ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 12, December 2017

Data Compression in Dictionary Based secure Provenance for Wireless Sensor Networks

[1] Vishwa Prasath T S, [2] Rajasekar P
[1] M.Tech, Information Security and Cyber Forensics, SRM University, Kattankulathur, Chennai, India
[2] Assistant Professor, (Sr.G), IT Department,SRM University, Kattankulathur, Chennai, India

***Abstract*: Remote sensor systems are spatially dispersed sensor hubs that monitors the physical or ecological conditions like sound, push waves, temperature of encompassing and so on, and furthermore send the detected information back to the sink or base station.WSN are utilized as a part of numerous applications like military regions, debacle administration in remote regions, in building keen urban communities and so forth. In this manner security is an essential angle in WSN. These systems can be inclined to different sad assaults or programmers that have the intention to upset the whole system. In a commonplace WSN organize, the minor hubs introduced on various areas sense the encompassing condition, send the gathered information to their neighbours, which thusly is sent to a sink hub. The sink hub total the information got from various sensors and send it to the base station for additionally handling and vital activities. The data's source, also packet and provenance binds with AM-FM sketch and uses secure packet sequence generation technique. Everything is fine here, but the data is not compressed, hence data compression part will be done using Lempel-Ziv method.**

***Key Words:* Sensor Networks, Data Provenance, Data compression.**

I. INTRODUCTION

Large scale sensor systems are sent in various application areas, including restorative observing, ecological checking, reconnaissance, home security, military operations, mechanical machine checking, and so on. In these application areas, sensors change from smaller than expected, body-worn sensors to outside sensors, for example, camcorders or situating gadgets. The assorted variety of such system situations requires embracing procedures that can guarantee the dependability of information over the system [1]. Since provenance records the historical backdrop of the two information securing and transmission, it is considered as a successful instrument to assess the dependability of information. It additionally gives the data about the operations performed on information [2]. Be that as it may, decreasing the measure of the provenance is pivotal in substantial scale sensor systems.

Sensor hubs in these systems will be unable to record and control vast provenance information because of capacity and computational asset requirements [4]. In this paper, we research the issue of effective and secure pressure of provenance data in remote sensor systems (WSNs) [6]. The issue forces an arrangement of difficulties:

- The pressure of provenance requires to be as conservative as conceivable so that for substantial scale WSNs the provenance measure does not increment with the quantity of hubs crossed by the system packets.
- The pressure or encoding ought to guarantee that the framework does not lose any provenance data subsequent to unravelling.
- The reliability of the provenance must be guaranteed.
- The information is to be compacted utilizing solid calculation; Lempel-Ziv is the calculation which will be actualized.

In a multi-jump sensor organize, information provenance enables the BS to follow the source and sending way of an individual information bundle [7]. Provenance must be recorded for every packet, except vital difficulties emerge because of the tight stockpiling, vitality and transmission capacity requirements of sensor hubs. Thusly, it is important to devise a light-weight provenance arrangement with low overhead. Besides, sensors frequently work in an untrusted situation, where they might be liable to assaults. Thus, it is important to address security necessities, for example, secrecy, respectability and freshness of provenance [9]. Provenance all through the system. Provenance of a data thing can be spoken to as a tree which is inserted as meta-information with the data thing, and refreshed along the way used to forward the thing to the base station. Consequently, every halfway hub conveys provenance of length corresponding to the bounce tally between that hub and the wellspring of the related data thing. In a network[15] with an expansive distance across (jump check), this expanded meta-information length brings about a broadened time of radio correspondence and vitality dissemination at each moderate hub.

2. PROVENANCE

Provenance is defined as origin or source of something that is to be proved, in a network there will be n number of networks that will send data to the sink node, sink node cannot find the source of the data, to overcome this data provenance is applied. Here each node will have an path index in which the data has to be transferred, which will help in tacking the data easily. This process has encoding method and decoding method as well.

