www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 7 July 2017, Page No. 21957-21962
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i7.10

Fuzzy Adaptive Selection of Votes in Probabilistic Filtering Scheme in WSNs

Muhamad Akram^{1,*}, Tae Ho Cho^{2}
[1]Sungkyunkwan University, College of Information and Communication Engineering,
Suwon 16419, Republic of Korea
akram.khan@skku.edu
*Corresponding Author
[2]Sungkyunkwan University, College of Software,
Suwon 16419, Republic of Korea
thcho@skku.edu

Abstract: ***Wireless senor networks comprise of tiny nodes with limited energy and computational resources. Sensor networks are usually installed in unattended volatile environments where they are exposed to sever security attacks. Attackers can generate two main types of attacks i.e. injection of false reports and false MACs. PVFS is a singular en-route filtering scheme which counters both the attacks. However, the number of MACs attached to each report is fixed in PVFS. In this paper, we propose a scheme that supports fuzzy adaptive choice of votes to be included in the report before forwarding it to the base station in probabilistic voting based filtering scheme. The proposed method attains better energy saving when the attacks are not very hostile in terms of their frequency of occurrence, which helps extend network lifetime.***

Keywords: energy, PVFS, false data, false vote.

1. Introduction

The energy and computational resources of wireless sensor networks (WSNs) are normally very scarce and limited as the sensor nodes are very tiny and hardware restricted. Furthermore, they are deployed in exposed environments that increases their proneness to threats such as compromise of sensor nodes. [1]- [4]. Attackers endeavor to compromise sensor nodes and consequently exploit them to generate various attacks. Since the energies of sensor nodes are very scarce and limited, therefore their energies can be depleted very fast by injecting false reports through compromised nodes in the network [3], [5]-[7]. The injection of false report attack is shortly referred to as FRI attack. Similarly, compromised nodes are exploited to attach false message authentication codes (MACs), alternatively referred to as votes, to legitimate data reports containing event information. When an intermediate verification node finds a false MAC, the same report is immediately dropped en-route during the process of verification [5]. The en-route filtering of legitimate reports stops them from reaching the base station (BS). Thus the injection of false MACs i.e. FMI attack bars delivery of the critical and legitimate information to the BS. Therefore, it is of utmost importance that both the attacks be tackled simultaneously so that the energy of the network can be saved from draining as well as the legitimate reports are ensured to reach the BS.

Figure 1 shows the two types of attacks i.e. false positive (FRI) and false negative (FMI) attacks. It is evident that the propagation of false reports depletes the energy of the intermediate nodes whereas false MACs cause true reports to be dropped before reaching the destination.

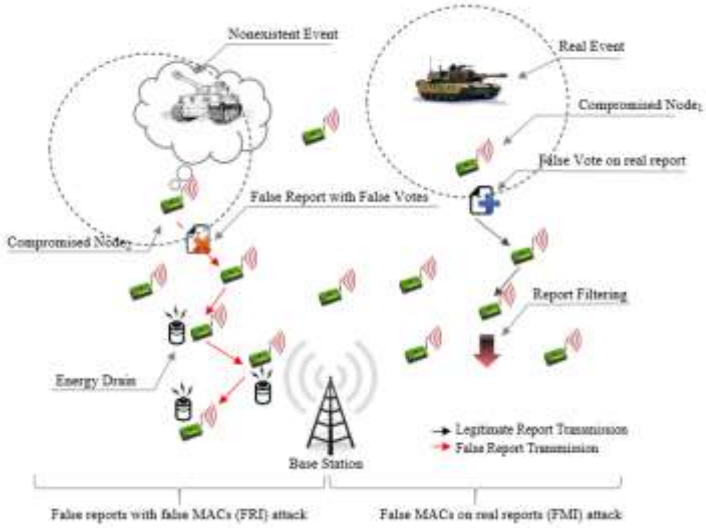
Nonexistent Event
Real Event
Compromised Node$_1$
False Vote on real report
False Report with False Votes
Compromised Node$_2$
Report Filtering
Energy Drain
Legitimate Report Transmission
False Report Transmission
Base Station
False reports with false MACs (FRI) attack
False MACs on real reports (FMI) attack

