

Secure data fusion technique to mitigate Byzantine Attack in WSN

Reshma Begum G H¹, Manjesh B N²

M.Tech(CNE)Student, Dept of computer science Engg, Akshya Institute of Technology, Tumkur, India¹.
Asst.Prof, Dept of computer science Engg, Akshya Institute of Technology, Tumkur, India².

Abstract— Wireless Sensor Networks (WSNs) have played a vital role, and is believed to be one of the vast and emerging technologies as there are several innovative applications both for public sector and military systems. The type of sensing technology used in WSNs combined with the processing power continues to be rich and get to be utilized in abundance in the forthcoming applications. Thus, due to their unlimited prospective views, WSNs are currently receiving substantial attention in several fields of inquiry. Additionally, despite the fact that WSNs are characterized by severely constrained computational and energy resources, they are still complemented by their limitless potential and hence they are currently getting trivial interest. Nevertheless, it is yet too early in the life of such systems to get established, as there are many research challenges that are still to be seen. Therefore, much research has been concentrated on making sensor networks feasible and useful rather than concentrating much on the security aspects of the deployment.

Index Terms— Byzantine Attacks, Distributed Detection, MANETs Sensor Security in WSNs.

I. INTRODUCTION

A WSN is a wireless network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, force per unit area, motion or pollutants, at different placements. Fig.1 shows the typical Multi-hop WSN architecture. The development of WSNs was originally prompted by military applications such as battlefield surveillance.

However, WSNs are now employed in many industrial and civilian application areas, including industrial process monitoring and control, machine health monitoring, environment and habitat monitoring, health care applications, home automation, and traffic control.

In addition to one or more sensors, each node in a sensor network is typically equipped with a radio transceiver or other wireless communications device, a small microcontroller, and an energy source, usually a battery. The envisaged size of a single sensor node can vary from shoebox-sized nodes down to devices the size of grain of dust, although functioning 'motest' of genuine microscopic dimensions have yet to be created. The monetary value of sensor nodes is similarly variable, running from hundreds of pounds to a few pence, depending on the size of the sensor network and the complexity required of individual sensor nodes. Size and

price constraints on sensor nodes result in corresponding constraints on resources such as vitality, memory, computational speed and bandwidth [1].

Byzantines intend to deteriorate the detection performance of the network by suitably modifying their decisions before transmission to the Fusion Center (FC). Byzantine Attacks are not only considered to be the most severe threat to WSNs, but they tend to make it more challenging to protect it from gaining full control over some of the authenticated nodes, eventually, which may lead to the uninformative behavior to disrupt and collapse the system.

Distributed Detection is a classical subject in signal processing and has attracted recent interest due to the possible deployment of wireless detectors for a diversity of applications from environmental monitoring to military surveillance. Most of the research in the field of Distributed Detection has been carried out under the assumption of a secure network. But in the recent past, researchers have looked into the problem of security threats on sensor networks [1].

II. ISSUES IN DISTRIBUTED DETECTION

The classical problem of Distributed Detection considered in the work [2], limits the sensors to get compromised by an intruder. As a result, all the compromised sensors which refer to as Byzantine tend to get reprogrammed by the intruder to attack the FC by transmitting fictitious observations. The uncompromised sensors that are referred to as honest can then follow the expected rules of operation. But, in the context of distributed detection, sensors are more vulnerable to tampering due to the Byzantine Sensor problem which is particularly motivated by the applications of envisioned WSNs. However, the wireless sensors then can be made of low cost devices adhering to the severe constraints on battery power. But, this requires that such practical limitations make use of sophisticated encryption, which eventually makes it more unrealistic.

Furthermore, the wireless transmission medium is more vulnerable to eavesdropping, which makes it possible for the attacker to extract information from sensor transmissions. As a result, the adversary can employ a wide range of strategies, including deploying its own sensors aimed at jamming the transmission of honest sensors or, in a more sophisticated way, transmitting optimally designed signals to confuse the FC.

The analytical characterization of the ability that Byzantine Sensors can affect the decision at the FC is considered is further elaborated in this study [2]. Specifically, this work is proposed from the intruder's perspective, and suggests the most effective attacking strategies by the Byzantine Sensors. As a result, it is evident that when too many sensors are compromised, the FC will lose its ability to detect the underlying phenomenon. But this work proposed lack in defining 1. The minimum population size of the Byzantine Sensors such that the fusion network is rendered completely ineffective. 2. The achievable performance without knowing which sensor is compromised in a situation where a decision maker is bound by an upper bound or with the given sensor population.

