-42, (2006).

[10] J. Gudmundsson, M. V. Kreveld, et al, Efficient detection of motion patterns in spatio-temporal data sets, *the 12th annual ACM international symposium on Advances in geographic information systems*, 250-257, (2004).

[11] H. Jeune, M. Yiu, et al, Discovery of convoys in trajectory databases, *the VLDB Endowment*, 1(1):1068-1080.

[12] H. Jeune, H. Shen, et al, Convoy queries in spatio-temporal databases, *the 24th International Conference on Data Engineering*, 1457-1459, (2008).

[13] L.A. Tang, Y. Zheng, et al, A Framework of Traveling Companion Discovery on Trajectory Data Streams, *ACM Transaction on Intelligent Systems and Technology*, 5(1):3.

[14] L.A. Tang, Y. Zheng, et al, Discovery of Traveling Companions from Streaming Trajectories, *the 28th IEEE International conference on Data Engineering*, 186-197, (2012).

[15] K. Zheng, Y. Zheng, et al, Online Discovery of Gathering Patterns over Trajectories, *IEEE Transactions on Knowledge and Data Engineering*, 26(8): 1974-1988.

[16] Y. Tong, L. Chen, et al., Mining frequent itemsets over uncertain databases, *VLDB Endowment*, 5(11): 1650-1661.

[17] M. Muzammal, R. Raman, Mining sequential patterns from probabilistic databases, *the 15th Pacific-Asia conference*, 210-221, (2011).

[18] Z. Zhao, D. Yan, et al, Mining probabilistically frequent sequential patterns in uncertain databases, *the 15th International Conference on Extending Database Technology*, 74-85,(2012).

[19] H. Wang, H. Su, K. et al, An Effectiveness Study on Trajectory Similarity Measures, *the 24th Australasian Database Conference*, 13-22, (2013).

[20] J. Bezdek, R. Ehrlich, et al, FCM: The fuzzy c-means clustering algorithm, *Computers Geosciences*, 10(2):191-203.

[21] C. Hwang, F. C.-H. Rhee, Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means, *Fuzzy Systems*, 15(1):107-120.

[22] T. Bernecker, H.-P. Kriegel, et al, Probabilistic frequent itemset mining in uncertain databases, *the 15th ACM SIGKDD on Knowledge discovery and data mining*, 119-128,(2009).

[23] Y. Theodoridis, J. R. O. Silva, et al, On the generation of spatiotemporal datasets, *the 6th International Symposium on Advances in Spatial Databases*, 147-164, (1999).

INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 10(3):441-451, June, 2015.

Power-Aware Relay Selection and Routing Scheme for Multi-Interface Sensor Networks

M. Zheleva, H. Lee

Mariya Zheleva
Department of Computer Science
University at Albany, State University of New York, Albany, NY, USA
mzheleva@albany.edu

HyungJune Lee*
Department of Computer Science and Engineering
Ewha Womans University, Seoul, Republic of Korea
*Corresponding author: hyungjune.lee@ewha.ac.kr

Abstract: We present a joint relay and radio interface selection algorithm with packet deadline under limited power usage in multi-interface sensor networks. We find route optimization techniques in multi-interface networks: 1) selecting the most conservatively lowest-power interface to guarantee timely transmission considering the remaining hops to destination, and 2) searching detouring paths when the power level of an involved relay node is too low to use the necessary interface that guarantees timely delivery. We aim to achieve data delivery with packet deadline requirement while minimizing energy consumption at each node, and further prolonging network lifetime by selecting cost-effective relay nodes and wireless interfaces.
We evaluate our proposed algorithm in terms of total power consumption and packet delivery performance, compared to homogeneous radio interface scenarios of only Wi-Fi interface and only 802.15.4 ZigBee interface. Simulation results show that the proposed algorithm fully exploits the given packet delivery time to conserve power consumption by selecting the interface that consumes the least power, and spreading out network traffic over the network. Our proposed algorithm achieves reliable packet delivery without delivery failures due to power outage and missed deadline.
Keywords: Relay Selection, Routing, Multi-Interface Sensor Networks, Energy Efficiency.

1 Introduction

According to Gartner analyst firm, information technology (IT) accounts for 2% of the world's carbon emissions as of 2007 [3], and reducing network energy demands has been much more important than before. Devising energy-aware network services is regarded as one of the most critical design principles for making Green IT networks realizable.

As System-on-Chip (SoC) technology integrates a larger number of transistors in a given physical space with cheaper manufacturing cost, embedded networked systems are now installed with a variety of wireless network interfaces such as Wi-Fi, Bluetooth, 802.15.4 ZigBee, and even 3G, 4G LTE baseband. Thus, network devices can choose one of these network interfaces to send and receive data packets depending on data rate, energy consumption, and delay.

