ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

A Novel Approach of Data Compression in Wireless Sensor Networks

Beema K.S , Mitha Rachel Jose

M.Tech Student, Dept. of CSE Mahatma Gandhi University, Mangalam College of Engineering, Kottayam, India

Assistant Professor, Dept. of CSE , Mahatma Gandhi University, Mangalam College of Engineering, Kottayam, India

ABSTRACT: Data provenance organization for sensor networks introduces quite a few demanding requirements, such as low energy and bandwidth consumption, competent storage and safe transmission. So data provenance for these networks as compact as possible. If the size of the provenance increases with the number of nodes traversed by the network packets. In order to reduce the measure of sensing data, necessitate compressing them inside the network. So we can compress the data by a dictionary based provenance scheme. The major objective of data aggregation is to bring together and aggregate data in an energy efficient way so that network lifetime is enhanced. Also we can introduce a threshold value to the sensor nodes and the set of nodes are waiting until the threshold values meted. Also introduce a congestion control mechanism. So the total time to be taken to travelled from sink to base station is reduced. Trustworthiness of sensor data is also assured through an AM-FM sketch, it can defend against most of the known provenance attacks. Reduce the time complexity and security overhead using pipelined Hash Tables, pipeline be able to attain a high throughput via the interstage parallel access to hash tables**.**

KEYWORDS: Provenance, Hash Tables, Dictionary based compression, Aggregation Nodes.

I. INTRODUCTION

Wireless sensor network have a variety of applications like environmental monitoring, building monitoring, health monitoring, military surveillance and target tracking and the data they collect are used in decision-making for critical infrastructures. Wireless sensor network is a resource restraint if we consider energy, computation, memory and limited communication capabilities. Every sensor nodes in the wireless sensor network are interacting with every other node or by intermediate sensor nodes. Data are streamed from numerous sources through in-between processing nodes that collective information. A hateful adversary may introduce extra nodes in the network or negotiation existing ones. Therefore, assuring high data trustworthiness [2] [5] is vital for right decision-making.

Plummeting the amount of the provenance is crucial in WSN as it is composed of a numerous sensor nodes. The restriction of provenance in WSN is stretched storage, deficient energy and increased bandwidth utilization of the sensor node. Also sensors operate in an malicious environment, where they may be focus to attacks. Provenance is l deals with the detecting malicious node in network and to notice the packet drop in network.

In this paper, we propose a novel approach for compressing the data and lightweight scheme to securely transmit provenance for sensor data through an AM-FM sketch [8]. Wireless sensor networks (WSN) offer less power for processing rather than transmitting data. It is preferable to do in network dispensation within network and diminish packet dimension. One such approach is data aggregation [3] which attractive technique of data gathering in distributed scheme architectures and lively access by the use of wireless connectivity. Wireless sensor networks have limited computational power and limited memory and battery power, this leads to increased intricacy for application developers and often results in applications that are closely coupled with network protocols.

Data aggregation structure on wireless sensor networks is a nodes that generates data, based on its sensing mechanisms surveillance and broadcast sensed data packet to the base station (sink). This procedure on the whole straight broadcast since the base station may place extremely far away from sensor nodes needs. Additional energy to broadcast data over extended distances so that a better method is to have less nodes send data to the base station. These nodes called aggregator nodes and processes is called data aggregation in wireless sensor network [3][4][6] .The configuration mechanism as a middleware for aggregating data measured by a number of nodes inside the network. The

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

data aggregation techniques choose a subset of sensor nodes in the network to be accountable for fusing the sensing data from other sensor nodes to diminish the total data transmissions.

This paper is to intend a high-throughput as well as memory-efficient multimatch packet classification through a Pipelined Hash Tables [9] [10] [11]. The elegant pipeline architecture combines the outcome from single-dimensional searches to find all matched rules. Traversing the provenance graph and store each edge of each partition into an individual hash table. To supplementary partition the hash table at each stage diminish the time complexity and security overhead and can attain a soaring throughput via the interstage parallel right of entry to hash tables extend intrastage parallelism.