The work proposed in [3], a standard model in Distributed Detection under binary Hypotheses H_0 versus H_1 with known distributions is studied. In such a model, all the sensors are assumed to draw observations that are independent, and identically distributed and conditioned on the unknown hypothesis. But, as studied in [4], the classical assumption of conditional independent and identically distributed observations may not always be valid in practice and the literature is evident of such complications of correlated observations. Therefore, the work studied in [3], recognizes the limitations, and makes the conditional independent and identically Distributed assumption for analytical tractability and gain insights into how Byzantine Sensors can affect the overall performance.

Recent researches in wireless communication and portable computers with two or more mobile nodes already issued a temporary network without the use of network infrastructure or centralized management of the sensor mobile network you can create. Source and destination mobile node to communicate with each other, if not within the scope of the data packets to the mobile parts which is between the relaying transmissions through other mobile hosts in the mobile host is not allowed to do. A sensor node, the node captured from the same identity (ID) number of copies of the lead, and malicious activities, and strategic levels of the network makes the replicas. Possible security metrics and efficiency metrics of distributed detection in WSN are given below.

Distributed Detection in the presence of Byzantine Sensors created by an intruder is studied. Further, this work characterizes the power of attack analytically. As a result, this work is able to provide closed-form expressions for the worst detection error exponent of an optimized NP detector at the FC, and for the corresponding attacking distributions. The work further gives an expression of the minimum attacking power above which the ability to detect is completely destroyed. As in the case of vector observations, they find that an intruder infecting less than 50% of the nodes cannot completely impair the system, regardless of the distributions of the sensors' observations. Metrics need to be considered in distributed detection are presented at a lower place.

Security Metrics

1. Node revocation: When replicated nodes are detected, the WSN should be capable of revoking them quickly. To prevent

a node replication attack, an efficient scheme has to detect whether the nodes are compromised or not. Therefore, an intruder cannot use this malicious nodes eavesdrop on the communication of other sensors, or inject false data reports that make a server misjudge.

2. Collusion resistance: If a number of nodes leave the network or are compromised by the intruder, the intruder cannot use these nodes' security element to compromise the whole network. A good detection mechanism must resist the collusion of malicious nodes.

3. Resilience: When an intruder physically captures several sensor nodes and collects all secret elements. Then he inserts the secret element in his malicious sensor nodes and deploys the malicious sensor nodes into the network. If resilience is high, the network is still available. Otherwise, if resilience is low, it may make the whole network broken down.

4. Lightweight: Sensor nodes are usually composed of low memory, energy and computation. Sensors cannot afford the heavy replication detection mechanism which will consume large power and complex computation. Therefore, a lightweight detection scheme is an important principle for resource-constrained wireless sensor network.

Efficiency Metrics

1. Memory: Sensor's memory always stores node's identity that can identify each legitimate sensor node in the network, security element, such as node's public, private key and session key.

2. Energy: Energy consumption is one of the most important things that have to be concerned in wireless sensor network. Complex computation will lead to an amount of energy consumption, so designing an efficient detection scheme is a necessary task.

III. ARCHITECTURE OF SYSTEM

The solution for distributed detection problem will start from developing a randomly distributed simulation for wireless sensor network. The study of the solution to be adopted can be discussed under following stages:

- **Stage-1:** This stage mainly involves in developing a simulation test-bed in wireless sensor network considering the fusion node and sink mainly. The simulation environment considers the random deployment of the sensor nodes thoroughly. Various simulation parameters like number of nodes, node ID, transmission area, selection of fusion node, positioning of sink, and mobile access point.
- **Stage-2:** The design of this stage is mainly focused on large scale wireless sensor network. The basic idea is to find the optimal scheme parameters at relatively small network sizes through exhaustive search, and then obtain the fusion parameters for large network size by exploiting the approximately linearrelationship between the scheme parameters and the network size. It is expected that the proposed linear approach can achieve satisfying accuracy with low false alarm rate. However, there are chances of

violating the problem constraint. To enforce the miss detection constraint and improve the data fusion accuracy, the proposed system will use the linear approximation as the initial point for the optimal exhaustive search algorithm.