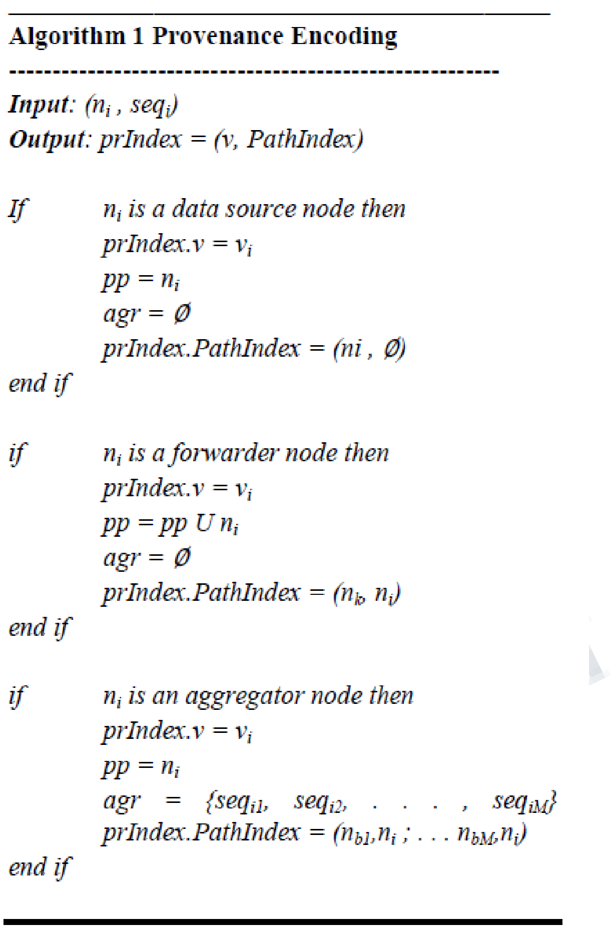
2.1 PROVENANCE SCHEME

A conveyed system to encode provenance at sensor hubs and a unified way to deal with disentangle provenance at the BS. With the help of the PPD at every hub, our proposed approach incorporates a PathIndex in the provenance record rather than the way itself and along these lines packs provenance to a considerably littler size than that of the first one. Likewise, to guarantee the insurance against any unapproved adjustment, the packet and its provenance are bound together by a protected message validation code (MAC). While translating, the BS checks the MAC and after that concentrates the provenance diagram for the packet it gets

2.2 PROVENANCE ENCODING

Our approach encodes the provenance inside the information bundle in a circulated way, and translates it at the base station. Every datum bundle comprises of a succession number, its own particular information, and an in-Bloom channel IBF field containing the provenance. Each sensor hub stores the area of its neighbour hubs that can interface with specifically and in addition the bundle grouping number of the last observed packet for each source. This will serve to identify if any packet has been dropped amid the following round of bundle transmission, and to confine the capable hub. In what tails, we will talk about how provenance is encoded and decoded, too how dropped packets are recognized and the dependable hubs found.

$$Vid_i\text{=generate VID } (n_i, \text{seq}, \text{pSeq}_i)$$
$$=E_{ki}(\text{seq}||\text{pSeq}_i)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$



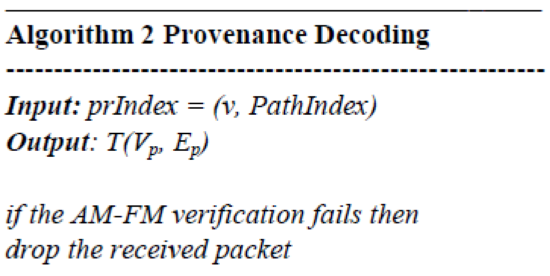
Algorithm 1 Provenance Encoding

```
Input: ($n_i$ , $seq_i$)
Output: prIndex = (v, PathIndex)

If      $n_i$ is a data source node then
        prIndex.v = $v_i$
        pp = $n_i$
        agr = Ø
        prIndex.PathIndex = (ni , Ø)
end if

if      $n_i$ is a forwarder node then
        prIndex.v = $v_i$
        pp = pp U $n_i$
        agr = Ø
        prIndex.PathIndex = ($n_k$, $n_i$)
end if

if      $n_i$ is an aggregator node then
        prIndex.v = $v_i$
        pp = $n_i$
        agr = {$seq_{i1}$, $seq_{i2}$, . . . , $seq_{iM}$}
        prIndex.PathIndex = ($n_{b1}$,$n_i$ ; . . . $n_{bM}$,$n_i$)
end if
```

