Original Research Paper

The Effect of Packet Redundancy Elimination Technique in Sensor Networks

Deepa Priya and Sharmini Enoch

[1]*Department of Computer Science and Engineering, Noorul Islam University, Kumaracoil, Tamilnadu, India*
[2]*Department of Electronics and Communication Engineering,*
Noorul Islam University, Kumaracoil, Tamilnadu, India

Corresponding Author:
B.S. Deepa Priya
Department of Computer
Science and Engineering,
Noorul Islam University,
Kumaracoil, Tamilnadu, India
Email: deepapratheeb@gmail.com

Abstract: A lot of sensor nodes which are able to sense, process and communicate with external base station with the data obtained from external environment belong to a class of adhoc networks called wireless sensor networks. The main challenge in sensor network is to increase the life time of the network thereby reducing duplicate by detecting the redundant information. While the communication energy-efficiency is necessary to increase the lifetime of the sensor network, an important implementation is to reduce duplicate packets which are considered as a serious issue in these networks. Here the network is hierarchical. The nodes which are intermediate collects information from the source node. These intermediate nodes have their own as well as additional information about the sensed data .This causes redundancy. The redundancy propagates further to the nodes above the network. The main purpose is to detect the packet level redundancy and then to eliminate it. As the nodes are continuously transferring data from sender to receiver and also the sensor nodes consume lot of energy for transmission the data redundancy must be avoided sometimes in order to avoid the loss of needed packets for some critical applications. The packet redundancy elimination hashing algorithm by Rabin Karp is to focus on the identification and elimination of the packet level redundancy with less energy consumption thereby receiving the data with reduction of duplicity. The performance analysis is on energy level of the network and on packet delivery. The level of energy possessed by the nodes in varying time period is noted. The energy levels for existing and proposed methods are compared. The comparison is made with time and energy level of the nodes. The bandwidth of the network is also compared and improved bandwidth is up to 65% in sensor networks when compared with conventional networks.

Keywords: Packet Redundancy Elimination, Locality Sensitive Hashing, Data-Centric Storage, Fingerprinting, Duplicity

Introduction

A network which consists of a number of bulky sensor nodes that are distributed over a region in an adhoc fashion is known as Wireless Sensor Networks (WSN's). In order to compute any corporal circumstances or to calculate any temperature, light intensity, humidity, noise level, locality hundreds of sensors are spread out in an external or internal environment. The attractive feature of all these bulky node of sensors lies in their self-organization and interconnected to form a network. Anand *et al*. (2009) stated that two-way processing can be done by these sensors in order to finish the work which may not done by them individually if they are interconnected together. More number of sensor nodes identify a single target of interest simultaneously and highly interrelated and redundant information is collected. More power will be worn out and due to this group or network force is distributed at a faster rate if each sensor node sends data to the base station. Wide Area Networks have many techniques developed in order to eliminate redundancy across packets. Data de-duplication is the name of the Packet Redundancy Elimination (PRE) which is a packet

data lessening method and a plagiaristic of data density. Reducing the size of the organizer by eliminating duplicate packet data limited within an object is by compressing the packet data. Identifying and eliminating the spreading of duplicate content from both objects within and outside is by DRE. Zohar *et al*. (2010) stated that, to diminish the quantity of packet data to be moved or accumulate remained as the whole folder and information chunk this technique is used .To get better performance over WAN's to the application and to diminish the relief delay and bandwidth utilization data redundancy elimination has been developed. Zohar *et al*. (2011) Packet-level data duplicity removal is implemented thereby identifying as well as eliminating duplicity within chunks from corner to corner in the packets thereby dropping the data quantity to be moved in a network.