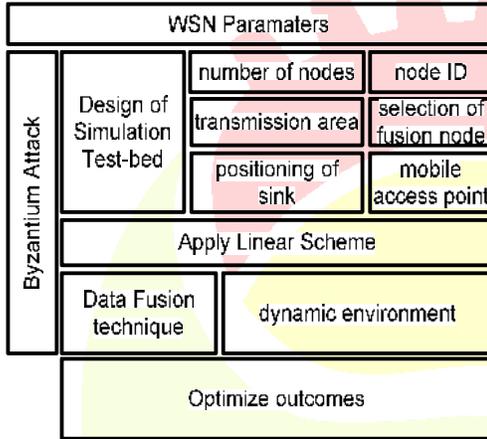


Figure 3.1: Architecture of the Proposed System

- Stage-3:** This stage of development will focus on designing an architecture of data fusion technique that can withstand the most dynamic as well as unpredictable environment in wireless sensor network. The prime focus is to perform identification of the adversarial module. The system will also derive a closed-form solution for novel fusion scheme based on the central limit theorem. It is expected that the closed-form solution is a function of the network size, the percentage of malicious users, the malicious nodes' behavior, and the detection accuracy of the sensor nodes. It is also expected to show that the closed-form solution delivers comparable results with that of the near-optimal solution obtained from the enhanced linear approach.
- Stage-4:** This stage mainly performs enhancement to the proposed system by observing the rate of false alarm in order to minimize it. This stage ensures effective generation of positive alarm with reliable detection probability of the malicious nodes in the wireless sensor network.

3.1 DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its *process* aspects. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design). A DFD shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing

of process or information about whether processes will operate in sequence or in parallel. This section presents the data flow diagram of the system.

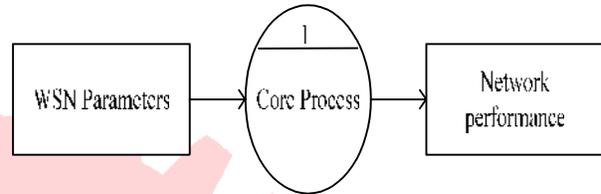


Figure 3.1.1 Level zero data flow diagram

Figure 3.1.1 shows the level zero data flow diagram, where it can be seen that the proposed system considers the input (WSN Parameters), and the core process {1} then performs incorporation with the a proposed a novel Malicious node detection and adaptive fusion algorithms.

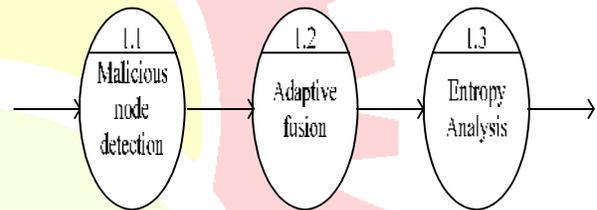


Figure 3.1.2 Level one data flow diagram

Figure 3.1.2 shows the level one data flow diagram, where it can be seen that proposed process {1} is decomposed into Malicious node detection {1.1}, the adaptive fusion algorithms {1.2} and analysis from the entropy point of view {1.3}.

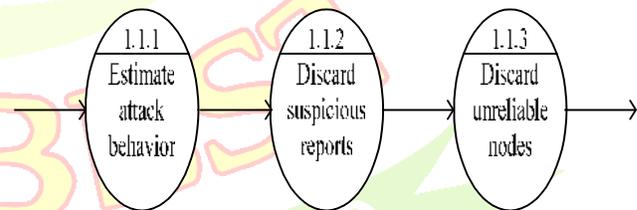


Figure 3.1.3 Level two data flow diagram

Figure 3.1.3 shows the level two data flow diagram, where the sub-process {1.1.1}, in this Estimate attack behavior procedure, we propose a simple malicious node detection scheme, where the sensor decision reports are used to identify the malicious nodes and estimate their attack behavior. The malicious node detection procedure has two levels namely, discard the suspicious reports {1.1.2} and discard the unreliable nodes {1.1.3}.