As multi-hop wireless sensor networks scale in size, their energy consumption grows as well. Each wireless network interface consumes different amount of energy depending on the design characteristics of physical, MAC, and network layer such as data rate and transmission range. For example, Wi-Fi interface can transmit data with a longer link compared to 802.15.4 ZigBee interface, but in turn consumes more power. Furthermore, using a different wireless interface incurs different transmission delay, which influences the achieved data rate. If a network device uses a lower-power wireless interface to relay data packets to the next-hop relay node, it will

consume less power, but will take longer to transmit. In embedded network systems, there exist real challenges and constraints of available remaining power to use (in battery-powered devices and sensor networked devices), and packet deadline for quality-of-service (QoS) requirement.

In this paper, we study the problem of relay selection and multi-hop routing using multiple wireless interfaces in sensor networks where each sensor node has limited power to use, and data packet should be delivered within a given packet deadline. We assume that all nodes in the network are stationary and have multiple radio interfaces for transmitting and receiving data. We consider all possible wireless interfaces that are expected to satisfy the packet deadline, and then choose the most energy-efficient interface among them. By selecting cost-effective relay nodes and wireless interfaces, we aim to achieve packet delivery within deadline while minimizing energy consumption and prolonging the overall network lifetime.

Prior work on relay selection in multi-interface ad-hoc networks [2, 9, 13, 14] have mainly focused on improving throughput or mitigating loss rate. However, they have not explicitly considered energy-awareness and packet deadline scenarios in multi-radio interface environments.

We propose a joint relay and radio interface selection algorithm with packet deadline under limited power availability in multi-interface sensor networks. To meet the packet deadline, each intermediate node en route picks up the lowest power radio interface among all possible interfaces that can guarantee timely delivery, expecting to use the same lowest interface over the remaining hops to the destination. In this way, intermediate relay nodes make their own distributed decision of relaying to the shortest next-hop node through the most energy-efficient interface while achieving packet delivery within the given deadline.

Further, we consider out-of-battery situations where a selected intermediate relay node runs out of power. To avoid the termination of packet transmission, we apply a local greedy forwarding technique that searches alternative paths (i.e., detouring paths) as long as the entire selected route can satisfy the packet deadline constraint. Although the high-level principle is similar to [11], we explicitly consider routing with packet deadline based on simple power cost calculation and consequent route decisions. Before the shortest route breaks, and the packet transmission is terminated, an intermediate node estimates that the next hop on the shortest-path would not have the necessary power for packet delivery within the allowed time limit, and then finds a detouring next-hop node that has enough power. Thus, the proposed algorithm diffuses network traffic by making more nodes involved with data routing, and accordingly balancing the network.

The rest of this paper is organized as follows: In Sec. 2, we discuss related work, and Sec. 3 provides system model. In Sec. 4, we describe the proposed scheme, and evaluate the approach in Sec. 5. Finally, we conclude this paper Sec. 6.

2 Related Work

There is a large body of work on relay selection in wireless multi-hop networks. Our focus is on energy-aware routing schemes in multi-interface sensor networks. Recent work on this problem can be categorized in three areas: 1) relaying over multi-interface networks, 2) energy-aware relay scheme, and 3) relay selection algorithm.

Relay selection in multi-interface networks have mainly focused on achieving high throughput or mitigating loss rate by proposing a new routing metric [2,9], or by exploiting the properties of multiplexing and diversity in the physical layer [13,14]. These previously proposed relay schemes incorporate fundamental ideas from MIMO (multiple-input and multiple-out) into multi-interface networks. They aim to select better relay paths with multiple radios to improve the corresponding performance metrics (i.e., throughput, loss rate).

Regarding energy-aware relay scheme, previous works have been studied mostly in the context of sensor networks. Due to inevitable requirement of battery usage in sensor networks, a variety

of energy-efficient routing techniques are proposed in [10, 12, 15, 18] to prolong network lifetime. They find efficient and low overhead relay nodes for data delivery to destination through data mining-based network optimization [10], network clustering [18], cooperative beamforming [12], or balancing between shorter high-quality links, and longer lossy links [15].

Previous works [6, 8, 16, 17, 19] have proposed relay selection schemes in cooperative networks that utilize game theory [8, 16, 19] or Markov modeling [6, 17]. The proposed schemes choose the optimal cooperative relay nodes that maximize their own utilities, while contributing to the entire network benefit.

Our proposed work is different from the previous works summarized above, in that it considers all of three design principles together, and thus develops an energy-aware relay scheme in multi-radio interface sensor networks.

