WITNESS BASED AND VOTING BASED DATA FUSION ASSURANCE MECHANISM IN SENSOR NETWORKS

M. Umashankar¹ & C. Chandrasekar²
¹K.S. Rangasamy College of Technology, Tiruchengode, Tamilnadu, INDIA
²Periyar University, Salem, Tamilnadu, INDIA
¹Email: tmus2009@gmail.com

ABSTRACT

Security is a very important issue when designing or deploying any network or protocol. The nature of large, ad-hoc, wireless sensor networks presents significant challenges in designing security schemes. One or several sensors then collect the detection results from other sensors. The collected data must be processed by the sensor to reduce the transmission burden before they are transmitted to the base station. This process is called data fusion. Data fusion nodes will fuse the collected data from nearby sensor nodes before they are sent to the base station. If a fusion node is compromised, then the base station cannot ensure the correctness of the fusion data sent to it. Various methods are proposed, that deal with providing an assured data transfer to the Base Station A Witness based approach the base station receives the data and votes from the randomly selected sensor node. The vote comes from other sensor nodes called witness, to verify the correctness of the fusion data. The base station to collect the vote through the selected node, if the selected node is compromised it could forge the vote. the witness node must encrypt the vote to prevent this forgery. Compared with the vote, the encryption requires more bits, increasing transmission burden from the chosen node to the base station. The chosen node consumes more power. the witness-based approach using direct voting mechanism scheme has better performance in terms of assurance, overhead, and delay. The witness node transmits the vote directly to the base station

Keywords: Sensor Network, Data Fusion, Fusion Assurance, Security

1. INTRODUCTION

Wireless Sensor Networks

A wireless sensor networks (WSNs) are consists of inexpensive sensor nodes, each node having continuous sensing capability with limited communication power [6]. They can be used for several applications such as Commercial, civil, and military applications including vehicle tracking, climate monitoring, intelligence, medical and agriculture, etc. Sensor nodes with inbuilt chips and Software for processing specific function. The security application of a Wireless sensor network would give some one the ability to collect and analyze data remotely and detect any kind of attack. In the Military applications they are used wireless sensor networks to collect such sensitive data the information passed over the nature would have to be secure. However, Sensor networks are relatively more insecure repository and routers of data, which increased the need of new security schemes. Their deployment in environments disaster areas, earthquake/rubble zones or in military battlegrounds can be seriously affected by any kind of sensor failure or malicious attack/security threats from an enemy. Sensor nodes are self powered and equipped with low computational power CPU allowing the sensor to execute some kind of treatment before sending a report to the centralized authority

Data fusion assurance problem.

In order to avoid heavy traffic and conserve energy in a sensor network caused by the transmission of raw data back to the base station from each sensor, data fusion nodes can be deployed in the network. In the data fusion process, a data fusion node receives data from a number of sensors, conducts data fusion, and then sends the result (decision) to the base station. One example of such a system is distributed detection using multi sensor networks as described by Varshney in [3] and further discussed in [4], [5], [6]. The sensor nodes collect data from the environment and make their binary decisions based on some detection rules. Then they send these decisions to the data fusion node. The fusion node decides on the presence or absence of the event in that environment, based on the binary data it received, and then sends this result to the base station. One of the key advantages of this distributed detection and fusion scheme is that it reduces the transmission burden between
sensor nodes and the data fusion node. While much effort has gone into the design of fusion algorithms [3], to our knowledge, security and assurance aspects of data fusion systems have not been studied. The current data fusion system puts a great deal of trust on the nodes conducting data fusion. However, if the data fusion node is compromised and becomes malicious, it can send an arbitrary fusion result to the base station. Since the original data are not forwarded to the base station, it is difficult for the base station to verify whether the result is valid. Moreover, sensors might also be compromised. If a sensor is compromised and becomes malicious, it can send incorrect sensing results to the fusion node. However, because some fusion algorithms can tolerate certain number of malicious sensors, we will assume that the number of compromised sensor nodes is tolerable.