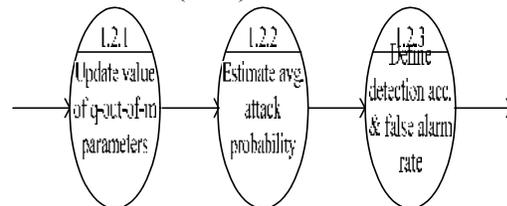


Figure 3.1.4 Level two (1) data flow diagram

Figure 3.1.4 shows a level two (1) data flow diagram, here adaptive fusion can be achieved by updating the value of the q-out-of-m fusion parameters {1.2.1} based on the average probability of attack. By using the set of detected malicious nodes, the total number of sensors detected to be malicious. Then estimate average attack probability {1.2.2}. Then define the detection accuracy and false alarm rate of the malicious node detection scheme {1.2.3}.

3.2 FLOW CHART DIAGRAM

Figure 3.2.1 shows a flowchart of malicious node detection and adaptive fusion algorithm. In this a simple malicious node detection scheme, where the sensors decision reports are used to identify the malicious node and estimate their attack behavior. More specifically, for node 'i' to represents the number of times nodes sends '0' or '1'. The malicious node detection procedure has two levels, namely, discard the suspicious report and discard the unreliable nodes. Then updates the estimated attack probability. It should be noted that 'N' needs to be greater than or equal to a certain threshold 'Nth' before taking the decision to discard any node. Adaptive fusion can be achieved by updating the value of the q-out-of-m fusion parameters based on the average probability attacks. Then estimate the value of average attack probability. The proposed malicious node detection scheme is equivalent to the detection approach based on the entropy-based trust model. This implies that the proposed malicious node detection scheme is optimal from the information theory point of view.

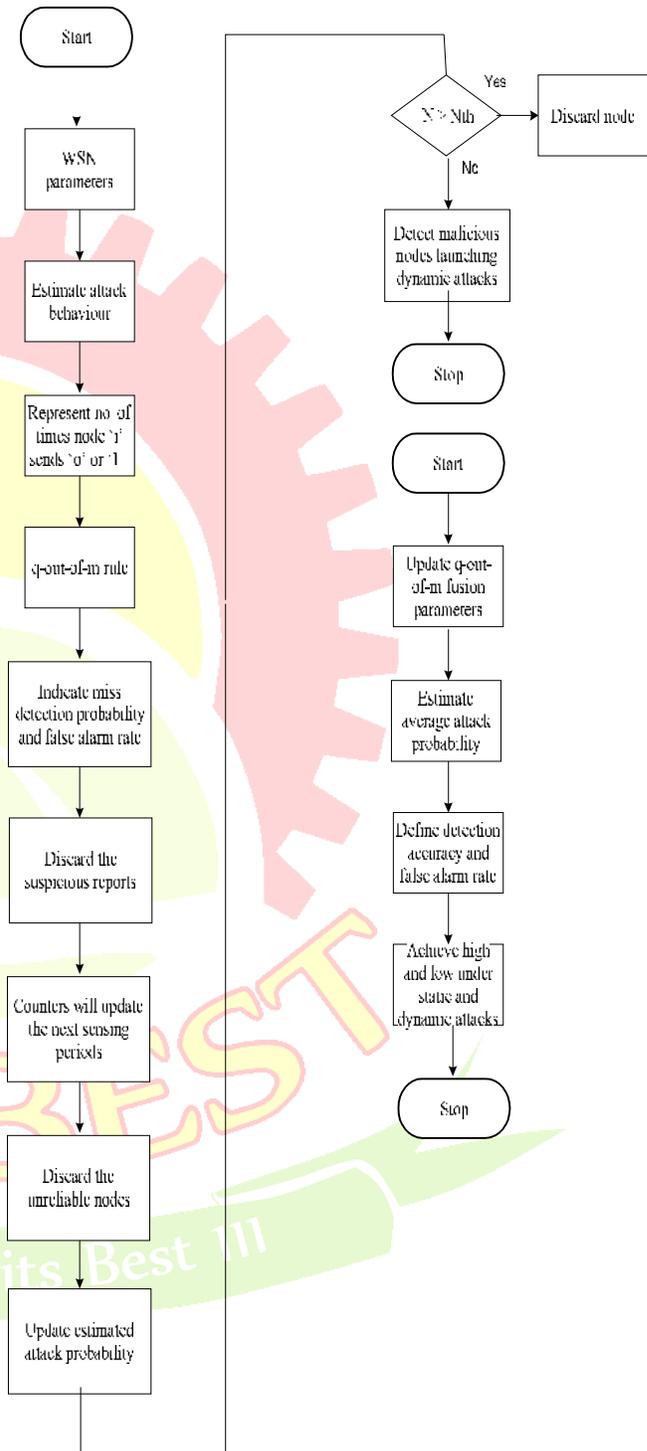


Figure 3.2.1 Flowchart of proposed of malicious node detection and adaptive fusion method