3 System Model

This paper considers the problem of relay and interface selection in multi-radio interface sensor networks. The objective of our proposed method is to minimize power consumption at each sensor node and prolong network lifetime when delivering data packets from a source node to a destination node, while satisfying the packet deadline constraint. Each sensor node is stationary and is installed with multi-radio interfaces, e.g., Wi-Fi and 802.15.4. A sensor node can determine *where to relay* (i.e., next-hop) the received data packet and through *which radio interface*. A high-power radio interface (e.g., Wi-Fi) will consume more power, but can transmit farther and faster compared to a lower-power radio interface such as 802.15.4.

We consider data delivery applications under delay constraints in battery-powered sensor networks. Effectively, our algorithm diversifies route paths in the network in order to prolong the overall network life-time. We find an alternative path for robust data delivery before the route path breaks due to energy depletion. We assume fixed packet size and power transmission for each interface type. We focus on the single hop transmission delay since it is largely affected by data rate of that interface compared to queueing delay and MAC access time. We use the distance-vector algorithm as the underlying routing mechanism. We do not constrain ourselves with a specific one-to-one routing protocol; other advanced routing protocols can be used with our proposed algorithm. We assume that control beacon messages for maintaing up-to-date topology include the current power status of the originator. In this way, a sensor node knows information of power availability of neighboring nodes and can use it for its route decision of selecting the next relay node and the radio interface.

4 Relay Selection Scheme

In our network setup, a packet is transmitted from a source node to a destination node. Upon packet generation at the source node, the packet is assigned a delivery deadline. Each node along the routing path updates the remaining time before deadline in the packet header before relaying it further. Each relay node makes a local decision for interface type and next hop when transmitting the packet, while carefully considering the remaining time before deadline. Various factors such as transmission time and power consumption as well as power availability influence the success of packet delivery. We design three route selection algorithms that take these factors into account. We now describe our routing schemes in turn.

Each of our route selection algorithms takes as an input the packet delivery deadline R, as well as the time D and transmission power P through each interface $j \in \{1, \ldots, M\}$ (where M is the number of available radio interfaces). D and P are vectors of size M, and are ini-

tialized in increasing order of transmission power (i.e., $D_1 > D_2 > \ldots > D_j > \ldots > D_M$ and $P_1 < P_2 < \ldots < P_j < \ldots < P_M$). Each sensor node has topology information for each interface type j ($Topology$) and can calculate the shortest path between source and destination. To do this, a sensor node utilizes Dijkstra's algorithm, where each edge between nodes in the topology is weighted with the expected transmission count ETX [1] as per-hop cost through the corresponding link. Considering this information, each node along the path makes a local routing decision to select interface and next hop in order to meet the packet delivery deadline while minimizing the overall power consumption in the system. Informally, our route selection algorithms take advantage of the provided packet delivery deadline and prolong the packet transmission time in order to conserve energy by selecting low power transmissions. We now describe in more detail our proposed route selection schemes.

4.1 Naive Method

We first describe a *Naive* counterpart inspired by Longest-Job-First scheduling algorithm [5] such that the longest-time-spending interface is assigned first due to the least power consumption. This algorithm only considers the remaining time before packet deadline in the scope of the local neighborhood in a *myopic* view. As detailed in Algorithm 1, each decision-making node (i.e. $currentNode$) finds the lowest power interface that takes less time than the remaining time before deadline, $R_{currentNode}$ and selects this interface. Naturally, this algorithm favors low power interfaces, which may result in missed deadlines.