II. RELATED WORK

Changda Wang, Syed Rafiul Hussain [1] proposed a dictionary based approach to encode the sensor data provenance to address the drawbacks of lossy compression techniques and to address the limitation of entropy lower bound. Pradhan et al. [12] proposed an outline for distributed compression using joint source and channel coding. This approach minimizes the quantity of inter-node communique for compression using both a quantized source and correlated side information within each individual node.

Hyo Sang Lim, Yang Sae Moon, Elisa Bertino [2] proposed propose a cyclic framework that generates trust scores of data substance from folks of network nodes and trust scores of network nodes from those of data substance. Trust scores are gradually evolved in the cyclic framework. Ian Foster, Jens Vöckler, Michael Wilde, Yong Zhao [13] proposed a virtual data directory (based on a relational virtual data schema) provides a dense and mobile representation of the computational events used to derive data, as well as invocations of those events and the datasets created by those invocations. Virtual data language interpreter executes necessities for constructing and querying database entries.

You-Chiun Wang [14] discusses the various data compression techniques in Wireless sensor networks. Ranganathan Vidhyapriya and Ponnusamy Vanathi [12] proposed an Energy Efficient Data Compression in Wireless Sensor Networks using data compression algorithms incorporated through the shortest path routing method.

Nandini. S. Patil, Prof. P. R. Patil [3] discuss with data aggregation in wireless sensor networks, data aggregation which attractive method of data assembly in distributed system architectures and dynamic right to use via wireless connectivity.Neeshma K K and Renisha P Salim [10] proposed a distributed hash based architecture. This hash tables be helps to get better packet traffic rate and decrease packet loss. The packet classifier performs multimatch packet classification for getting best matched rule. By using hash based asynchronous pipeline architecture helps to get better throughput and memory access rate.

III. PROPOSED SYSTEM

A. *Data Provenance Management:*

Dipping the dimension of the provenance is crucial in WSN as it is composed of a large number of sensor nodes. Provenance contains data and transmission details. The limitation of provenance in WSN is tight storage, limited energy and increased bandwidth consumption of the sensor node.

In order to diminish the quantity of sensing data, we require compressing them inside the network. We can classify the data compression schemes into two categories: lossless, lossy. A lossless compression is after executing the decompression process, we can obtain exactly the same data as those before executing the compression operation. A lossy compression means that some detailed features of data may be lost due to the compression operation. So data provenance for this network is compact as possible. In dictionary based provenance scheme [1], every sensor node in the network stores a packet path dictionary, by the hold up of this dictionary a path index in its place of path itself is enclosed among each packet.

Fig. 1 shows the provenance graph of sensor networks. In fig 1(a) having five nodes, with no compression, paths in such a provenance graphs can be prearranged as <n5,n4, n3, n2, n1> Where ni is the node id. In fig 1(b) having eight nodes and manifold paths are linked as branches in a tree. Dealing with this type of graph use semicolon as a delimiter to split the twigs of a tree, paths in such a provenance graphs can be encoded as <<n6, n3>; <n7, n5>;< n8,n5>;<n5,n3>;<n3,n1>>.

These linear paths and their equivalent indexes are stored in a dictionary as shown in Table 1.If the graph is alinear path $\{n_x,n_{x-1},\ldots\ldots n_2,n_1,\}$, it can be simply represented by the index $< n_x , n_1>$.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

