An Energy Efficient Data Transmission And Aggregation of WSN Using

Data Processing In MapReduce

M.Mahalakshmi¹, A. Mohamoodha bi², B. Thamarai Selvi³, M. J. T. Vasantha Priya⁴

¹eswarimrugan@gmail.com, ²mohamoodhabi@gmail.com, ³thamaraibupalan@gmail.com, ⁴vasanthapriyamjt@gmail.com ⁴M. J. T. Vasantha Priya, Assistant Professor, Dept. Of Computer Science and Engineering, Velammal Institute of Technology, TamilNadu, India

Abstract : In this Paper, We construct a wireless sensor network using software defined network(SDN) framework. The sensors are grouped as three different clusters and cluster head elected between each cluster based on their distance, memory and battery to reduce the Energy Consumption of Sensor nodes .The sensed data is sent to the cluster head of each cluster by the other sensor nodes in encrypted form. Now, the data received in the head of cluster are aggregated and signature is appended to the data by Privacy Homomorphism Encryption Scheme using Ecc-Elgamal Signature in a Binary transmission for three completely different Network Clusters and sent to the base station. Here, again the signature is verified and stored in hadoop distributed file system and query processing using Mapreduce.

Key Words: Wireless sensor network, software defined network, big data, MapReduce, Elgamal signature.

1.INTRODUCTION

The basic architecture of Wireless Sensor Networks is usually a Hybrid type where it is a of Infrastructure combination Oriented and Infrastructure less Networks. A wireless sensor network (WSN) is a wireless network consisting of spatially distributed autonomous devices using sensors to monitor physical or environmental conditions. A WSN system incorporates a gateway that provides wireless connectivity back to the wired world and distributed nodes (see Figure 1). The wireless protocol you select depends on your application requirements. Some of the available standards include 2.4 GHz radios based on either IEEE 802.15.4 or IEEE 802.11 (Wi-Fi) standards or proprietary radios, which are usually 900 MHz. The Communication from sensor to sensor head takes place through p2p Architecture (Infrastructure less) and the communication from Cluster Head to Base

Station involves Broadcast Based (Infrastructure Oriented). This Hybrid Architecture is to reduce the Energy Consumption of Sensor nodes as it will be depleted soon when each sensor broadcasts sensed data to Base station as and when it senses. Hence a Cluster Head will be elected for each cluster by considering the battery, Memory and processing ability. All the Sensors will be sending their sensed data to the Cluster Head in a p2p manner.

Bigdata is a term for data sets that are so large or complex that traditional data processing application softwares are inadequate to deal with them. Challenges include capture, storage, analysis, datacuration, search, s haring, transfer, visualization, querying, updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. Hadoop MapReduce (Hadoop Map/Reduce) is a software framework for distributed processing of large data sets on compute clusters of commodity hardware. It is a sub-project of the Apache Hadoop project. The framework takes care of scheduling tasks, monitoring them and re-executing any failed tasks. The process of moving map outputs to the reducers is known as shuffling. Sort: Each reduce task is responsible for reducing the values associated with several intermediate keys. The set of intermediate keys on a single node is automatically sorted by Hadoop before they are presented to the Reducer.

Today, big data is strongly associated with MapReduce [6]. Its simplicity along with its scaling capabilities have rendered MapReduce the de facto



standard for data processing and analytics implementation at large scale. Several frameworks have been based on it, ranging from the integrated platforms, such as hadoop file systems, to proprietary applications spanning in several domains. We introduce a comprehensive solution to support MapReduce inside WSNs. In Software Defined Networking(SDN), the routing of packets can be easily modified according to the current deployment of the MapReduce functions, implemented by the big data application developers, inside the network. Furthermore, according to the SDN approach, all network management functions are centralized in a server, called the Controller, which has a global view of the network and can interact closely with the application. It follows that the Controller is in the best position to decide where to deploy the MapReduce functions and change the routing accordingly. In this course, we also develop an analytical framework which can be employed by the Controller to select the nodes, which will execute these functions. More specifically, three significant cases are considered: (i) the case in which the deployment is not affected by the resource limitations of nodes of the WSN, (ii) the case in which the limitations on the processing and memory resources of the sensor nodes are the cause of constraints in the deployment of the MapReduce functions, and (iii) the case in which the strictest constraints are due to the bandwidth limitations.