Algorithm 1 Naive

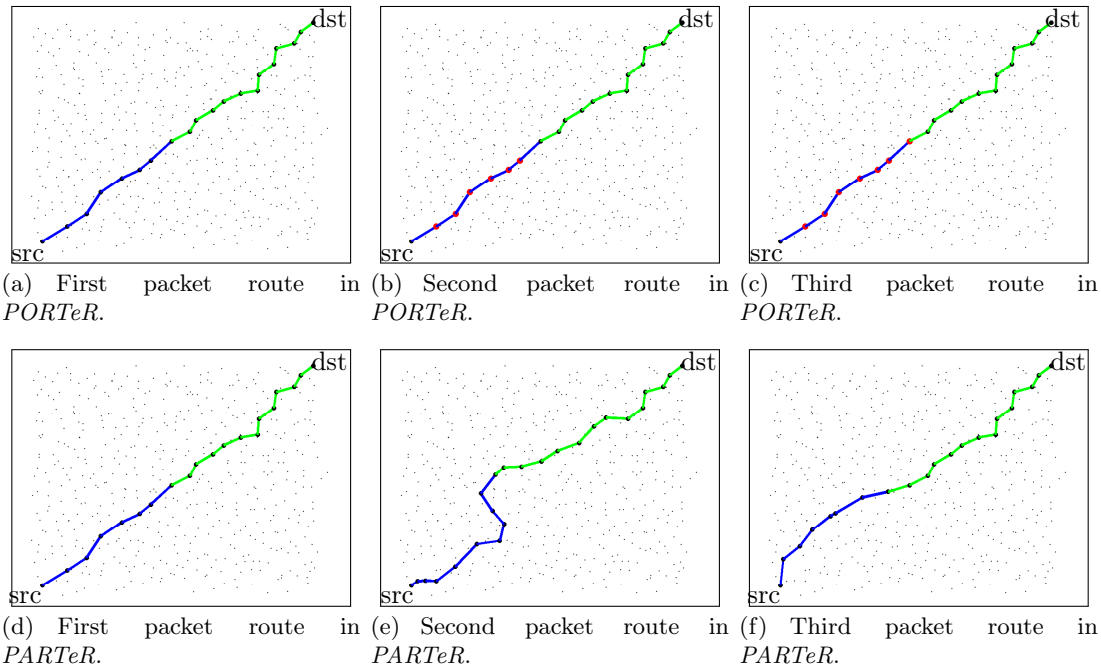
Input:src, dst, $D, P, Topology$
Output: Route from src to dst
currentNode=src
while $currentNode \neq dst$ **do**
  for all interfaces $j \in \{1, \ldots, M\}$ **do**
    $[path_j, cost_j] = Dijkstra(currentNode, dst, Topology_j)$
  end for
  Select $\arg\min_{j \in \{1,\ldots,M\}} P_j$, s.t. $D_j \leq R_{currentNode}$
  $currentNode = path_j[indexOf(currentNode) + 1]$
  Transmit packet (update remaining time)
end while

4.2 Power-Oblivious Routing with Strict Time Requirements

Our second routing scheme, called Power-Oblivious Routing with strict Time Requirements ($PORTeR$), enhances *Naive* by incorporating knowledge about the number of future hops N_j to the destination through each interface type $j \in \{1, \ldots, M\}$ in a *far-sighted* view. Particularly,

Algorithm 2 PORTeR

Input:src, dst, $D, P, Topology$
Output: Route from src to dst
currentNode=src
while $currentNode \neq dst$ **do**
  for all interfaces $j \in \{1, \ldots, M\}$ **do**
    $[path_j, cost_j] = Dijkstra(currentNode, dst, Topology_j)$
  end for
  Select $\arg\min_{j \in \{1,\ldots,M\}} P_j$, s.t. $N_j * D_j \leq R_{currentNode}$
  $currentNode = path_j[indexOf(currentNode) + 1]$
  Transmit packet (update remaining time)
end while



(a) First packet route in *PORTeR*. (b) Second packet route in *PORTeR*. (c) Third packet route in *PORTeR*.

(d) First packet route in *PARTeR*. (e) Second packet route in *PARTeR*. (f) Third packet route in *PARTeR*.

Figure 1: **Route selection.** Figures (a), (b) and (c) present routes selected by the *PORTeR* algorithm, whereas (d), (e) and (f) are routes taken by the *PARTeR* algorithm. Green links present 802.15.4 links, and blue links correspond to Wi-Fi links. Red dots denote the nodes that reach the maximum power limit. *PARTeR* successfully finds a detouring route path before the route path is broken due to energy depletion.

PORTeR checks if the number of remaining hops N_j through interface j multiplied by the transmission time D_j through interface j is lower or equal to the remaining time before packet deadline $R_{currentNode}$ (line 8 from Algorithm 2). By doing so, this scheme enforces in-time packet delivery in all feasible cases.[1] This scheme, however, does not consider power availability in nodes in the process of route selection.

4.3 Power-Aware Routing with Strict Time Requirements

Finally, we propose a route selection scheme that considers power availability in nodes when making route decisions. We call this scheme Power-Aware Routing with strict Time Requirements (*PARTeR*). Similarly to *PORTeR*, *PARTeR* also takes into account the number of future hops N_j and the transmission time D_j to enforce timely packet delivery. However, it takes one additional step to check if the selected next hop has enough power to receive and transmit the packet (lines 13-15 from Algorithm 3). If not, *PARTeR* selects an alternative next hop from its neighborhood.

Algorithm 4 provides details for the selection of alternative next hop. First, we identify all immediate neighbors of *currentNode* reachable through interface type j. We then calculate the path and cost to destination through each of the alternative next hop candidates. We select a random next hop candidate whose path cost ($costNxt_j$) is lower than the path cost of the current node ($cost_j$) (i.e., selecting a next-hop node closer to destination than the current node). The latter guarantees a loop-free route.

[1]We call a case feasible if transmission through the highest power interface M throughout the whole route guarantees timely delivery, i.e. $N_M * D_M \leq R$.