2.3 PROVENANCE DECODING

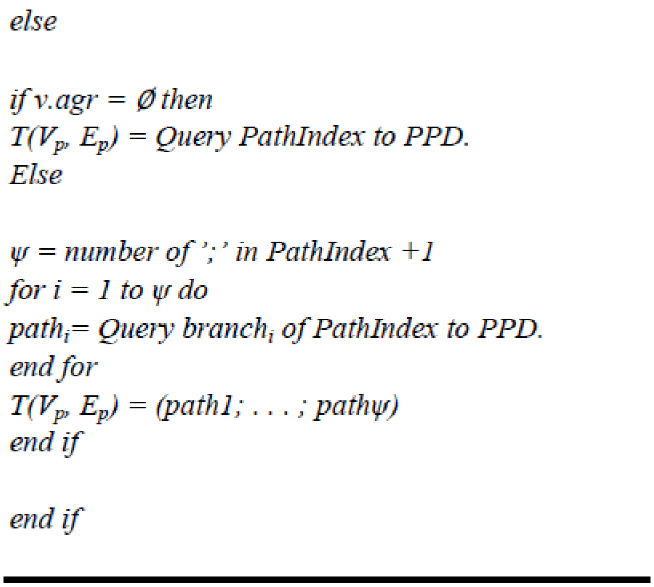
At the point when the BS gets a packet, it confirms the honesty of the ensured information through the AM-FM assessment process. On the off chance that the check affirms that the secured information are reliable, the BS acknowledges the packet, generally, the bundle is dropped. For an acknowledged packet p, the provenance diagram spoke to as T(Vp, Ep) is recovered by looking into its PathIndex in the PPD of the BS. To clarify the unravelling procedure, we again make reference to Fig. 1(c). We expect that hub n1 is the BS in the system and it beforehand got a packet with arrangement number seq1 created by hub n6. Thus, it has all the data of the way from n6 to itself put away in its PPD. Presently expect that hub n6 produces another packet p with grouping number seq4 and the BS gets that from n2. While sending the bundle to the BS, n2 implants the provenance record prIndex = (v2,n6, n2i) in the packet as portrayed. Here, v2 = (n2, seq4, Ø). The BS at that point translates the provenance of p utilizing the PPD put away in it. In this way, when the BS gets prIndex = (v2,n6, n2i) from hub n2, it frames hn6, nli by supplanting n2 with its own ID and gazes upward in the table it has for hub n6 in its PPD.At the point when the BS gets a packet, it confirms the honesty of the ensured information through the AM-FM assessment process. On the off chance that the check affirms that the secured information are reliable, the BS acknowledges the packet, generally, the bundle is dropped. For an acknowledged packet p, the provenance diagram spoke to as T(Vp, Ep) is recovered by looking into its PathIndex in the PPD of the BS. To clarify the unravelling procedure, we again make reference to Fig. 1(c). We expect that hub n1 is the BS in the system and it beforehand got a packet with arrangement number seq1 created by hub n6. Thus, it has all the data of the way from n6 to itself put away in its PPD. Presently expect that hub n6 produces another packet p with grouping number seq4 and the BS gets that from n2. While sending the bundle to the BS, n2 implants the provenance record prIndex = (v2,n6, n2i) in the packet as portrayed. Here, v2 = (n2, seq4, ∅). The BS at that point translates the provenance of p utilizing the PPD put away in it. In this way, when the BS gets prIndex = (v2,n6, n2i) from hub n2, it frames hn6, nli by supplanting n2 with its own ID and gazes upward in the table it has for hub n6 in its PPD.