Figure 1 depicts a wireless sensor network for distributed detection with N sensors for collecting environment variation data, and a fusion center for making a final decision of detections. This network architecture is similar to the so-called Sensor with Mobile Access (SENMA) [9], [14], Message Ferry, and Data Mule [8]. At the jth sensor, one observation \( y_j \) is undertaken for one of phenomena \( H_i \), where \( i = 1, 2, \ldots, L \). If the detection (raw) data are transmitted to the fusion nodes without any processing, then the transmission imposes a very high communication burden. Hence, each sensor must make a local decision based on the raw data before transmission. The decisions, \( v_j \), \( j = 1, 2, \ldots, N \), can be represented with fewer symbols than the raw data. The sensor then transmits the local decision to M fusion nodes using broadcast. The fusion node can combine all of the local decisions to yield a final result, and directly communicate with the base station. Finally, one of the fusion nodes is specified to send the final result to the base station. Unless all of the fusion nodes or all of the sensors fail, this detection and fusion scheme can guarantee that the base station can obtain the detection result. However, the accuracy of the result is not certain. Two problems must be solved to ensure that the base station obtain the correct result. First, every fusion node must correctly fuse all of the local decisions, which also implies that all of the fusion results must be the same. Several algorithms have been proposed to deal with this problem [7], [11]–[13]. This work assumes that this problem has been solved. The second problem concerns assurance of the fusion result. The transmission between the fusion node and the base station is assumed herein to be error-free. Since some fusion nodes may be compromised, the fusion node chosen by the base station to transmit the fusion result may be one of the compromised nodes. Malicious data may be sent by the compromised node, and the base station cannot discover the compromised nodes from the normal fusion nodes since the data detected by the sensor are not sent directly to the base station. Consequently, the result obtained at the base station may be incorrect.

2. WITNESS BASED DATA FUSION ASSURANCE

Witness nodes to enhance the assurance of data fusion. In order to prove the validity of the fusion result, the fusion node has to provide proofs from several witnesses. A witness is one who also conducts data fusion like a data fusion node, but does not forward its result to the base station; instead, each witness computes the Message Authentication Code (MAC) of the result (we call the MAC a proof), and then provides it to the data fusion node, who must forward the proofs to the base station. If the data fusion node is compromised, and wants to send an invalid fusion result to the base station, it has to forge the proofs on the invalid result. There could be various ways to achieve what we described. Let \( F \) denote the data fusion node. Assume that we have chosen M witnesses, \( w_1, \ldots, w_m \) and \( K_1, \ldots, K_m \) represent the MAC keys they share with the base station. above. We assume that the data fusion node and witness

![Fig. 1. Structure of a wireless sensor network for distributed detection using N sensors and M fusion nodes.](image-url)
nodes share a secret key with the base station. After receiving the data from the sensor node, each witness conducts data fusion, and obtains the result $S_i$; it then sends $MAC_i = MAC(S_i, W_i, K_i)$.

**The Witness Based Data Assurance Algorithm**

This algorithm to ensure the validity of the data fusion result, here they developed a witness-based mechanism, the base station uses the n-out-of-(m+1) voting strategy[2].

- Let there be $m$ witnesses + 1 data fusion node.
- Each witness $W_i$ share an unique key with the BS, $K_i$.
- After receiving reports from the sensor nodes, each witness performs data fusion and obtains the result $r_i$.
- It then sends a MAC (Message Authentication Code) to the data fusion node:
  \[ MAC_i = MAC(r_i, W_i, K_i) \]
- The data fusion node computes its result and sends its MAC key with its witnesses to the BS.
- The BS exercises a voting scheme to determine the validity of the report.
- If the report is corrupted, the BS discards it and polls one of the witness nodes for the correct report.
- The Base Station can employ two voting schemes to determine the validity of the fused report.
  - $m+1$ out of $m+1$: the result is valid if supported by all the witnesses.
  - $n$ out of $m+1$: ($1 \leq n \leq m+1$) the result is valid if supported by at least $n$ witness.