Algorithm 3 PARTeR

Input:src, dst, $D, P, Topology, PowerMap$
Output: Route from src to dst
currentNode=src
while $currentNode \neq dst$ **do**
  for all interfaces $j \in \{1, \ldots, M\}$ **do**
    $[path_j, cost_j] = Dijkstra(currentNode, dst, Topology_j)$
  end for
  Select $\arg\min_{j \in \{1,\ldots,M\}} P_j$, s.t. $N_j * D_j \leq R_{currentNode}$
  $P_{txCurNode} = P_j$; $P_{rxNextNode} = P_j$
  $nextNode = path_j[indexOf(currentNode) + 1]$
  Find $\arg\min_{j \in \{1,\ldots,M\}} P_j$, s.t. $N_j * D_j \leq R_{nextNode}$
  $P_{txNextNode} = P_j$
  if $PowerMap[nextNode] < P_{rxNextNode} + P_{txNextNode}$ **then**
    altNextHop($cost_j$,currentNode,dst,$Topology_j$)
  end if
  Transmit packet (update remaining time)
end while

Algorithm 4 altNextHop

Input:$cost_j, currentNode, dst, D, P, Topology, PowerMap$
Output: Alternative next hop
$ImNeigh = FindImmediateNeighbors(currentNode, j)$
for all nodes n in $ImNeigh$ **do**
  for all interfaces $j \in \{1, \ldots, M\}$ **do**
    $[pathNxt_j, costNxt_j] =$
    $Dijkstra(nextNode, dst, Topology_j)$
    Select $\arg\min_{j \in \{1,\ldots,M\}} P_j$, s.t. $N_j * D_j \leq R_{nextNode}$
    $P_{txNextNode} = P_j$; $P_{rxNextNode} = P_j$ {Next statement guarantees a loop-free route}
    if $costNxt_j < cost_j$ and $PowerMap[nextNode] < P_{rxNextNode} + P_{txNextNode}$ **then**
      $neighbors = neighbors.Append(n)$
    end if
  end for
end for
{Select a random alternative next hop from $neighbors$}
$altNextHop = SelectRandom(neighbors)$
return $altNextHop$

5 Evaluation

We validate our route selection schemes. We start with a description of our simulation environment and continue with analysis of route selection, and then evaluate network performance. We compare the routing schemes described in Sec. 4 with two basic counterparts that use only Wi-Fi and only 802.15.4 ZigBee (Sensor) interfaces; we call these counterparts *WiFi* and *Sensor*, respectively.

5.1 Simulation environment

We evaluate our routing algorithms in a MATLAB simulated network, which consists of 518 nodes in a $500 \times 500\text{m}^2$ grid. Each node has two interfaces: one 802.11 (Wi-Fi) and one 802.15.4 (Sensor). For our simulations, we use a realistic propagation model, which is based on a combined path-loss and shadowing model [4]. We use a path-loss exponent of 3, a reference loss of 46.67 dB and additive white Gaussian noise of $\mathcal{N}(0, 5^2)$ dB. We transmit packets of 10 KBytes at data rate of 11Mbps and 250 Kbps for Wi-Fi and Sensor, respectively. With these packet size and transmission rates, the transmission delays are $D_w = 0.89$ms through Wi-Fi and

D_s = 40ms through Sensor. We assume that power consumption for transmission is equal to that for reception and is C_w = 100mW for Wi-Fi and C_s = 1mW for Sensor [7].

For each experimentation, we transmit K packets in the network (specified below). As we evaluate power and time-related aspects of the system, we carefully select the initial power charge in nodes as well as the tolerated packet delivery deadline to demonstrate the benefits of our proposed algorithms. Particularly, we assign initial power of $K \cdot C_w$ at each node, which is enough for the transmission or reception of K packets through Wi-Fi, and vary the transmission delivery deadline between 0.2s and 1.6s to force only Wi-Fi and only Sensor transmission, respectively. We present our evaluation results below.

5.2 Route selection

We start our evaluation by presenting two examples of route selection by our *PORTeR* and *PARTeR* algorithms. Figure 1 presents the selected routes for a toy-example in which three consecutive packets are transmitted from node 1 to node 518. Each node in the mesh topology has 300 mW of power charge and nodes' battery levels are updated after each packet transmission and reception. As can be seen in Figure 1(a), 1(b), 1(c), *PORTeR* always selects the same route regardless of the power level of nodes along the path. This results in nodes running out of power as early as the second packet transmission. At the same time, *PARTeR* modifies the transmission route based on the prior knowledge of power change of the neighboring nodes as in Figure 1(d), 1(e), 1(f). As a result, *PARTeR* is capable of finding alternative routes without exhausting any of the nodes in the network.

5.3 Effects of packet delivery deadline

Next we explore the effects of packet delivery deadline on network performance. We present results for a single packet transmission from node 1 to node 518. For each run, we double the packet deadline, starting at $0.2s$.