The proposed system we construct a wireless sensor network using software defined network(SDN) framework. The sensors are grouped as three different clusters and cluster head elected between each cluster based on their distance, memory and battery to reduce the Energy Consumption of Sensor nodes .The sensed data sent to the cluster head of each cluster by the other sensor nodes in encrypted format. Now, the data received in the head of cluster are aggregated and signature is appended to the data by privacy Homomorphism Encryption Scheme using Ecc-Elgamal Signature in a Binary transmission for three completely different Network Clusters and sent to the base station. Here, again the signature is verified and stored in hadoop distributed file system and query processing using Mapreduce.

2. BIG DATA PROCESSING IN WSNs

MapReduce is presently viewed as the accepted standard in enormous information handling, for the most part because of its effortlessness and its scaling abilities. Amid its ten years of life, since its unique production [6], MapReduce has ruled the examination what's more, improvement in the region of huge information. Apache Hadoop incorporates the most broadly utilized execution of MapReduce, while a few different frameworks, for example, NoSQL databases, have been founded on its worldview to upgrade their scaling out performance.

MapReduce is based upon the supposition that the info informational index can be part into key-esteem matches that will be passed to the examination capacities. The extraction of the key-esteem sets from the first information is executed in the guide work. On a basic level, this operation is completely circulated, since information can be part in a few littler pieces, which are given to various procedures executing a case of the delineate. Handling of the created key-esteem sets is performed by the lessen work. Specifically, all key value sets with a similar key are prepared by the same lessen work occurrence. Along these lines, there can be a few forms (up to the quantity of various keys) executing a diverse occurrence of the diminish work, empowering parallel handling of info information.

Despite the fact that more capacities and operations may take put in a MapReduce-based application, here we are thinking about the two essential ones (guide and lessen), since

we are concentrating on their in-system execution in WSNs, where the assets are in any case constrained and along these lines, operations ought to stay as basic as could be expected under the circumstances. In whatever remains of the paper, we will allude to a system hub executing a outline a diminish work as mapper or reducer, separately.

Execution of guide and diminish capacities is doing the enormous information application designer, since it is utilize case particular. MapReduce just gives the particular of the capacities, so that the fundamental foundation can relegate them to a few procedures, subsequently parallelizing the in general enormous information examination operation.

Regardless of the first server farm arranged plan of Map- Diminish, there have been endeavors to bolster its operations in cell phones. Specifically, the Misco framework [10] empowers cell phones to perform enormous information handling with MapReduce. Aside from the framework level support, a booking calculation for constant preparing utilizing Misco has been proposed in [11]. Nonetheless, the proposed approach centers on the productive treatment of a few distinct applications, treating the gadgets more like low-end servers of a server farm as opposed to data sources. The last is for the most part tended to by works concentrate the utilization of MapReduce in performing examination on data originating from sensors. In this specific situation, the creators in [12] are thinking about the issue of incorporating information originating from expansive scale heterogeneous WSNs. They are concentrating on complex spatio-transient situations where cutting-edge investigation and monstrous measures of information are required, for example, an Earth-wide temperature boost examine, and they determine information demonstrate for productively applying MapReduce operations in the cloud. Another fascinating and normal utilization of WSN information is activity examination in urban situations; truth be told, creators in [13] devise a structure for demonstrating furthermore, anticipating activity wonders utilizing sensor information, while they are likewise considering in-system circulated execution of their calculation. Be that as it may, despite the fact that this approach can be upheld on a fundamental level by MapReduce, no solid plan is given, it is not examined how the system would bolster such operations and moreover, the model is carefully fit for the specific situation contemplated all through that paper.