**Over head of Witness Based Data Assurance**

In the n out of m+1 scheme, where $m+1 \geq n$ there are two scenarios in which the base station may not receive a valid result: + Current data fusion node compromised but enough honest (uncompromised) witnesses; + Not enough honest nodes (data fusion node or witnesses). In the first scenario, in which the data fusion node is compromised or malicious, it may send an arbitrary fusion result together with the MAC messages to the base station, which rejects the result without enough endorsements from the witnesses. The base station then polls other witnesses and activates one of them as the new data fusion node and obtains its data fusion result. In the second scenario, in which the number of compromised or malicious nodes (including the current data fusion node and witnesses) is larger than $m-n+1$, there are at most $n-1$ honest nodes in the system. So it is impossible for the base station to receive a valid result. Under such circumstance, the base station may till poll other witness nodes and activate one of them as the new data fusion node. The difference between these scenarios is that, with the polling scheme, the base station will be able to obtain a valid result in the first scenario but not in the second scenario.

### 3. VOTING BASED FUSION ASSURANCE MECHANISM

As in the witness-based approach, a fusion node is selected to transmit the fusion result, while other fusion nodes serve as witnesses. Nevertheless, the base station obtains votes contributing to the transmitted fusion result directly from the witness nodes. Only one copy of the correct fusion data provided by one uncompromised fusion node is transmitted to the base station. No valid fusion data are available if the transmitted fusion data are not approved by a pre-set number of witness nodes. Analytical and simulation results reveal that the proposed scheme is up to 40 times better on the overhead than that of the witness based approach.

The voting mechanism[1] in the witness-based approach is designed according to the MAC of the fusion result at each witness node. This design is reasonable when the witness node does not know about the fusion result at the chosen node. However, in practice, the witness node is in the communication range of the chosen node and the base station, and therefore can overhear the transmitted fusion result from the chosen node. The witness node then can compare the overheard result with its own fusion result.

Finally, the witness node can transmit its vote (agreement or disagreement) on the overheard result directly to the base station, rather than through the chosen node. The base station has to set up a group key for all fusion nodes to ensure that the direct voting mechanism works.

When a fusion node wishes to send its fusion result to the base station, it adopts the group key to encrypt the result, and other fusion nodes serving as witness nodes can decode the encrypted result. The witness node then starts to vote on the transmitted result. A Polling Scheme based on the voting mechanism using a public key is proposed to ensure data fusion assurance.
The voting mechanism in the witness-based approach is designed according to the MAC of the fusion result at each witness node. This design is reasonable when the witness node does not know about the fusion result at the chosen node. However, in practice, the base station can transmit the fusion result of the chosen node to the witness or the witness node is in the communication range of the chosen node and the base station. Therefore, the witness node can obtain the transmitted fusion result from the chosen node through the base station or overhearing. The witness node then can compare the transmitted fusion result with its own fusion result. Finally, the witness node can send its vote (agreement or disagreement) on the transmitted result directly to the base station, rather than through the chosen node. The base station has to set up a group key for all fusion nodes to ensure that the direct voting mechanism works. When a fusion node wishes to send its fusion result to the base station, it adopts the group key to encrypt the result, and other fusion nodes serving as witness nodes can decode the encrypted result. The witness node then starts to vote on the transmitted result. Two data fusion assurance schemes are proposed based on the voting mechanism using a group key.