During event detection the sensor nodes within a range becomes active and sudden traffic burst occurs towards the sink. Akyildiz *et al*. (2002) has shown that buffer overflow can occur due to traffic burst at the nodes which leads to energy wastage at nodes. Necessity of communication with high efficiency in energy level is needed in order to boost the duration of the sensor network. This work is mainly focused on hop-by-hop system. This method needs the monitoring of storage space overflow with the help of each and every node along the communication path. The intermediate nodes along the path stores the packet records in its remembrance and this technique is more efficient in terms of energy energy efficient in order to collect information and to eradicate the duplicity in WSNs. The existing data redundancy elimination is on WiFi, WLAN's. The objective of the technique using Rabin Karp hashing algorithm is to take full advantage of the duration of network with redundancy removal by increasing reliability, reducing delay and reducing energy consumption. Here the division of sensor networks is into three levels. The tree structure is hierarchical and is divided into three levels. 1. Level1- Data Collection nodes 2. Level2- Nearby sink nodes 3. Level3- Gateway (Final Sink) node. From the source node the nodes which are intermediate collect the records. These intermediate nodes have their own records about the sensed data as well as additional records from the source node. These leads to redundancy and this type of redundancy have to be reduced. The protocol used for transmission is sensor TCP.

Related Works

In this section, a few vital and already available packet duplication removal methods are conferred in sensor networks and also in various other networks and their original supposition and problems are measured here.

Aktas *et al*. (2010) stated a content-type based selective DRE (SDRE) method is proposed thereby applying data duplication elimination discriminately on the contents to diminish the requirement of supply in DRE algorithms where there is more possibility and opportunity to identify the redundant contents. Traffic traces are evaluated in thereby organizing SDRE on smart phone and results obtained shows, SDRE and standard DRE achieves approximately at most equal bandwidth gain with fewer calculation and tiny memory size. The drawback considered here is high computation time is needed for selecting the traffic based on contents when comparing this with standard redundancy elimination method even though the resource requirement is lessened.

Graph-based duplicity deletion algorithm is proposed by Aktas *et al*. (2010) to perceive and decide redundancies automatically in multiple cross-layer interactions. A theoretical graph-based explanation of the crisis provided by building the method suitable for modular systems or networking scenarios at a wider area. Based on that a general algorithm to automatically detect duplicate element concerto is suggested. Li *et al*. (2003) stated a method detects the redundancy in different node compositions, node services and its interactions based on the behavior and equality in modules. It also improves energy efficiency and saves battery power. Mandagere *et al*. (2008) shows a Support Vector Machine (SVM) based duplicity eradication with data aggregation in Wireless Sensor Networks. At earlier for the given network size an aggregation tree is built. To eliminate the redundant data SVM method is applied on the tree. A hashing technique known as Locality Sensitive Hashing (LSH) is used in this study to decrease the duplicity of data and to eradicate the bogus data based on similarity. The LSH code is generated during each session, on the latest data readings of sensor nodes. The codes of LSH which is transferred to the superior data collection nodes, maintains redundancy count for similar LSH code. The superior data collection nodes detects sensor nodes having similar records and picks up only one sensor node amongst them to send real data. Aggregation supervisor also eliminates the outliers and it did not accept the data sent from any other, other than selected node. The profit of this technique is that it lessens the duplicity and eliminates fake records, thus improving the overall performance of the WSN. SVM technique is then practised on the tree to eradicate the duplicate record. To eradicate the data duplicity Locality Sensitive Hashing (LSH) is used here during each session, the LSH code is generated on the latest data readings of sensor nodes. The drawback of this method is that if the code calculation process is mistaken the whole process will be stopped thus reducing the data rate of the process. A data aggregation algorithm Redundancy Elimination for Accurate Data Aggregation (READA) is suggested here. A combining and compressing technique is applied in