3. SDN FOR WSNS

The SDN worldview is drawing in the expanding enthusiasm of both the mechanical and scholastic research groups as it fundamentally disentangles arrange control and administration

[14]. Its qualities can in this manner lessen administration costs for system administrators and permit the presentation of

new arrangements in a system as effectively as introducing a product program is, so encouraging advancement. The excitement for SDN has brought about countless commitments

for both wired and remote cell systems.

Less consideration has been paid to the augmentation of the SDN way to deal with WSNs up until this point. To be sure, early endeavors to make organizing conventions connected by WSNs programmable have been introduced in [15]. Later on, the benefits of the SDN approach in WSNs have been talked about in [16], [17], [18].

As of late SDN-WISE [8] has been presented, which progresses the best in class by receiving a stateful approach that is helpful to decrease the measure of data traded between the sensor hubs and the Controller [19]. Moreover, SDN-WISE presents the WISE-Visor which is a layer empowering the creation and administration of a few virtual WSNs in light of the same physical sensor hubs.

The proposed system expands on top of SDN-WISE also, hence, in the accompanying we give some foundation data on SDN-WISE which is helpful to get it whatever is left of this paper. As per the SDN approach, SDN-WISE unmistakably isolates the information plane keep running by the sensor hubs and the control plane executed by a product program, called Controller, running on a server.

The conduct of sensor hubs is for the most part dictated by the substance of the supposed WISE Flow Table which is filled with data originating from the Controller. Like in Open-

Stream, Entries of the WISE Flow Table can be separated in

three segments: Rules, Action, and Statistics. In the Rules area

up to three principles can be characterized. Each lead can investigate any byte (indicated by the Address field in the section) of the parcel. On the off chance that such byte is equal2 to the esteem determined in the Esteem field, then the control is fulfilled.

On the off chance that a parcel fulfills every one of the standards, then it is dealt with by the hub as indicated in the Action area. The sort of activity to execute on the bundle is determined in the Type field. Conceivable sorts are drop, forward, change, and so on. On the off chance that the activity is forward, at that point the hub the bundle ought to be sent to, is determined in the Value field.

At long last, in the Statistics segment data is put away about the use of such section. Different fields are characterized in the WISE Flow Table sections which be that as it may, are past the extent of this paper. Besides, sensor hubs execute a basic convention which empowers them to take in their best next jump towards the sink (or the nearest sink, if a few sinks are conveyed). At long last, SDN-WISE sensor hubs gather data about their neighbors and report such data to the Controller intermittently.

Like in the OpenFlow case, when a SDN-WISE hub gets a bundle for which none of its entrances applies, it sends this parcel to the Controller inside a Rule ask for message. Because of the occasional reports got by the sensor hubs, the Controller has a total perspective of the system topology also, conditions, which it uses to choose how to react to the gotten Rule asks. Take note of that thusly, the Controller can offer standards to sensor hubs in a manner that bundles are directed in the system relying upon any byte they contain, that is, not considering the goal address as it were



Figure1:SDN Detailed Architecture

4.PROPOSED ARCHITECTURE

The basic architecture of Wireless Sensor Networks are usually a Hybrid type where it is a combination of Oriented Infrastructure and Infrastructureless Networks. The Communication from sensor to sensor takes place through p2p head Architecture (Infrastructureless) and the communication from Cluster Head to Base Station involves Broadcast Based (Infrastructure Oriented). This Hybrid Architecture is to reduce the Energy Consumption of Sensor nodes as it will be depleted soon when each sensor broadcasts sensed data to Base station as and when it senses. Hence a Cluster Head will be elected for each cluster by considering the battery, Memory and processing ability. All the Sensors will be sending their sensed data to the Cluster Head in a p2p manner. We Propose an Efficient hybrid data prediction technique with data Aggregation which can drastically reduce energy consumption of sensor nodes during communication. In data prediction the communication can be significantly reduced by avoiding transmission of each raw sample to the sink. This is achieved by using a model to estimate the sensed values, and by communicating with the sink only when there is a change in the sampled data when the Aggregation time out is triggered in cluster head. Each node is using a model to predict its own sensor data, and compares the predicted values with those actually observed and generates a confidence value.