In this scheme[1], the base station needs to ask the witness node whether it agrees or disagrees with the transmitted fusion result. The witness node then sends its vote to the base station. If the transmitted fusion result is not supported by at least T witness nodes, then the base station might have to select a witness node that does not agree with the transmitted result as the next chosen node. The detail steps of the scheme are given as follows:

**Step 1:** The base station chooses a fusion node. Other fusion nodes serve as witness nodes.

Define a set of witness nodes that includes all witness nodes and let the nodes in the set be randomly ordered. Denote \( M' = M - 1 \) as the size of the witness set in the current round.

**Step 2:** The chosen node transmits its fusion result to the base station.

**Step 3:** The base station polls the node in the witness set by following the order of the witness nodes. The polling-for-vote process does not stop until

- \( T \) witness nodes agree with the transmitted fusion result (agreeing nodes), where \( 1 \leq T \leq M - 1 \),
- \( M' - T + 1 \) witness nodes disagree with the transmitted fusion result (disagreeing nodes), or
- all witness nodes have been polled.

**Step 4:** Represent A as the number of witness nodes that agree with the transmitted fusion result.

- If \( A = T \), then the transmitted fusion result passes the verification of the fusion result. Stop the polling.
- If \( M' - T + 1 < A < T \), then no reliable fusion result is valid. Stop the polling.
- If \( A \leq M' - T - 1 \), then exclude the A agreeing witness nodes from the witness set. Let the first node that disagrees with the transmitted fusion result be the chosen node to transmit its fusion result.

### 4. PROPOSED DATA FUSION ASSURANCE USING SILENT NEGATIVE VOTING

As in the Direct Voting Mechanism based approach, a fusion node is selected to transmit the fusion result, while other fusion nodes serve as witnesses. But in this case, witnesses nodes will be silent if there is no compromised nodes. If a compromised node is sending false data, then one or more witnesses nodes will put a negative vote.

- In the proposed method, a fusion node is randomly selected for forwarding the fusion data as in the previous methods. But, instead of sending the data, the fusion node will send a MAC (Message Authentication Code) by encrypting it with its private key provided by the BS.
- The BS will receive the encrypted MAC and decrypt it with the private key of the selected Fusion Node.
- The BS will broadcast the MAC after encrypting it using a Public key or Group key and wait for Negative votes from the fusion nodes which will not compromise with the MAC.
- All the Fusion nodes will receive the Encrypted MAC given by BS and calculate another MAC using the locally available Fusion Data and compare it with the Decrypted copy of Received MAC.
- If the Received MAC and the Newly created MAC differ, then the fusion node will prepare a Negative-Vote along with newly calculated MAC encrypt it with its private key and pole it to BS.
- If there will not be sufficient Negative-votes from fusion nodes, then the BS will ask the selected Fusion Node for real Fusion Data and Receive it.

*Advantages of the Proposed Mechanism.*
Since small size MAC is only used to validate the data, and only one time it is transmitted from one selected fusion node to BS, the power will be preserved at other fusion nodes.

Since the Fusion Data transmission will consume lot of power, obviously the proposed method will preserve lot of transmission power by avoiding retransmission.

Since Negative-voting mechanism is used, the power will be used for Negative-voting if and only if there is a invalid MAC at BS. So the power at the Fusion nodes will not be wasted for voting/Negative-voting during normal operations.

5. CONCLUSION

We have studied the data fusion assurance problem in this paper. To ensure the validity of the data fusion result, To reduce energy consumption in our scheme, we have analyzed and computed the minimum length needed for the Message Authentication Code to achieve a pre-defined level of security. Our results show that the number of bits used for MACs does not increase linearly with the number of witnesses. We have also proposed a polling scheme and studied the overhead of our scheme when the invalid data from fusion nodes are rejected.

In this data fusion scheme the base station in the sensor network collects the fusion data and the votes on the data directly from the fusion nodes. This scheme is more reliable with less assurance overhead and delay than the witnessed based approach. Here Polling Scheme is an overhead. Use of a public key is a threat to security. These type of problems to be eliminated in our proposed Data fusion assurance using silent negative voting method.

6. REFERENCES