Figure 2(a) presents our results for path length with increasing delivery deadline. As expected, *WiFi* and *Sensor* maintain the same path length irrespective of packet deadline. At the same time, all of *Naive*, *PORTeR* and *PARTeR* take advantage of the increasing delivery deadline by increasing the number of hops to the destination. This, as indicated in Figure 2(b), results in reduced power consumption as nodes favor lower power transmissions where time allows.

Finally, in Figure 2(c) and 2(d), we measure the remaining time till the deadline each algorithm incurs as the permitted delivery deadline increases. For low delivery deadline values, *Sensor* and *Naive* always miss the deadline (which appears as negative remaining time in the figures), whereas *WiFi*, *PORTeR* and *PARTeR* are well within deadline requirements. As we will see in subsection 5.4, *WiFi* and *PORTeR* do so at the cost of exhausting power resources in nodes, which results in packets being lost due to nodes along the route running out of power.

5.4 Network behavior over consecutive packet transmissions

We extend our evaluation by exploring performance trends when multiple consecutive packets are transmitted in the network. Particularly, we transmit 30 packets (K = 30) in a network where each node has 3 W of power to start with. The power level at each node is updated as the transmission of packets progresses. For each of the five routing schemes, we evaluate the number of nodes along a route that ran out of power, the path length taken, and the total consumed power. We also report the number of lost packets in two categories: (i) due to nodes running out of power and (ii) due to transmission taking longer than the permitted delivery deadline.

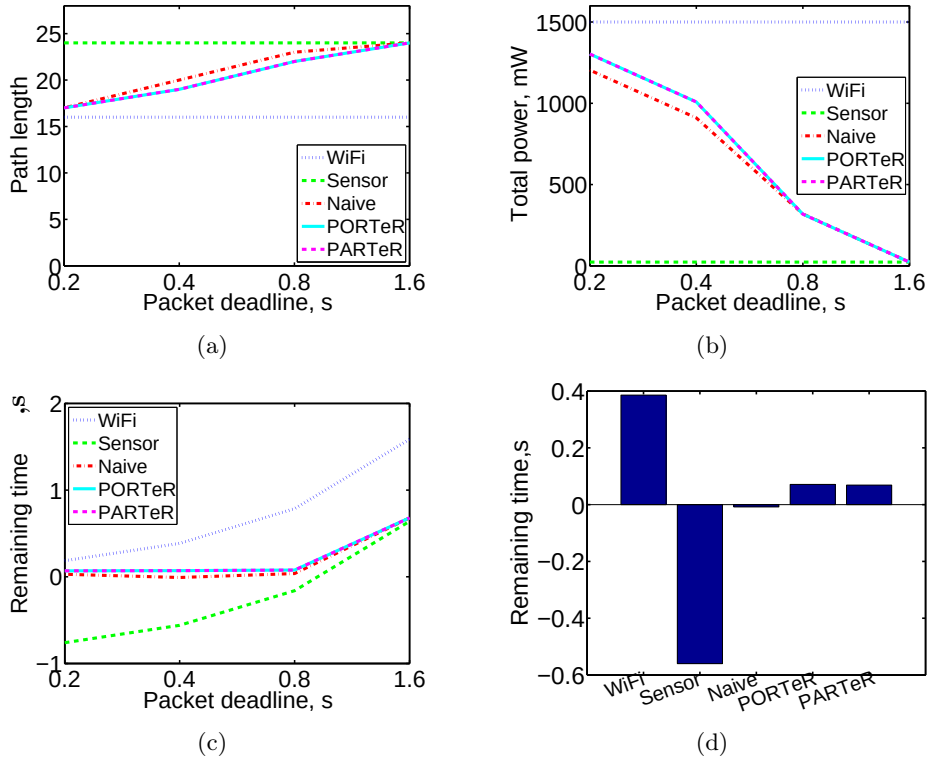


Figure 2: **Performance over increasing packet delivery deadline:** (a) Path length; (b) Total TX power; (c) Remaining time; (d) Remaining time for permitted delivery deadline of 0.4s.

We start with evaluation of the number of nodes that ran out of power for each algorithm. As Figure 3(a) shows, *Sensor* and *PARTeR* do not cause nodes to run out of power. As detailed in subsection 5.3, however, *Sensor* does this at the cost of missed packet delivery deadlines. *WiFi*, *Naive*, and *PORTeR* exhaust 15, 9 and 8 nodes respectively after the transmission of the 15th packet. In our simulations, we let the nodes out of power continue with packet transmission while counting the number of nodes with negative power, just to effectively show power consumption dynamics of other nodes. This trend persists or increases as packet count progresses, because these three counterpart schemes do not search for alternative routes, and keep using the same route while exhausting the remaining power of the nodes en route. On the other hand, our proposed *PARTeR* scheme does not cause any nodes with negative power because it finds detouring paths dynamically as consecutive packet transmissions continue. This means that in real battery-powered scenarios, packet transmission for *WiFi*, *Naive*, and *PORTeR* will be terminated as soon as one of the nodes en route runs out of power, whereas our proposed *PARTeR* scheme will continue packet transmissions without any node failures due to energy depletion.