IV. PRIOR STUDY WORK

Wang and Tseng [5] proposed systematical solutions to the k-coverage sensor placement problem and distributed sensor dispatch problem. Their placement solutions allow an arbitrary relationship of sensors' communication distance and their sensing distance, and can work properly under both binary and probabilistic sensing models. It is verified that the interpolating placement scheme requires fewer sensors to ensure k-coverage of the sensing field and connectivity of the network as compared with the duplicate placement scheme. Their dispatch solutions are based on the competitive nature of a distributed network. Simulation results have shown that the competition-based dispatch scheme performs better than the greedy and pattern-based dispatch schemes. However, by selecting sufficient seed locations, the pattern-based scheme can work as efficient as the competition-based scheme.

Bharathidasan et al. [6] have done a survey on the various issues in sensor networks like energy efficiency, routing and localization and the various schemes proposed for these issues and have given brief descriptions of these schemes and conventional routing protocols being used in sensor networks also simulators for sensor networks are discussed. Further work is necessary in the areas of media access control, security and privacy.

Chong et al. [7] introduced more than two decades ago, it was more a vision than a technology ready to be exploited. The early researchers in DSN were severely handicapped by the state of the art in sensors, computers, and communication networks. Even though the benefits of sensor networks were quickly recognized, their application was mostly limited to large military systems. Technological advances in the past decade have completely changed the situation. MEMS technology, more reliable wireless communication, and low-cost manufacturing have resulted in small, inexpensive, and powerful sensors with embedded processing and wireless networking capability.

Perrig et al. [8] presented a suite of security building blocks optimized for resource constrained environments and wireless communication. SPINS has two secure building blocks: SNEP and μ TESLA. SNEP provides the following important baseline security primitives: Data confidentiality, two-party data authentication, and data freshness. A particularly hard problem is to provide efficient broadcast authentication, which is an important mechanism for sensor networks. μ TESLA is a new protocol which provides authenticated broadcast for severely resource-constrained environments. They implemented the above protocols, and show that they are practical even on minimal hardware: the performance of the protocol suite easily matches the data rate of their network. Additionally, they demonstrate that the suite can be used for building higher level protocols.

Rodhe et al. [9] proposed a protocol for query authentication in a sensor network where there is multi-hop communication and the queries are broadcasted by the base station into the network. Authenticating the queries is important so attackers cannot modify existing queries because this would lead to wrong readings; or insert new ones into the network because this would lead to waste of energy. They

propose a layered query authentication protocol that ensures that, in the presence of less than n captured nodes, unauthorized queries are stopped after a small number of hops.

Zhang et al. [10] demonstrated in this paper a novel message authentication approach which adopts a perturbed polynomial-based technique to simultaneously accomplish the goals of lightweight, resilience to a large number of node compromises, immediate authentication, scalability, and non-repudiation. Extensive analysis and experiments have also been conducted to evaluate the scheme in terms of security properties and system overhead.

Chan et al. [11] presented the first algorithm for provably secure hierarchical in-network data aggregation. Their algorithm is guaranteed to detect any manipulation of the aggregate by the adversary beyond what is achievable through direct injection of data values at compromised nodes. In other words, the adversary can never gain any advantage from misrepresenting intermediate aggregation computations.

Awerbuch et al. [12] illustrated a detailed description of several Byzantine attacks (black hole, flood rushing, wormhole and overlay network wormhole), analyze their mechanisms and describe the major mitigation techniques. Through simulation, they perform a quantitative evaluation of the impact of these attacks on an insecure on-demand routing protocol.