the network by READA to eradicate redundant record. The duplicate record is removed from the collected set of records to be sent to the sink node without any big lose in its exactness of the finally collected record by utilizing the series of spatial correlations of record. READA adopts strategy called combining and compressing in order to create a compressed but exact collection in records. Since the set of records intellected shows increase in spatial correlation, READA divides them into sets. A set consists of an id for the group, the shrinked value and the amount of nodes joining here to get the value compressed. The determination of pivot which is the set id is done by the sink. Scarcity of sensor nodes lies in their capability of memory usage. Due to this when there is surplus amount of sets, the values in the two sets are combined. Based on the proximity of the collected records in the two sets the value merging is done. Halepovic *et al*. (2011) stated that determination of the group eviction and maintenance of the set retention is by executing a weighted average. The drawback of this method is in Spring and Wetherall (2000) stated a new last-to last Traffic Redundancy Elimination (TRE) method called PACK for customers in cloud computing. It is based on a new technique of TRE, which allows the customer to use the portion of data's which are at present and to recognize previously received data records. TRE is used to reduce the transfer of duplicate records and infact considerably reduces the cost of sensor network. The sender and the receiver checks up and match up to the names of data chunks in almost all of the Traffic Redundancy Elimination techniques. Before transmission the records are parsed based on the record contents. The chunks, when are found with duplication, the dispatcher restores the transfer of each repeated chunk with its strong signature. Traffic traces from various sources gets the advantage of PACK mainly meant for cloud computing. The drawback of this method is redundancy from the same server but with different protocol. He *et al*. (2014) stated a technique used to remove redundancy in web traffic and for web performance improvement by assuming that there are no collisions. Here the chunking approach is based on content of the transmission. The distribution of traffic is non uniform and the computational overhead is also high. The traffic type is dynamic and streaming media. A fraction of fingerprint values are selected based on their values and not on location. The same content that is packetized in different ways thereby interspersing with other content is also considered as repetitive data by using this technique. The equivalent region in the incoming data chunk is compared with the cached packet to observe that there occurs any collision. Then the equivalent region is extended byte by byte to find the largest matching region in each packet**.** The total repeated content is the union of all largest matching regions. The drawback of this method is only intended for resource constrained hosts.

Prim (2014) stated a PACK chunking approach which is content based and the interfingerprint distance is non uniform distribution. It is a universal TRE (Traffic redundancy Elimination) mechanism, which is implemented for web, mail and video streaming. It is receiver oriented and end-to-end transport level approach in which most of the computation is on receiver side TRE. PACK operates at the transport layer and is designed as a TCP extension, all applications built over TCP. The advantage of this method is that it has a low latency and low overhead. The major drawback of PACK chunking is faulty node cannot be detected and eliminated. Yan *et al*. (2012) has shown a Redundant Traffic Elimination (RTE) a protocol-independent technique for finding and eradicating duplicate chunks of data by scanning it byte by byte from network-layer. The packets are assumed to traverse on a constrained link or path. Here a lively sampling algorithm for finding out duplicate content which works on heterogeneous traffic is proposed. The parameters are self-configuring and no precomputation of traffic is needed. The byte value for data is set in the lookup table and this byte of data is assumed as boundary marker for the corresponding chunk and a hash code will be generated for that chunk. The dynamic lookup table makes the computation of chunk boundary easier at desired sampling rate.

Network Model for Proposed Duplicate Packet Elimination

Using In-network storage is followed as the existing way of avoiding redundancy is by the application of Data Centric Storage Schemes. This scheme is suitable to data that is neither very matured and existing data's and those are also considered to be query nor present at the sensor node which is used for measure. Proficient data access is done in network storage by target nodes and sensor nodes. Open resolution in this class includes Data-Centric Storage (DCS) plans. These plans care for each record as a row of attributes and accumulates record or meta-data about table entries that lie within given column series at low collected set of sensor nodes.

The method using hash code generation is compared with the method for conventional networks without packet chunking, in order to avoid redundant packets reaching the head node and the algorithm is shown as below.

Algorithm for Redundant Data Elimination using hashing technique:
Start
Build the tree node structure hierarchically
Divide the incoming packets into chunks using Rabin Hashing
Generate the representative chunks for each source generating packets
Compare the representative chunks with received packets for redundancy
If redundancy exists
Do

{   Encode the redundant packet with fixed size metadata

Shrink the packet

Decode the redundant packet at receiver end

Retrieve the original data

}

Repeat the process for all redundant packets incoming.