In our Proposed Design we promise to give high authenticity of each sensing data and Integrity of the same in a recoverable environment for concealed Data Aggregation (CDA) by privacy Homomorphism Encryption Scheme using Ecc-Elgamal Signature in a Binary transmission for three completely different Network Clusters Architecture. Base Station can recover each sensing data as well as can compute on it. Overhead is greatly reduced as Cluster Heads of High and Heterogeneous Sensors can respond for Base Station Requests. So communication cost is drastically reduced that a Low Cluster Network can also be deployed to a WSN.



Figure2: Shortest Path Finding By Comparing The Neighbour Nodes



Figure3: Overall Architecture Diagram

International Research Journal of Engineering and Technology (IRJET)e-ISSN: 2395 -0056INJETVolume: 04 Issue: 03 | Mar -2017www.irjet.netp-ISSN: 2395-0072

5. METHODOLOGY USED

There are four separate modules to develop the proposed project:

5.1. Network Formation & electing Cluster Head

Three Completely different Clusters are formed in a

Wireless Sensor Network named

- 1. Homogenous Low Sensors Cluster.
- 2. Homogenous High Sensors Cluster.
- 3. Heterogeneous Sensor Cluster.

Cluster Heads are elected based on the battery, Memory and processing ability for each cluster and only the cluster head aggregate the received data from its cluster sensors and sends to the Base Station .The Homogeneous High and Heterogeneous clusters have the high capability sensors and hence can do computations on aggregated data and send the corresponding result and also respond to Base Station (BS) request thereby reducing overhead drastically. So Homogeneous Low Clusters can also be incorporated into WSN since it contains Low Sensors which just aggregate and sends the cipher text to BS. Each Node is equipped with 3 keys 1.Cluster Key 2.Private Key 3.public Key in which bit length varies depending on type of cluster.

5.2. Data transmission and Aggregation

Sensors send its own sensing data to its cluster head as each and every sensor knows its own cluster head and generates a shortest path to reach it and transmits through it. Each Sensed data is converted into a packet and is encrypted and the cipher is subjected to signature generation process. The Cluster Head receives the encrypted cipher text and signature is verified and the data is aggregated. Cluster Head recovers the data and in Homogeneous High and Heterogeneous and generated the signature using elgamal. For Homogenous Low and Heterogeneous Clusters only Aggregation process takes place as Homogeneous low cluster is memory constrained.

5.3. Recovering data using Signature

The Aggregated Data's are Converted into a Single Packet when Aggregation Time out is triggered in Cluster Head. Now a cipher text is created by encryption and a signature is also generated for the same and the cipher is concealed in signature by compressing and converting the whole compressed content to binary. In Homo Low Cluster the Head Just verifies signature sends the data to BS without aggregating packets. BS once again verifies each and every data by verifying signature thereby ensuring data Integrity and Authenticity. Base Station can also send request to Cluster Heads of High & Heterogeneous Clusters and can receive the recovered data from Cluster head by verifying signature.

5.4. Data Prediction and Query Processing:

The Low Confidence data from the sensor nodes are dropped in cluster head by data prediction strategy on each cluster head. Hence the redundant data's are free from communication reducing overhead. The dropped packets are shown in a graph for redundancy evaluation by time vs. drop count. The Broadcasted compressed encrypted binary packets are being accumulated on Base Station and it is fed to a Database after Verification of Packets from various Clusters. The Historical thus formed by time are subjected to four kinds of Query Processing.

- 1. Top-K Based Query Processing (Top Ranked Values on each Cluster).
- 2. Necessary Set Based Query Processing (Values Should be Present).
- 3. Sufficient Set Based Query Processing (Values that are more than enough).
- 4. Boundary based Query Processing (Ranked values in a range)

Above query processing are implemented using Hadoop, the Historical data's to be stored in HDFS. The query submitted to the master which will retrieve the historical data from the data node. It is faster than the normal DB query processing and large set of data's can be maintained in the HDFS.

6.PROBLEM STATEMENT

We propose the ECC-Elgamal scheme to securely send the data through the sensors in the network clusters. The signature is appended to the binary data in the cluster head sent by the sensor. After appending data reach the base station here again the signature is verified (Lwo Level Authenticity).