The pp it recovers from its PPD is {n6, n4, n3, n2, n1}. Along these lines the BS unravels the provenance data and recovers the way of the got bundle utilizing the lexicon PPD put away in it. In the event that the BS gets a bundle with a non-exhaust conglomeration record (agr), i.e., the PathIndex of the packet's provenance record has semicolon(s), the BS finds the aggregator node(s) by finding the semicolons in the PathIndex. At that point every way isolated by the semicolon(s) is considered exclusively and turned upward in the PPD so as to disentangle. The means of provenance interpreting are exhibited in Algorithm 2.



Algorithm 2 Provenance Decoding

```
Input: prIndex = (v, PathIndex)
Output: $T(V_p, E_p)$

if the AM-FM verification fails then
drop the received packet
```

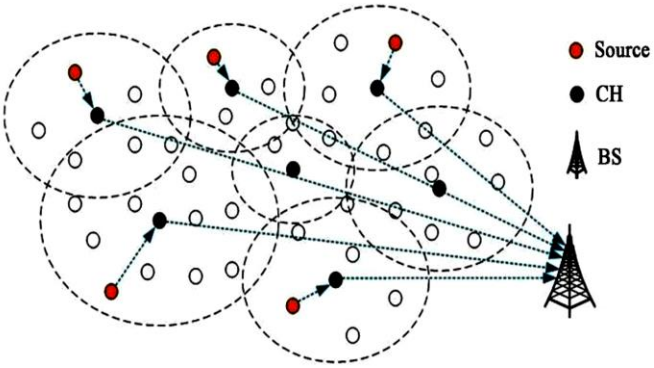


```
else

if $v.agr = \emptyset$ then
$T(V_p, E_p)$ = Query PathIndex to PPD.
Else

$\psi$ = number of ';' in PathIndex +1
for i = 1 to $\psi$ do
$path_i$= Query $branch_i$ of PathIndex to PPD.
end for
$T(V_p, E_p) = (path1; . . . ; path\psi)$
end if

end if
```

3.EXISTING SYSTEM

Existing Systems takes a shot at a probabilistic way to deal with encode the hubs' IDs into the provenance. What's more, a portion of alternate works utilize Bloom channel to encode the IDs of the hubs that are on a packet's way. These methodologies limit the measure of provenance data by keeping just the hub's ID's. Existing Works, for example, lightweight secure provenance conspire in view of in-packet Bloom channel. This approach ties information and its provenance together and furthermore chains the packet succession numbers neighbouring identify provenance falsification and bundle dropping assaults. Nonetheless, the edges which point indirectly the bundle transmissions are disposed of. Consequently, those methodologies are lossy provenance pressure strategies. Just hubs' IDs are recorded in the information provenance. Unavoidable false positive in provenance deciphering, and increment in the provenance estimate with the quantity of hubs navigated so as to keep the false positive rate under a given edge. The word reference based provenance plot which is the most minimal and lossless plans made. Configuration will be productive and appropriated calculation for encoding the provenance data and in addition a concentrated approach for its deciphering. Here they presents a safe packet succession number age component and utilize the AM-FM outline method to secure the provenance likewise play out a formal security examination and a broad execution assessment of our proposed provenance plot.