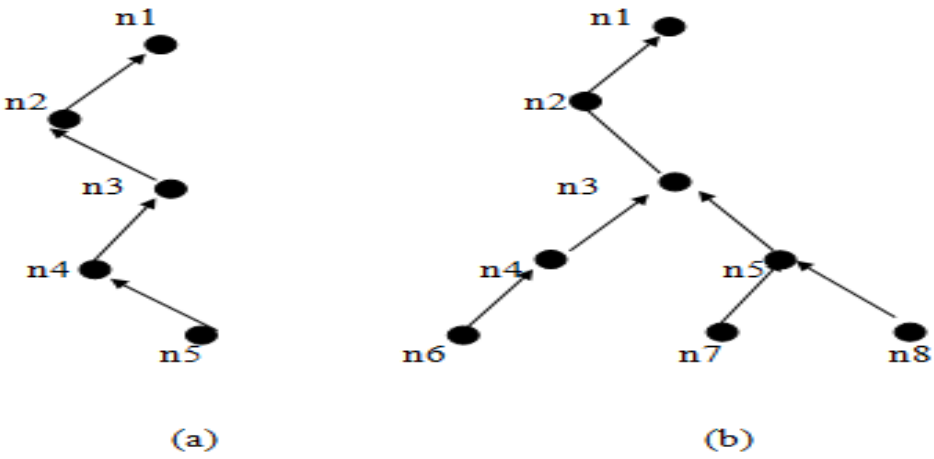

Fig.1 Provenance graph of sensor networks

TABLE 1
Dictionary of Indexes

Linear Path	Index
{n4, n3, n2, n1}	<n4,n1>
{n6,n4,n3}	<n6,n3>
{n3,n2,n1}	<n3,n1>

Intend a high-throughput and memory-efficient multimatch packet classification through a Pipelined Hash Tables [9].Tthe elegant pipeline architecture combines the outcome from single-dimensional searches to find all matched rules. Traversing the provenance graph and store each edge of each partition into an individual hash table. To supplementary partition the hash table at each stage diminish the time complexity and security overhead and can attain a soaring throughput via the interstage parallel right of entry to hash tables extend intrastage parallelism.

B. *Trustworthiness of Data:*

Since the need of data distribution between numerous organizations like governmental organizations, financial corporations and medical hospitals, military purpose it is critical to make sure the data integrity so that effective decisions can be made based on these data. Significant part of any solution for assessing data veracity is represented by techniques and tools to assess the trustworthiness of data provenance. The accessibility of complete data makes it possible to take out more accurate and whole knowledge and thus ropes more knowledgeable decision making. Devoid of high-assurance integrity, information extracted as of available data cannot be trusted.

To make sure the safety of provenance, make use of the AM-FM sketch scheme [8] which binds the packet content and its provenance mutually. The AM-FM sketch is a distributed, node-level digital signature scheme. Using this approach when a sensor node generates or ahead a data packet, it creates the digest of the data and signs this digest prior to sending the packet to the next node.

C. *Data Aggregation:*

The major objective of data aggregation is to gather together and aggregate data in an energy competent way so that network lifetime is enhanced. Wireless sensor networks present more and more sensor nodes need a smaller amount

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

power for processing the sensor data since compared to broadcasting the data. Though the data aggregation results in smaller number of transmissions, if there is a greater delay in the case of some aggregation functions because data from earlier sources may have to be detained rear at an intermediate node in order to be aggregated with data coming from sources that are beyond absent. In the most horrible case, the latency due to aggregation will be relative to the number of hops between the sink and the uttermost source.

The data aggregation method is to select a subset of sensor nodes in the network to be accountable for fusing the sensing data from other sensor nodes to decrease the amount of data transmission. But a set of sensor nodes are wait until the complete set of nodes are reached at the aggregation node. So this is the main disadvantage of aggregation. In this paper we can introduce a threshold value to the sensor nodes and the set of nodes are waiting until the threshold values meted. So the total time to be taken to travelled from sink to base station is reduced. Also there is no congestion control mechanism in the existing data aggregation process. So we can control the congestion in the data aggregation process by setting out the threshold value.

IV. SIMULATION RESULTS

The Dictionary based provenance scheme in wireless sensor networks, every sensor node in the network stores a packet path dictionary, by the hold up of this dictionary a path index in its place of path itself is enclosed among each packet. The major objective of data aggregation is to gather together and aggregate data in an energy competent way so that network lifetime is enhanced.

Fig 2 shows time taken against the total packet sent. Wireless sensor networks present more and more sensor nodes need less power for processing as compared to transmitting data. Though the data aggregation results in smaller number of transmissions. Without any aggregation (represent old in graph) time taken to sent the packet through the network have greater transmission time than with aggregation (represent new in graph). So aggregation results in smaller number of transmissions.