Marano et al. [13] considered distributed detection in the presence of Byzantine sensors created by an intruder, and characterized the power of attack analytically. They are able to provide closed-form expressions for the worst detection error exponent of an optimized NP detector at the fusion center, and for the corresponding attacking distributions.

Kosut et al. [14] demonstrated an explicit characterization of the region of achievable rates for a Byzantine attack on distributed source coding with variable-rate codes, deterministic fixed-rate codes, and randomized fixed-rate codes. They saw that a different set of rates were achievable for the three cases, and gave converse proofs and rate achieving coding schemes for each. Variable-rate achievability was shown using an algorithm in which sensors use randomness to make it unlikely that the traitors can fool the coding process.

In the work proposed in [15], the problem of FDR based Distributed Detection in the presence of Byzantines is discussed. Firstly, it studies the work proposed in [2] and is observed that deflection coefficient is not the best heuristic for the design of FDR based Distributed Detection framework in non-asymptotic cases. Hence, in this work, there are several observations made:

1. It is observed that the optimization is performed offline by finding the optimal parameter value through brute-force search.

2. Through empirical studies and analytical justifications, it is observed that system performance can be improved by the use of Kolmogorov-Smirnov distance as the design heuristic.

IV. RESEARCH ISSUES

Distributed Detection in the presence of Byzantine Sensors created by an intruder is studied. Further, this work characterizes the power of attack analytically. As a result, this work is able to provide closed-form expressions for the worst detection error exponent of an optimized NP detector at the FC, and for the corresponding attacking distributions. The work further gives an expression of the minimum attacking power above which the ability to detect is completely destroyed. As to the case of vector observations, they find that an intruder infecting less than 50% of the nodes cannot completely impair the system, regardless of the distributions of the sensors' observations.

Therefore, there are number of future research challenges:

1. It may be of interest to consider a Bayesian formulation with a priori probabilities assigned to the hypotheses, so that the asymptotic performance can be measured in terms of the Chernoff information in such a setting.
2. The tools and schemes that were used in this work can be further exploitable for studying the attacks of a less dangerous intruder that does not know the true state of the nature.

Variable-Rate Distributed Source Coding in the Presence of Byzantine Sensors is studied assuming that the traitors have access to all the source values. However, such an assumption is considered to be vital in many of the converse proofs that are dealt in this work. But, this is a significant assumption that may not be all that realistic. Therefore, it would be worthwhile, though perhaps it appears to be more difficult, to perform the following in the future research line.

1. Firstly, to characterize the achievable rate region without this assumption.
2. Secondly, performing the variable-rate source coding by assuming that the traitors have access only to their own source values or possibly degraded versions of those of the honest sensors. Finally, considering the Byzantine Attacks on other sorts of multi-terminal source coding problems, such as the rate distortion problem still remain as the future research challenge.

Interest in WSNs across industry and academia continues to be very high, even though now a day's experiencing a bit of a "backlash" due to the large number of academic research groups getting involved and few successes commercially to date. The vast popularity of WSNs as a research field for academia has left some to feel that it is becoming difficult to make fundamental contributions although the field is still very young. There is also a sense of ossification behind the Tiny OS and mote platforms which are premature since many application domains involve quite different hardware and software demands than provided by that system. This "second system effect" will likely subside over the next year or so and it will become clearer where the lasting contributions and research directions lie.

Some of the research issues are:

- Transducer design: Developing new sensor transducers that are compact, low power, and cost effective. Bio-degradable / environment-friendly sensor design.

- Electronic system design: The system design is one of the promising challenge areas where several new breakthroughs are possible in the near term leading to fundamentally new design directions. Integrating sensors with the appropriate electronic circuitry to extract digital data, using sensor feedback to enhance the data collection within the electronics, and providing low noise outputs using sensor arrays.
- Node design: Developing low power sensor nodes with appropriate processing and networking capabilities.
- System Design: Developing sensor networks of several nodes and integrating them with application specific information systems.
- Protocol: Distributed algorithms, Power Aware Routing, Dissemination, Time Synchronization, Security, Middleware, Localization of sensors, Data aggregation Techniques, Multimodal sensor fusion, Energy-Efficient real-time Scheduling.

1. CONCLUSION

In this paper a detailed review of Multi-hop WSN Architecture is provided. The problems in distributed detection in wireless sensor networks are presented; an extensive literature survey on distributed detection of byzantine attack in wireless sensor network is summarized. Grounded on this literature survey and issues in distributed detection, subject research issues are presented.