6.1. Elgamal algorithm

It is a digital signature scheme which is based on the difficulty of computing discrete logarithms. It was described by Taher ElGamal The ElGamal signature scheme allows a third-party to confirm the authenticity of a message sent over an insecure channel. system parameters

Let *H* be a collision-resistant hash function.

• Let *p* be a large prime such that computing discrete logarithms modulo *p* is difficult.

• Let *g* < *p* be a randomly chosen generator of the multiplicative group of integers

modulo p

These system parameters may be shared between users.

6.1.1. Key generation

- Randomly choose a secret key *x* with 1 < x < p 1.
 - Compute $y = g^x \mod p$.
 - The public key is *y*.
 - The secret key is *x*.

These steps are performed once by the signature

6.1.2. Signature generation

- Choose a random k such that 1 < k < p 1and gcd(k, p - 1) = 1.
 - Compute y=g^k (mod p)
 - Compute s=(H(m)-xr)k ⁻¹ (mod p-1)

If s=0 start over again. Then the pair (r,s) is the digital signature of m. The signer repeats these steps for every signature

6.1.3. Verification

A signature (*r*,*s*) of a message *m* is verified as follows.

• 0 < r < p and 0 < s < p -1 • g^{H(m)} ≡ y^r r^s (mod p)

The verifier accepts a signature if all conditions are satisfied and rejects it otherwise.

6.1.4. Correctness

The algorithm is correct in the sense that a signature generated with the signing algorithm will always be accepted by the verifier.The signature generation implies

 $H(m) \equiv xr + sk \pmod{p-1}$

Hence Fermat's little theorem implies

 $g^{H(m)} \equiv g^{xr} g^{ks}$

$$\equiv (g_x)^r (g_k)^s$$

 $\equiv (y)^r(r)^s \pmod{p}$

A third party can forge signatures either by finding the signer's secret key *x* or by finding collisions in the hash function

 $H(m) \equiv H(M) \pmod{p-1}$

Both problems are believed to be difficult. However, as of 2011 no tight reduction to a computational hardness assumption is known.



Figure4: Working Of Elgammal Algorithm

7. CONCLUSION

We design, develop and evaluate a energy efficient, light weighted, Secure and reliable system for Hybrid WSN to transmit and process Historical data on different types of clusters using Elgammal Signature Scheme, Aggregation methods and Data Prediction Strategies. In this paper, a solution supporting big data processing in Wireless Sensor Networks based on MapReduce has been presented. Leveraging and extending the functionality of SDN-WISE, an architecture, which allows the dynamic loading and execution of the user-specified map and reduce functions in the nodes, has been proposed. Given that, in a WSN, all nodes are considered mappers, the optimal selection of the reducers has been addressed. In this context, three different cases have been studied. The first, baseline case, does not consider any restrictions in the resources of the sensor nodes and it is used to provide the foundations for more complex scenarios. The future work can be predicting the temperature and humidity automatically based on the historical data and graphical representation.

8. REFERENCES

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Inst., May 2011.

[2] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," IEEE Commun. Mag., vol. 49, no. 11, pp. 32–39, Nov. 2011.

[3] Y. Tian and E. Ekici, "Cross-layer collaborative in-network processing in multihop wireless sensor networks," IEEE Trans. Mobile Comput., vol. 6, no. 3, pp. 297–310, Mar. 2007.

[4] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "Innetwork aggregation techniques for wireless sensor networks: A survey," IEEE Wireless Commun., vol. 14, no. 2, pp. 70–87, Apr. 2007.

[5] M. Gregorczyk, T. Pazurkiewicz, and K. Iwanicki, "On decentralized in-network aggregation in real-world scenarios with crowd mobility," in Proc. IEEE Int. Conf. Distrib. Comput. Sens. Syst., May 2014, pp. 83–91.