Figure 1: False positive and false negative attacks

2. Existing En-route Filtering Schemes

Safeguard against FRI and FMI attacks in WSNs has been an important theme of research and motivation behind development of en-route filtering schemes. Several static and dynamic key based filtering schemes have been proposed to counter both or either of the two attacks [1]-[3], [5], [8]-[11]. Probabilistic voting-based filtering scheme (PVFS) [5] is based on static key sharing method and it offers to tackle both the attacks i.e. to filter fabricated reports and deliver legitimate report with few false MACs to the BS. Apart from PVFS, rest of the filtering schemes only focus on countering false data injection attack.

Numerous cluster based message authentication and en-route filtering protocols are effective choices for saving energy. To this date, PVFS is the only scheme that tackles FRI and FMI attacks at the same time and provides an advantage of lesser key storage overhead [5]. PVFS averts the likelihood of filtering original reports while countering FMI attack.

Ye et al. have proposed statistical en-route filtering (SEF) [3]. In SEF, an intermediate verification node can only probabilistically verify the MAC in the report if it has the corresponding key and drops the report if it finds the MAC to be false. SEF has a very limited filtering capability measured by the probability with which an intermediate node detects false reports. In SEF, a legitimate report is subjected to repeated verifications en-route before it reaches the BS, which causes an energy inefficiency. SEF doesn't provide a mechanism to distinguish between false reports and valid reports with false MACs. In CCEF [10], the false report still manages to reach the BS, even if the event reporting CH is compromised because each forwarding node only verifies the session key shared by the BS with the nodes between the CH and the BS to authenticate the source. The secure ticket-based en-route filtering scheme (STEF) [9] drops false reports en-route and provides a safeguard against Path-based Denial of Service attacks (PDoS). In STEF, a single compromised node can attach a false MAC to the legitimate report which is immediately filtered out en-route after being detected with a false MAC. Zhu et al. proposed interleaved hop-by-hop authentication scheme (IHA) that aims at detecting false reports but fails to detect the FMI attack [11]. In IHA, the messages from the BS to the CH and from the CH to BS are routed through the same path. The report is verified by each intermediate node using their session keys shared with the lower associated nodes. However, only BS can verify the report using individual keys of the cluster member nodes. The problem with this verification scheme is that a report with a single false MAC, calculated by a member node travels all the way to the BS and BS drops it after detecting the false MAC. In IHA, not only the energy resource of the intermediate nodes is invested in forwarding such reports but also the BS fails to the get the real information about the event.

The aforementioned schemes offer effective protection against FRI attack in Wireless sensor networks. However, these methods do not afford an effective energy saving answer.

3. Fuzzy Based Inferencing in WSNs

Fuzzy inferencing techniques have been used in WSNs in the past. Fuzzy inferencing aids sensor network in achieving adaptability according to varying situations and variables. Fuzzy control has mostly been proposed for choosing between routing paths, re-clustering and selecting among security-algorithms in the network.

Fuzzy logic systems provide a superlative alternative when mathematical models are impossible to develop. In WSNs, mathematical solutions with adaptive capabilities to respond to the dynamic network behaviors often become complex to develop. Sometime they also prove to be impractical and unviable methods as they are inextensible. Not only network topologies and configurations tend to be dynamic, but also the environments in which they are deployed, exhibit variable behaviors. In such circumstances, Fuzzy rule-based inference solutions stand out as a natural alternative to deal with ambiguous and vague values [12].