4. SYSTEM ARCHITECTURE



5. BACKGROUND STUDY

Broad research has been completed on the subject of system provenance. Be that as it may, because of restricted computational capacity, vitality limitations, and low data transmission, these ordinary provenance plans [20], [21] can't be connected in WSNs straightforwardly. Alam et al. propose a probabilistic approach [5] to encode the hubs' IDs into the provenance. Shebaro et al. utilize Bloom channel [6] to encode the IDs of the hubs that are on a packet's way. These methodologies limit the measure of provenance data by keeping just the hubs' IDs. Be that as it may, the edges which points indirectly to the bundle transmissions are disposed of. Subsequently, those methodologies are lossy provenance pressure systems. Salmin et al. [2], [16] propose a lightweight secure provenance conspire in view of in-bundle Bloom channel. This approach ties information and its provenance together and furthermore chains the packet arrangement numbers adjoining distinguish provenance fraud and bundle dropping assaults. Lossy provenance encoding plans [2], [5], [6], [16] share the accompanying attributes: (I) just hubs' IDs are recorded in the information provenance, (ii) unavoidable false positive in provenance deciphering, and (iii) increment in the provenance estimate with the quantity of hubs navigated so as to keep the false positive rate under a given edge. The pressure procedure utilized as a part of this paper depends on the pressure calculations proposed by Ziv et al. [11], [12]. To assemble the word reference they utilize the substrings that seem various circumstances in a message. As cyclic ways are not permitted in sensor organizes, a hub appears at most once on a bundle's way i.e., there is no rehashed hubs on a way. Subsequently, these conventional word reference based calculations can't be connected specifically to provenance pressure. Chen et at. [18] diminish the subtrees recursively into their underlying foundations. Along these lines, the subsequent diagram speaks to a coarse granular perspective of the provenance. By recursively contracting kids hubs into their aggregator hubs, the provenance tree can be spoken to as an arrangement of direct way bits. At that point each of these way bits is packed utilizing PPDs. This work is near our approach. In any case, a huge contrast between these two methodologies is that our approach can recuperate the entire provenance tree from the coarse granular view by turning upward the PathIndexes in their P Ds though the approach by Chen et al. can't. Hasan et al. utilize the provenance anchor model and MAC to guarantee honesty of the provenance. Contrasted and our security arrangements, this approach requires to transmit a lot of provenance data as it isn't particularly intended for WSNs.

6.PROPOSED SYSTEM

In existing framework does not have any information pressure part which is one of the significant downside in this venture, in the proposed arrangement the information's that is gathered by the sink hub from the every sensor hubs will be compacted before encoding and it will be sent to the provenance part. Additionally the proposed framework will have grouping idea and information detecting, this will expand the proficiency of the battery with the goal that life of the battery will increment. The current framework has safety efforts yet then the proposed framework will have extra security conspire called bunch approval. A definitive point of the undertaking is to send the information safely regardless of time. The information is compacted utilizing Lempel-Ziv Algorithm.LZW is the main letter of the names of the researchers Abraham Lempel, Jakob Ziv, and Terry Welch, who built up this calculation. LZW pressure is a lossless pressure calculation. Information pressure strategy usage is the most critical errands for any product engineer. LZW pressure strategy is straightforward and is word reference based. A document is examined for a grouping of redundant information happening in the program. These successions are put away in the lexicon inside the compacted document and references are embedded wherever the redundant information happens.

6.1 ADVANTAGES OF PROPOSED SYSTEM

The measure of documents for the most part increments, all things considered, when it incorporates heaps of monotonous information or monochrome pictures. LZW pressure is the best strategy for lessening the span of records containing more tedious information. LZW pressure is quick and easy to apply. Since this is a lossless pressure system, none of the substance in the record are lost amid or after pressure. The decompression calculation dependably takes after the pressure calculation. LZW calculation is proficient in light of the fact that it doesn't have to pass the string table to the decompression code. The table can be reproduced as it was amid pressure, utilizing the information stream as information. This dodges addition of substantial string interpretation table with the pressure information

7. CONCLUSION

In this paper, the current framework utilizes a word reference based secure provenance conspire for remote sensor systems. Utilizing packet way word references, which encase way lists rather than the way itself in the provenance. Consequently, the span of the compacted provenance in our lossless approach is littler than that of the current lossy provenance plans., by utilizing the AM-FM draw conspire and a protected packet arrangement number age method, additionally it guarantee the security targets of the plan. The information part is packed utilizing Lempel-Ziv calculation likewise grouping idea is connected with cluster approval for security upgrade.

8. REFERENCE

[1]H.- S. Lim, Y.- S. Moon, and E. Bertino, "Provenance-based trust-value appraisal in sensor systems," in Proceedings of the Seventh International Workshop on Data Management for Sensor Networks, ser. DMSN '10, 2010, pp. 2– 7.