Next, in Figure 3(b), we evaluate the path length for each of the algorithms when transmitting consecutive packets. All algorithms but *PARTeR* maintain the same path length over different packet transmissions, because they always select nodes from the shortest path. *WiFi*, *Naive* and *PORTeR* cause nodes to run out of power by the 15th packet transmission, at which point packet transmission is terminated. *Sensor* maintains an uninterrupted path of 24 relay nodes over continuous packet transmissions, but it sacrifices packet deadline requirements. *PARTeR*, however, manages to select an alternative route after nodes on the shortest path run out of power. This results in increase of 5 nodes of the route length. It also translates in increase of transmission power consumption, as indicated by Figure 3(c). At the cost of more power,

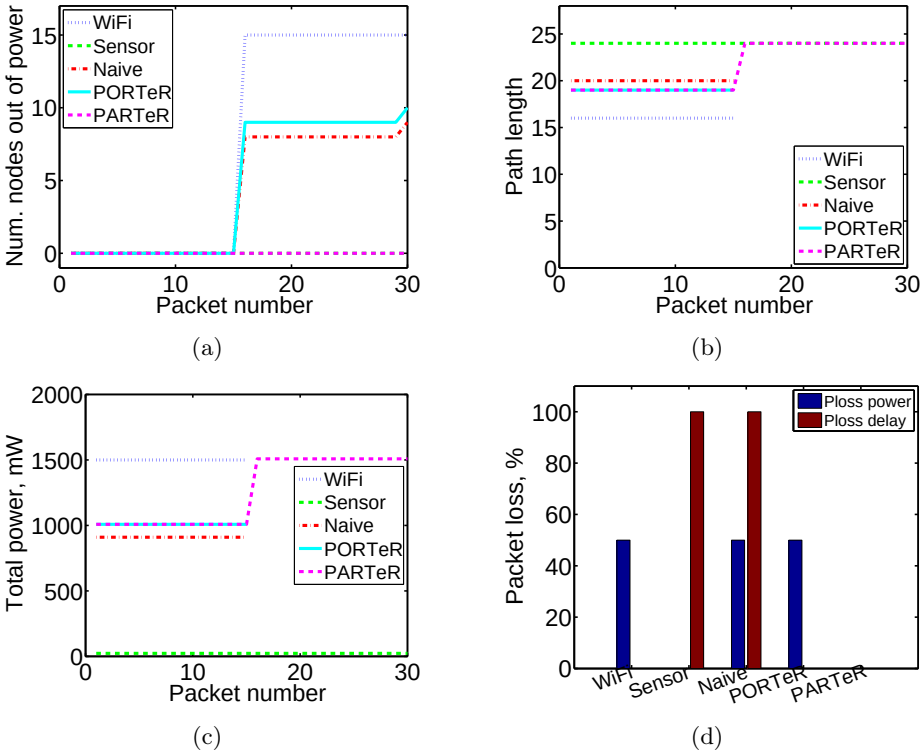


Figure 3: **Performance over consecutive packet transmissions:** (a) Number of nodes out of power; (b) Path length; (c) Total TX power; (d) Packet loss.

however, *PARTeR* is capable to successfully transmit longer packet sequences.

Finally, we summarize results for packet losses incurred by each algorithm. We differentiate between two types of losses: (i) such that were caused by nodes along the route running out of power and (ii) such that were caused by a missed deadline. Figure 3(d) presents for each algorithm packet loss rate for 30 consecutive packet transmissions. *WiFi* and *PORTeR* both have 0% loss due to missed deadline, but they both lose 50% of the packets due to nodes running out of power along the route. *Sensor* is most power conserving, and accordingly incurs 0% packet loss due to power outage. However, it causes 100% packet loss due to missed deadline. *Naive* causes 50% packet loss due to power outage and 100% packet loss due to missed packet deadline. In contrast, *PARTeR* successfully transmits all 30 packets.

6 Conclusion

Design of energy-efficient technologies is crucial both for reducing the footprint of technology on the environment as well as for successful adoption of technology in areas with limited availability of electricity. In line with this need, we design an energy-efficient route selection scheme for heterogeneous sensor networks that leverages knowledge for permitted packet delivery delay and finds routes with minimal power consumption. Each node in these networks is equipped with multiple interfaces that take different amount of time and power for a single packet transmission.