[6] J. Dean and S. Ghemawat. (2004). MapReduce: Simplified data processing on large clusters. Proc. 6th Conf. Symp. Opeart. Syst. Des. Implementation, p. 10 [Online]. Available: http://dl.acm.org/ citation.cfm?id=1251254.1251264

[7] A. Dunkels, B. Gronvall, and T. Voigt, "Contiki - a lightweight and flexible operating system for tiny networked sensors," in Proc. 29th Annu. IEEE Int. Conf. Local Comput. Netw., Nov. 2004, pp. 455–462.

[8] L. Galluccio, S. Milardo, G. Morabito, and S. Palazzo, "SDN-WISE: Design, prototyping and experimentation of a stateful SDN solution for Wireless Sensor networks," in Proc. IEEE Conf. Comput. Commun., Apr. 2015, pp. 513–521.

[9] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," Commun. ACM, vol. 43, no. 5, pp. 51–58, May 2000.

[10] A. Dou, V. Kalogeraki, D. Gunopulos, T. Mielikainen, and V. H. Tuulos, "Misco: A MapReduce framework for mobile systems," in Proc. 3rd Int. Conf. PErvasive Technol. Related Assistive Environ., 2010, pp. 32:1–32:8.

[11] A. J. Dou, V. Kalogeraki, D. Gunopulos, T. Mielikainen, and V. Tuulos, "Scheduling for realtime mobile MapReduce systems," in Proc. 5th ACM Int. Conf. Distrib. Event-based Syst., 2011, pp. 347– 358. [12] C. Jardak, J. Riihij€arvi, F. Oldewurtel, and P.
M€ahonen, "Parallel€ processing of data from very large-scale wireless sensor networks," in Proc.
19th ACM Int. Symp. High Perform. Distrib.
Comput., 2010, pp. 787–794.

[13] J. Chen, K. H. Low, C. K.-Y. Tan, A. Oran, P. Jaillet, J. Dolan, and G. Sukhatme. (2012). Decentralized data fusion and active sensing with mobile sensors for modeling and predicting spatiotemporal traffic phenomena. Proc. 28th Conf. Uncertainty Artif. Intell., pp. 181–191 [Online]. Available: http://www.auai.org/uai2012/ papers/110.pdf [14] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling innovation in campus networks," SIGCOMM Comput. Commun. Rev., vol. 38, no. 2, pp. 69–74, Mar. 2008.

[15] F. Dressler, I. Dietrich, R. German, and B. Kruger, "A rule-based€ system for programming self-organized sensor and actor networks," Comput. Netw., vol. 53, no. 10, pp. 1737–1750, Jul. 2009.

[16] T. Luo, H.-P. Tan, and T. Quek, "Sensor openflow: Enabling software-defined wireless sensor networks," IEEE Commun. Lett., vol. 16, no. 11, pp. 1896–1899, Nov. 2012.

[17] S. Costanzo, L. Galluccio, G. Morabito, and S. Palazzo, "Software defined wireless networks: Unbridling SDNs," in Proc. Eur. Workshop Softw. Defined Netw., Oct. 2012, pp. 1–6.

[18] D. Zeng, T. Miyazaki, S. Guo, T. Tsukahara, J. Kitamichi, and T. Hayashi, "Evolution of softwaredefined sensor networks," in Proc. IEEE 9th Int. Conf. Mobile Ad-hoc Sens. Netw., Dec. 2013, pp. 410–413.

[19] G. Bianchi, M. Bonola, A. Capone, and C. Cascone, "OpenState: Programming platformindependent stateful openflow applications inside the switch," SIGCOMM Comput. Commun. Rev., vol. 44, no. 2, pp. 44–51, Apr. 2014.

^[20] C. Buratti, A. Stajkic, G. Gardasevic, S. Milardo, M. D. Abrignani, S. Mijovic, G. Morabito, and R. Verdone, "Testing protocols for the internet of things on the EuWIn platform," IEEE Internet Things J.

[21] A. Awad, R. German, and F. Dressler, "Exploiting virtual coordinates for improved routing performance in sensor networks," IEEE Trans. Mobile Comput., vol. 10, no. 9, pp. 1214–1226, Sep. 2011.