Example

Else

Allow the normal packets at other end

In the proposed packet redundancy removal technique as shown in Fig. 1 as block diagram the tree structure is build hierarchically and the incoming packets are divided into small fixed chunks. Rabin Hash function is to identify redundant chunks which are used for multiple pattern matching. For a S byte packet payload with a sequence of bytes [t1, t2….tn] within S, where (S>=n), a total of S-n+1 fingerprints/patterns are calculated per packet. A set of representative chunks /fingerprint values are found by hash function. These fingerprint table hash values are matched with the incoming packets in the form of chunks. If value of incoming chunks matches with the fingerprint table hash values the duplicate records can be identified at a particular position. The length of pattern is same as that of length of text substring to be matched. Rabin Karp uses hash() to search whether the pattern is in incoming packet or not. Hash function is needed to efficiently calculate the hash values of all substrings for the given text. Initial hash h:d^(m-1)%q q is a prime number, m:length of the pattern, d:no.of characters in alphabet. Hash of next window:

hash (txt[s+1…s+m]) =d(hash(txt[s…s+m-1])-txt[s]*h)+txt[s+m])mod q. For each of the source generating packets generate a set of representative chunks and then compare the set of representative chunks with each receiving packets for redundancy. If redundancy exists while comparison, fixed size metadata is used to encode the redundant packets. After encoding the packets is shrinked and send to the receiver end. At the receiving node the packet is decoded and original data is received. Thus the entire incoming packets are checked for redundancy by repeating the same process above. Else if not, the normal packets are sending to the destination or text = ABDCB and pat = DC q = 11, N = 5,P = 2,d = General value for alphabets Initial hash h: 256^P-1%11=3 Hash of pat = (d*hash value of pat+68) %11=2 Hash of txt= (d*hash value of txt+65) %11 = 10 In the next iteration pat = 7 and txt = 8 If pat hash value not matches with hash (text) there is no redundancy Continue the process of matching by sliding the window.
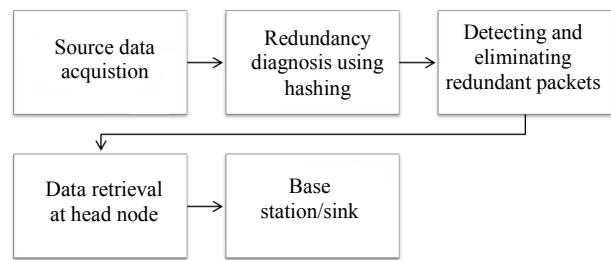

Fig. 1: PRE process block diagram

Results and Discussion

The simulation area of the network is 400×400.The number of nodes for simulation is taken to be fifteen. The entire amount of packets within the drop tail or priority queue is 50.

Deficient resources, energy supply are the two concepts of wireless sensor networks which often faces. To enhance the lifetime of sensor networks, it is vital to put away small amount of energy for the transfer of sensed data. The simulation results show that this scheme valuably eradicates quantity of traffic and energy usage with insignificant improvement of performance cost.

Energy Consumed by Network

The level of energy the nodes in the network, represents its energy consumption within the network. In any simulation the level of energy the node possesses is the initial energy value at the beginning and this initial energy value is passed as an input argument for the starting simulation. In similar way a particular amount of energy will be lost for every packet transmission and every packet reception by a node in the network. So there is a reduction in the preliminary energy of a node. Determination of the energy consumption of a network is by finding the difference between its initial energy value and current energy value. In the simulated work the number of nodes in the environment is 15. From Fig. 2 the initial energy of the network for the existing method without applying chunk selection and PRE methods are assumed to be same as 89.6mJ.

The Table 1 shows the values obtained finally by the network with varying time period. For 1ms as initial time period and with constant increase in time and the number of nodes, the energy consumption of network varies. As simulation proceeds, for a time period of 1ms, 2ms up to 5ms the level of energy for the existing sensor network will be decreased than that of the one which is proposed as redundancy removed sensor network by applying PRE technique, which indicates that the transmission of redundant packets within the network consumes less energy than the other which is proposed. The Table 1 indicates the energy consumed by the total no. of nodes within the network before applying PRE and after applying PRE.