In this paper, we outline three schemes for route selection in such heterogeneous networks and provide evaluation that brings out important trade-offs that need to be considered in the design of such routing schemes. One crucial design principle to guarantee timely packet delivery (i.e., no packet loss due to missed deadline) is that the most conservatively lowest-power interface needs

to be selected by taking into account the expected remaining hops to destination in a long-term view. Also, it is important that a routing algorithm considers the power availability of nodes along the routing path to ensure no packet loss due to power outages by adaptively taking alternative (detouring) paths. Such power-aware algorithm performs more computationally-intensive search for alternative routes, but returns back with increased network lifetime in comparison to shortest path-based counterparts.

We identify some limitations of our proposed schemes. While they are efficient in providing timely packet delivery at minimal power cost, as packet transmission progresses, they become unfair. The reason for this is that our proposed schemes favor high power transmissions in the beginning of the routes, in order to assure timely delivery, while consistently allowing nodes later in the path to use low power transmissions. With increasing number of packets between the same source and destination, the power charge of part of the nodes is exhausted, which ultimately results in network disconnection. To address this problem, as future work, we plan to design a more advanced route selection scheme that diversifies interface selection along the path with fairness, while still meeting the packet delivery deadline.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2013R1A1A1009854, NRF-2013R1A2A2A04014772).

Bibliography

[1] D. S. J. De Couto, D. Aguayo, J. Bicket, R. Morris (2003), A high-throughput path metric for multi-hop wireless routing, In *ACM MobiCom*, 1-13.

[2] R. Draves, J. Padhye, B. Zill (2004), Routing in multi-radio, multi-hop wireless mesh networks, In *ACM MobiCom*, 114–128.

[3] Gartner Inc., *Gartner Symposium/ITxpo*, April 2007.

[4] A. Goldsmith (2005), *Wireless Communications*, Cambridge University Press, New York, NY, USA.

[5] R. L. Graham (1996), Bounds on multiprocessing timing anomalies, *SIAM Journal on Applied Mathematics*, 17(2):416–429.

[6] H. Halabian, I. Lambadaris, C.-H. Lung, A. Srinivasan (2010), Throughput-optimal relay selection in multiuser cooperative relaying networks, In *MILCOM*, IEEE, 507–512.

[7] S. Han, T. Li, C. Qian, D. Leith, A. K. Mok, S. S. Lam (2011), HartFi: an energy-efficient localization system, In *GreenNets, ACM SIGCOMM workshop on Green networking.*

[8] H. Kwon, H. Lee, J. M. Cioffi (2009), Cooperative strategy by stackelberg games under energy constraint in multi-hop relay networks, In *IEEE GLOBECOM*, 1–6.

[9] P. Kyasanur, N. H. Vaidya (2006), Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks, *ACM SIGMOBILE Mobile Computing and Communications Review*, 10(1):31–43.

[10] H. Lee, M. Wicke, B. Kusy, O. Gnawali, L. Guibas, Data stashing: Energy-efficient information delivery to mobile sinks through trajectory prediction, In *ACM/IEEE IPSN*, 2010.

[11] H. Lin, M. Lu, N. Milosavljevic, J. Gao, L. J. Guibas (2008), Composable information gradients in wireless sensor networks, In *ACM/IEEE IPSN*, 121–132.

[12] R. Madan, N. Mehta, A. Molisch, J. Zhang (2008), Energy-efficient cooperative relaying over fading channels with simple relay selection, *Wireless Communications, IEEE Transactions on*, 7(8):3013–3025.

[13] A. Miu, H. Balakrishnan, C. E. Koksal (2005), Improving loss resilience with multi-radio diversity in wireless networks, In *ACM MobiCom*, 16–30.

[14] M.-G. Peng, J. Zhang, W.-B. Wang, H.-H. Chen (2009), Heterogeneous cooperative relay selection with maximal-ratio combining for multi-radio access networks, *International Journal of Communication Systems*, 23(6-7):732–750.

[15] K. Seada, M. Zuniga, A. Helmy, B. Krishnamachari (2004), Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks, In *ACM SenSys*, 108–121.

[16] B. Wang, Z. Han, K. R. Liu (2007), Distributed relay selection and power control for multiuser cooperative communication networks using buyer/seller game, In *IEEE INFOCOM*, pages 544–552.

[17] Y. Wei, F. Yu, M. Song (2010), Distributed optimal relay selection in wireless cooperative networks with finite-state markov channels, *IEEE Transactions on Vehicular Technology*, 59(5):2149–2158.

[18] M. Younis, M. Youssef, K. Arisha (2002), Energy-aware routing in cluster-based sensor networks, In *IEEE MASCOTS*,129–136.

[19] J. Zhang, Q. Zhang (2009), Stackelberg game for utility-based cooperative cognitive radio networks, In *ACM MobiHoc*, 23–32.

Author index