REFERENCES

- [1] Mohan, Akram Pasha, "Distributed Detection of Byzantine Attacks in WSNs: A Short Critical Survey", *International Journal of Computer Science and Information Technology Research* ISSN 2348-120X, Vol. 2, Issue 2, pp: (390-403), Month: April-June 2014.
- [2] Stefano Marano, Vincenzo Matta, and Lang Tong, IEEE "Distributed Detection in the Presence of Byzantine Attacks", *IEEE Transactions On Signal Processing*, Vol. 57, No. 1, January 2009.
- [3] J. N. Tsitsiklis, H. V. Poor and J. B. Thomas, Eds.—"Decentralized detection in Advances in Signal Processing", *JAI Press New York*. pp. 297–344, 1993.
- [4] P. Willett, P. Swaszek, and R. Blum, —"The good, bad and ugly: Distributed Detection of a known signal in dependent Gaussian noise", *IEEE Trans. Signal Process.*, vol. 48, pp. 3266–3279, Dec. 2000.
- [5] Y.-C. Wang and Y.-C. Tseng, "Distributed Deployment Schemes for Mobile Wireless Sensor Networks to Ensure Multilevel Coverage," *IEEE Trans. Parallel and Distributed Systems*, vol. 9, no. 9, pp. 1280-1294, 2008
- [6] A. Bharathidasas and V. Anand, "Sensor Networks: An Overview," technical report, Dept. of Computer Science, Univ. of California at Davis, 2002
- [7] C. Chong and S. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges," *Proc. IEEE*, vol. 91, no. 8, pp. 1247-2056, 2003



- [8] A. Perrig, R. Szewczyk, J.D. Tygar, V. Wen, and D.E. Culler, "Spins: Security Protocols for Sensor Networks," *Wireless Networks*, vol. 8, pp. 521-534, 2002
- [9] I. Rodhe, C. Rohner, and A. Achtzehn, "n-lqa: n-Layers Query Authentication in Sensor Networks," *Proc. IEEE Int'l Conf. Mobile Adhoc and Sensor Systems*, pp. 1-6, 2007
- [10] W. Zhang, N. Subramanian, and G. Wang, "Lightweight and Compromise-Resilient Message Authentication in Sensor Networks," *Proc. IEEE 27th Conf. Computer Comm.*, pp. 1418-1426, 2008
- [11] H. Chan, A. Perrig, and D. Song, "Secure Hierarchical in-Network Aggregation in Sensor Networks," *Proc. 13th ACM Conf. Computer and Comm. Security (ACM CCS '06)*, pp. 278-287, 2006.
- [12] B. Awerbuch, R. Curtmola, H.D., N.-R.C., and R.H., "Mitigating Byzantine Attacks in Ad Hoc Wireless Networks," *Technical Report Version 1*, Dept. of Computer Science, Johns Hopkins Univ., 2004.
- [13] S. Marano, V. Matta, and L. Tong, "Distributed Detection in the Presence of Byzantine Attack in Large Wireless Sensor Networks," *Proc. IEEE Military Comm. Conf.*, pp. 1-4, 2006
- [14] O. Kosut and L. Tong, "Distributed Source Coding in the Presence of Byzantine Sensors," *IEEE Trans. Information Theory*, vol. 54, no. 6, pp. 2550-2565, 2008.
- [15] P. Ray and P. K. Varshney, "False Discovery Rate based sensor decision rules for the Network-wide distributed detection problem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1785-1799, 2011.

BIOGRAPHY

Ms. Reshma Begum G.H, Completed B.E(CSE) in Shridevi Institute of Technology, Tumkur, Affiliated to VTU. Presently pursuing M.Tech in Computer Network and Engineering, Akshaya Institute of Technology, Tumkur. Affiliated to VTU, Belgaum, India.
Email: reshmagh2401@gmail.com.

Mr. Manjesh B.N, M.Tech in Computer Science and Engineering. Presently working as Asst. prof, Dept of Computer Science and Engineering, Akshaya Institute of Technology, Tumkur. Affiliated to VTU, Belgaum, India.
Email: manjeshbn@gmail.com.