The Fig. 3 shows the final energy level obtained by the network. The values taken from the Table 1 is plotted separately for a time period of 1, 2, 3, 4 and 5ms as time taken for transmission. The energy level of the network without chunking upto the time period of 5 ms is plotted in the graph. The level of energy in the network is decreasing gradually from 69 to 63 ms after the PRE technique as shown in Fig. 3 as the final energy level obtained by the network.

Redundancy Reduction

The amount of successful data transfer within the network represents the reduction of duplicated packets in sensors. The successful transfer of packets is determined by the bandwidth rate or throughput of the network. When the number of nodes increases, the time of transmission increases and the packets to be transferred to the sink increases with decrease in duplicate packets using the packet redundancy reduction technique. The throughput of the network is calculated with 15 nodes.

The time of transmission is increased for each node and its throughput is calculated for varying time period.

Valid Packet Delivery Ratio

Ratio of valid packets without redundancy delivered to the sink represents the number of successfully received packets by the total number of transmissions within the network by the nodes.

Packets received with redundancy = Total number of transmissions-No. of unique packets delivered to the sink. Table 2 shows the no of unique packets received for a time period of 0.5, 1, 1.5 etc upto 4.5 milli secs.

Redundancy ratio is calculated by dividing the number of redundant transmissions with the total number of transmission. The power retention ratio after removing duplicate packets is high and is calculated by the ratio between the uniquely transmitted packets with the total number of transmissions. Power restoration ratio here represents the amount of power saved after eliminating redundancy as shown in Fig. 4.

Table 1: Values of energy consumption by the networks

SI No	Time taken for transmission	Energy Level of the network without PRE(mJ)	Energy Level of Network after PRE(mJ)
1	1ms	89.60	89.6
2	2ms	86.00	90.5
3	3ms	82.70	91.7
4	4ms	83.57	95.3
5	5ms	81.50	93.0

Table 2: Table for packet delivery ratio

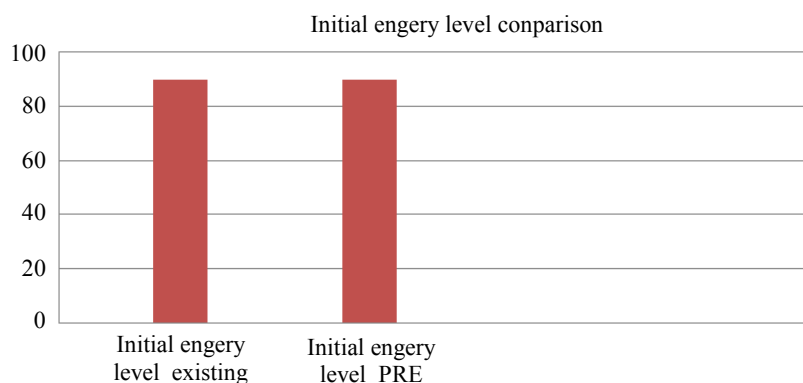| Time in msecs	Total no. of transmissions	No. of unique packets
0	0	0
0.5	32	24
1.0	20	16
1.5	120	64
2	160	80
2.5	130	92
3	360	220
3.5	198	178
4	412	349
4.5	340	280