By introduce a threshold value to the sensor nodes and the set of nodes are waiting until the threshold values meted. So the total time to be taken to travelled from sink to base station is reduced. Also we can control the congestion mechanism in the data aggregation process by setting out the threshold value. Fig 3 shows the Effect of applying a maximum limit or threshold value. When the maximum limit increases then time taken for transmission decreases. Also when the maximum limit or threshold value increases the speed of the transmission also increases.

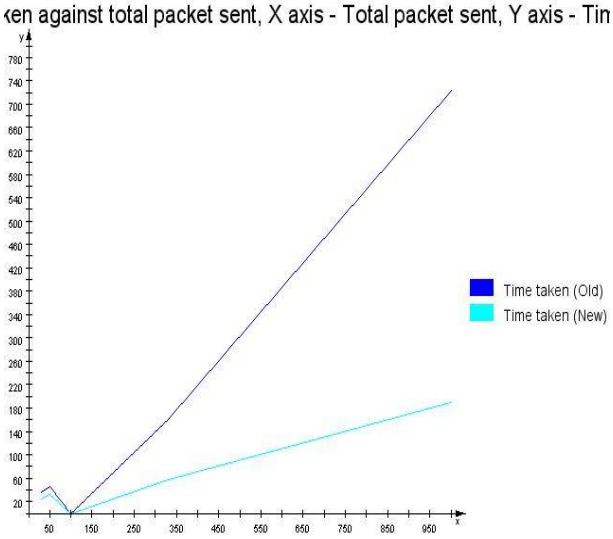

Fig. 2. Time taken against total packet sent

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

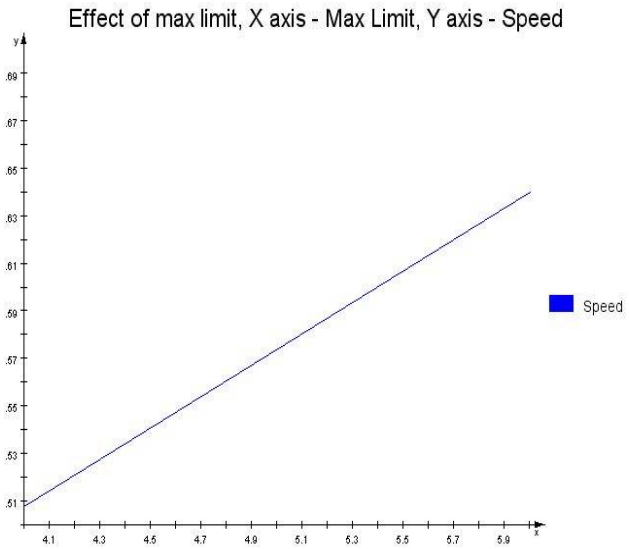

Fig 3. Effect of Maximum Limit

V. CONCLUSION

This paper put forward a dictionary based provenance compression for wireless sensor networks. Using packet path dictionaries encircle path indexes as an alternative of the path itself in the provenance. The size of the compressed provenance in lossless approach is smaller than that of the existing lossy provenance schemes. Data aggregation is to bring together and to aggregate data in an energy well-organized way so that network lifetime is enhanced. Trustworthiness of sensor data is also assured through an AM-FM sketch; it can defend against most of the known provenance attacks. We can introduce a threshold value to the sensor nodes and the set of nodes are waiting until the threshold values meted. So the total time to be taken to travelled from sink to base station is reduced. We can control the congestion in the data aggregation process by setting out the threshold value. Reduce the time complexity and security overhead using pipelined Hash Tables; pipeline is able to attain a high throughput via the interstage parallel access to hash tables. Experimental results demonstrate that our method can put aside more energy and time and show high performance.