[2]S. Sultana, G. Ghinita, E. Bertino, and M. Shehab, "A lightweight secure provenance conspire for remote sensor systems," in IEEE eighteenth International Conference on Parallel and Distributed Systems (ICPADS), 2012, pp. 101– 108.

[3]I. Encourage, J. Vockler, M. Wilde, and Y. Zhao, "Delusion: A virtual information framework for speaking to, questioning, and robotizing information induction," in fourteenth International Conference on Scientific and Statistical Database Management, 2002, 2002, pp. 37– 46.

[4]K.- K. Muniswamy-Reddy, D. A. Holland, U. Braun, and M. I. Seltzer, "Provenance-mindful capacity frameworks." in USENIX Annual Technical Conference, General Track, 2006, pp. 43– 56.

[5]S. M. I. Alam and S. Fahmy, "Vitality effective provenance trans-mission in vast scale remote sensor systems," in IEEE Inter-national Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011, pp. 1– 6.

[6]B. Shebaro, S. Sultana, S. Reddy Gopavaram, and E. Bertino, "Exhibiting a lightweight information provenance for sensor net-works," in Proceedings of the 2012 ACM meeting on Computer and interchanges security, 2012, pp. 1022– 1024.

[7]A.- H. Jallad and T. Vladimirova, "Information centricity in remote sen-sor systems," in Guide to Wireless Sensor Networks, ser. PC Communications and Networks. Springer London, 2009, pp. 183– 204.

[8]D. Mama, "Secure input benefit in remote sensor systems," in Information Security Practice and Experience, ser. Address Notes in Computer Science, vol. 4464. Springer Berlin Heidelberg, 2007, Book Section, pp. 116– 128.

[9]S. Sultana, E. Bertino, and M. Shehab, "A provenance based instrument to recognize malevolent bundle dropping enemies in sensor systems," in 31st International Conference on Distributed Computing Systems Workshops (ICDCSW), 2011, pp. 332– 338.

[10]S. Ihara, Information hypothesis for ceaseless frameworks. World Scientific, 1993

[11]J. Ziv and A. Lempel, "A general calculation for successive information pressure," IEEE Transactions on Information Theory, vol. 23, no. 3, pp. 337– 343, 1977.

[12]J. Ziv and A. Lempel, "Pressure of individual successions by means of variable-rate coding," IEEE Transactions on Information Theory, vol. 24, no. 5, pp. 530– 536, 1978.

[13]M. Garofalakis, J. M. Hellerstein, P. Maniatis, and IEEE, "Evidence portrays: Verifiable in-organize conglomeration," in IEEE 23rd Inter-national Conference on Data Engineering, pp. 971– 980.

[14]S. Enrage, M. J. Franklin, J. M. Hellerstein, and W. Hong, "Tag: A little conglomeration benefit for impromptu sensor systems," SIGOPS Oper. Syst. Rev., vol. 36, no. SI, pp. 131– 146, Dec. 2002. [Online]. Accessible: http://doi.acm.org/10.1145/844128.844142

[15]P. Levis, N. Lee, M. Welsh, and D. Culler, "Tossim: exact and adaptable reproduction of whole tinyos applications," in Proceedings of the first worldwide gathering on Embedded organized sensor frameworks, ser. SenSys '03, 2003, pp. 126– 137.

[16] S. Sultana, G. Ghinita, E. Bertino, and M. Shehab, "A lightweight secure plan for identifying provenance fabrication and packet drop assaults in remote sensor systems," IEEE Transactions on Dependable and Secure Computing, vol. 99, no. PrePrints, p. 1, 2014.

[17] R. Hasan, R. Sion, and M. Winslett, "The instance of the phony picasso: Preventing history imitation with secure provenance," in Proccedings of the seventh Conference on File and Storage Technologies, ser. Quick '09, 2009, pp. 1– 14

[18] S. Chen and J. H. Reif, "Productive lossless pressure of trees and diagrams," In IEEE Data Compression Conference (DCC), 1996, pp. 428.