Fig. 2: Initial energy level of network
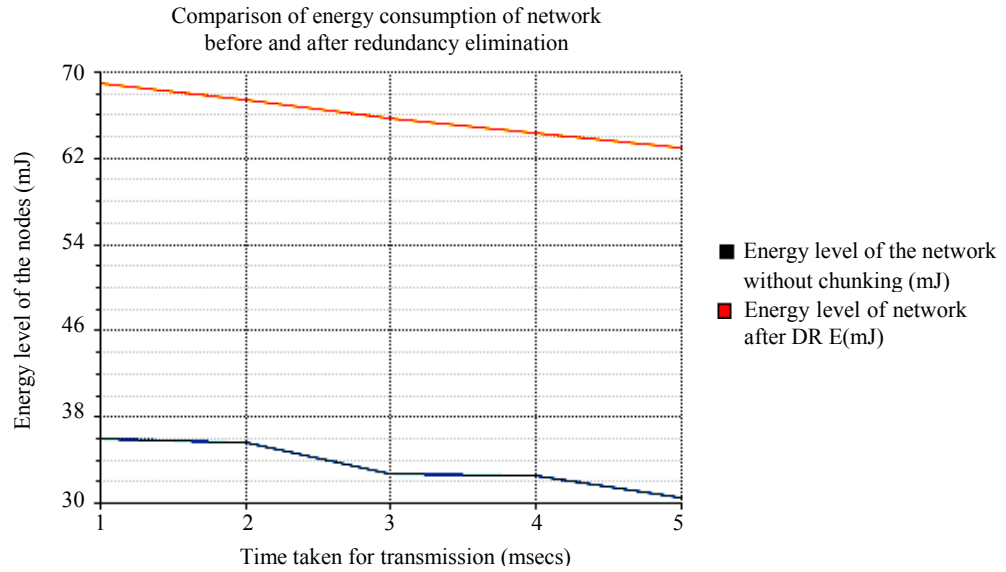
Fig. 3: Final Energy level obtained by the network
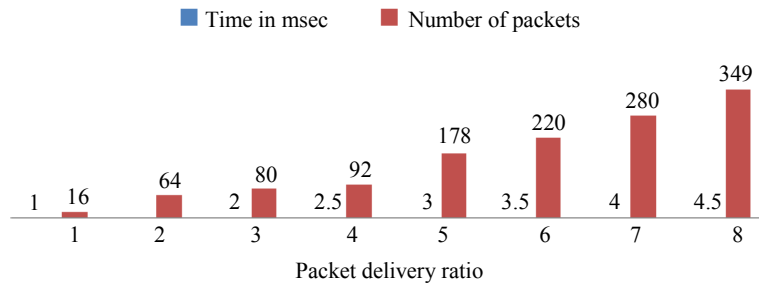

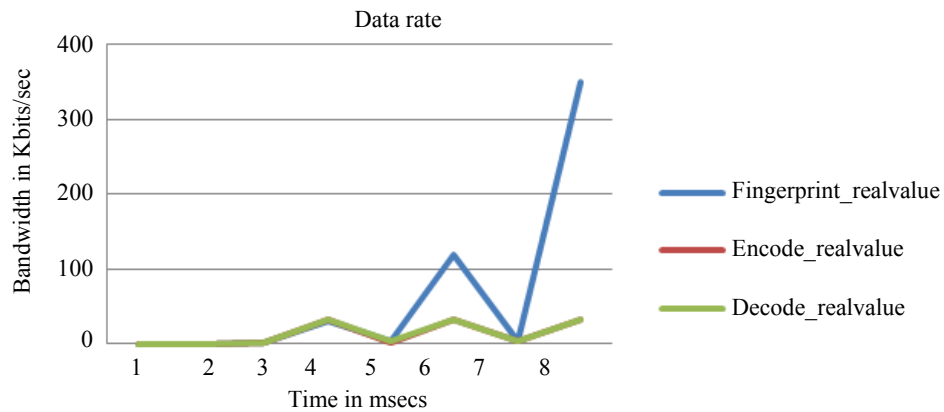**Fig. 4:** Packet delivery ratio of the network

Fig. 5: Bandwidth of the network

Bandwidth Consumption Rate

The rate of bandwidth represents the data rate of the network. The data rate or bandwidth of a network depends on the amount of data taken from the sensors or on the information kind to be transmitted. From Fig. 5 for a period of about 8 msecs the data rate obtained by the network is shown for the network.

The rate of bandwidth after performing redundancy elimination is compared with the existing technique without applying packet chunking or finger printing. The graph above shows the bandwidth comparison with varying time period in milliseconds. From that the proposed technique called fingerprinting have a high bandwidth rate of 300 Kbits per second when compared

with existing technique without fingerprinting as a range of about 25 Kbits per second.

Conclusion and Future Work

The hashing method used here by Rabin Karp hashing algorithm reduces the duplicate data by comparison. Thus redundancy among packets are efficiently reduced packet redundancy elimination method. The results obtained by shows a better performance in energy efficiency, having an improvement in the packet delivery ratio of the network. The future research direction is to focus on an algorithm with improvement over Rabin Karp. The advanced method focuses to obtain a better time complexity, in order to save the time needed for its execution.