REFERENCES

1. Changda Wang, Syed Rafiul Hussain, "Dictionary Based Secure Provenance Compression for Wireless Sensor Networks," in IEEE Transactions on Parallel and Distributed Systems, VOL. 27, NO. 2, Feb 2016.
2. H.S. Lim, Y.-S. Moon, and E. Bertino, "Provenance-based trustworthiness assessment in sensor networks," in Proc. 7th Int. Workshop Data Manage. Sensor Netw., pp. 2–7, 2010.
3. Nandini. S. Patil, Prof. P. R. Patil, "Data Aggregation in Wireless Sensor Network," in IEEE International Conference on Computational Intelligence and Computing Research, 2010.
4. Sumit Chaudhary, Neha Singh, Avinav Pathak,A.K Vatsa, "Energy Efficient Techniques for Data aggregation and collection in WSN ," International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.4, August 2012.
5. S. Sultana, E. Bertino, and M. Shehab, "A provenance based mechanism to identify malicious packet dropping adversaries in sensor networks," in Proc. 31st Int. Conf. Distrib. Comput. Syst. Workshops, pp. 332–338, 2011.
6. Bhaskar Krishnamachari, Deborah Estrin ,Stephen Wicker ," The Impact of Data Aggregation in Wireless Sensor Networks", 2001 ACM SIGCOMM Workshop on Data communication, April 2001.
7. S. Madden, M. J. Franklin, J.M. Hellerstein, and W. Hong, "Tag: A tiny aggregation service for ad-hoc sensor networks," SIGOPS Oper. Syst. Rev., vol. 36, no. SI, pp. 131–146, Dec. 2002.
8. M. Garofalakis, J. M. Hellerstein, P. Maniatis, and IEEE, "Proof sketches: Verifiable in-network aggregation," in Proc. IEEE 23rd Int. Conf. Data Eng., pp. 971–980, 2007.
9. Yang Xu, Member, IEEE, Zhaobo Liu, Zhuoyuan Zhang, and H. Jonathan Chao," High-Throughput and Memory-Efficient Multimatch Packet Classification Based on Distributed and Pipelined Hash Tables", IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 22, NO. 3, JUNE 2014.
10. Neeshma K K ,Renisha P Salim," Multi-match packet Classification Based on Distributed Hsah Table", International Journal of Science, Environment and Technology, Vol. 4, No 4, 1098 – 1106, 2015.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

11. Neeshma K K ,Renisha P Salim ," Perfect Hash Table Construction Using Pipelined Architecture Based on Signature Tree ",AEIJST - July 2015 - Vol 3 - Issue 7 ISSN - 2348 – 6732
12. Ranganathan Vidhyapriya, Ponnusamy Vanathi, "Energy Efficient Data Compression in Wireless Sensor Networks" , The International Arab Journal of Information Technology, Vol. 6, No. 3, July 2009.
13. I. Foster, J. Vockler, M. Wilde, and Y. Zhao, "Chimera: A virtual data system for representing, querying, and automating data derivation," in Proc. 14th Int. Conf. Sci. Statist. Database Manage., pp. 37–46, 2002.
14. You-Chiun Wang ,"Data Compression Techniques in Wireless Sensor Networks ",Pervasive Computing
15. S. Sultana, G. Ghinita, E. Bertino, and M. Shehab, "A light weight secure provenance scheme for wireless sensor networks," in Proc. IEEE 18th Int. Conf. Parallel Distrib. Syst., pp. 101–108, , 2012
16. K.K. Muniswamy-Reddy, D. A. Holland, U. Braun, and M. I. Seltzer, "Provenance-aware storage systems," in Proc. USENIX Annu. Tech. Conf., General Track, pp. 43–56, 2006.
17. S. M. I. Alam and S. Fahmy, "Energy-efficient provenance transmission in large-scale wireless sensor networks," in Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw., pp. 1–6, 2011.

BIOGRAPHY

Beema K.S doing M.Tech in Computer science And Engineering at Mangalam College of Engineering, Mahatma Gandhi university. She receives B.Tech degree in 2013 at Ilahia college of engineering and Technology, under Mahatma Gandhi university. Areas of interest are data mining, security and wireless networks.

Mitha Rachel Jose, Assistant professor in Mangalam college of engineering, under Mahatma Gandhi university. She receives B.Tech from Kerala university in 2008 and M.Tech from Anna university. She has a research experience of one year as research student in Chydenius university Consortium Finland (Jyvaskyla University).