Acknowledgment

Author's Contributions

B.S. Deepa Priya: Conceived of the presented idea and developed the theory and performed the computations.

Dr. Sharmini Enoch: Verified the analytical methods and encouraged the corresponding author to investigate the aspect of modifying the algorithm thereby achieving sound results.She also supervised the findings of the work,discussed the results and contributed to the final manuscript.

Ethics

I testify on behalf of co-author that our article submitted to Journal of Computer Science has no ethical issues.All authors have been personally and actively involved in substantive work leading to the manuscript and will hold themselves jointly and individually responsible for its content.

References

Aktas, I., M. Henze, M.H. Alizai, K. Möllering and K. Wehrle *et al.*, 2010. READA: Redundancy elimination for accurate dataaggregation in wireless sensor networks. Scientific Res., 2: 300-308. DOI: 10.4236/wsn.2010.24041

Akyildiz, F., W. Su, Y. Sankarasubramaniam and E. Cayirci, 2002. Wireless sensor networks: A survey. El-sevier Comput. Networks, 38: 393-422. DOI: 10.1016/S1389-1286(01)00302-4

Anand, A., C. Muthukrishnan, A. Akella and R. Ramjee, 2009. Redundancy in network traffic: Findings and implications. Proceedings of the 11th International Joint Conference on Measurement and Modeling of Computer Systems, (MCS' 09), pp: 37-48.

Halepovic, E., C. Williamson and M. Ghaderi, 2011. DYNABYTE: A dynamic sampling algorithm for redundant content detection. Proceedings of the 20th International Conference on Computer Communications and Networks, Jul. 31-Aug. 4, IEEE Xplore Press, Maui. DOI: 10.1109/ICCCN.2011.6005799

He, Q., Z. Li and X. Zhang, 2014. Data deduplication techniques. Proceedigns of the International Conference on Future Information Technology and Management Engineering, Oct. 9-10, IEEE Xplore Press, Changzhou, pp: 430-433. DOI: 10.1109/FITME.2010.5656539

Li, N., J. Hou and L. Sha, 2003. Design and analysis of an MST-based topology control algorithm. Proceedings of the Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, IEEE Societies, Mar. 30-April. 3, IEEE Xplore Press, San Francisco, pp: 1702-1712. DOI: 10.1109/INFCOM.2003.1209193

Mandagere, N., P. Zhou, M.A. Smith and S. Uttamchandani, 2008. Demystifying data deduplication. Proceedings of the Middleware Conference Companion, Leuven, Dec. 1-5, Belgium, pp: 12-17. DOI: 10.1145/1462735.1462739

Prim, R.C. , 2014. Shortest connection networks and some generalizations. Bell Syst. Technol. J., 36: 1389-1401. DOI: 10.1002/j.1538-7305.1957.tb01515.x

Spring, N.T. and D. Wetherall, 2000. A protocol-independent technique for eliminating redundant network traffic. Proceedings of the Conference on Applications, Technologies, Architectures and Protocols for Computer Communication, Aug. 28-Sept. 01, Stockholm, pp: 87-95. DOI: 10.1145/347059.347408

Yan, Z., N. Ansari, M. Wu and H. Yu, 2012. SDRE: Selective data redundancy elimination for resource constrained hosts. Proceedings of the IEEE Global Communications Conference, Dec. 3-7, IEEE Xplore Press, Anaheim. DOI: 10.1109/GLOCOM.2012.6503387

Zohar, E., I. Cidon and O. Mokryn, 2011. The power of prediction: Cloud bandwidth and cost reduction. Proceedings of the ACM SIGCOMM 2011 Conference, Aug. 15-19, Toronto, pp: 86-97. DOI: 10.1145/2018436.2018447

Zohar, E., O.O. Mokryn and I. Cidon, 2010. PACK: Speculative TCP traffic redundancy elimination. Tech. Rep. CCIT Report #770.