4. Rationale and Motivation

All filtering schemes solely focus on countering only the FRI attack with an exception to PVFS which provides protection against both the attacks i.e. FRI and FMI attacks. Moreover, Key information sharing overhead, continuous verification of the report at every intermediate node, greater number of MACs attached to every report make those schemes less energy efficient. Sensor nodes are supplied with limited energy, and energy efficient data routing protocols and filtering schemes need to be developed to save energy and extend network life. Sensor nodes closer to the BS often possess more verification keys than other node and they are expected to perform more report verifications than those nodes that are away from the BS [13].

We have chosen PVFS to further enhance it in terms of energy efficiency because PVFS ensures to deliver a legitimate report to the base station unless more than V_f MACs are invalid.

PVFS has three desirable characteristics:

1. Verification nodes are chosen probabilistically, therefore the attacker cannot predict which nodes are verification nodes. Predicting verification nodes makes it harder for the adversaries to avoid being detected.
2. Nodes around the BS will have less chances of possessing verification keys for the farther clusters which reduces the key overhead.
3. There is a higher probability that the report is either rejected or accepted in the first few hops.

The average number of keys possessed by a verification node in PVFS is

$$N_{ver\_keys} = c.\ (d_{max} - d_i) \qquad (1)$$

and the total number of keys is:

$$N_{ver\_keys} = \sum_{j=1}^{(d_{\max} - d_i)} c \qquad (2)$$

where

d_{max} = the distance between the CH and the BS,

d_i = the distance between the verification node and the BS.

Equations (1) and (2) reveal that the key storage overhead in PVFS is very low in comparison to other existing schemes. However, PVFS uses fixed number of MACs to authenticate a report. This value is selected pre-deterministically before the network organization and remains constant during the network lifetime. The selection of fixed MAC number encourages unnecessary consumption of the energy resource. Thus every report is required to be verified by a minimum of s verification nodes. Not only this, but also every report carries exactly s MACs along with the report. If the size of single MAC is 4 bytes and the number of MACs attached to each report is 5, the total MAC size becomes 20 bytes. Energy consumed during the transmission is directly proportional to the size of the data given by the following equation:

$$E_{Tx} = E_{elec(k)} + \varepsilon_{amp} \times k \times d^{\alpha} \qquad (3)$$

where

k = Size of data,

d = Distance between the sender and the receiver,

$E_{elec(k)}$ = Energy used to run the radio electronic,

ε_{amp} = Energy used to amplify the signal.

α = 2 for the free-space.

In this paper, we propose a fuzzy adaptive scheme that chooses s, the minimum required MACs to attach to a report after endorsement by L member nodes. The fuzzy adaptive selection of MACs saves energy when the attacks are very low or virtually nonexistent. It provides stronger protection when the attack frequency is higher by attaching increased number of MACs to the report for authentication. We observe that our scheme helps to prolong the network life-time. Fuzzy based inferencing derives the values for the inputs to calculate total MACs to send along the reports in each round.

5. Probabilistic Voting-based Filtering Scheme

PVFS uses static key management scheme [8]. In PVFS, verification nodes are chosen probabilistically during the network bootstrap phase when no nodes are compromised.

5.1 Initialization and Key Assignment

In PVFS, each node gets a single key from a global key pool $\{K_i:\ 0 \le i \ge n-1\}$. After sensor deployment, the key pool is divided into *cn* non-overlapping partitions $\mathcal{K}_{\text{Cid}} = \{K_i:\ L \times C_{id} \le i \ge L \times (C_{id}+1)-1\}$; where each partition consists of *L* keys and the number of clusters in the network is *cn*, therefore total number of nodes are $n = cn \times L$. C_{id} is the cluster id. Sensor nodes are organized into clusters and every member node selects a single key from the partition $\mathcal{K}_{\text{Cid}}$ and stores it in the format (i, K_i) where *i* is the key index. Every key K_i belongs to partition $\mathcal{K}_{\text{Cid}}$ or cluster C_{id} if $C_{id} \leftarrow \lfloor i/L \rfloor$. During the path discovery phase, a cluster head (CH) discovers multiple routing paths to the BS. The CH selects verification nodes with a probability given by:

$$P_{filtering\_node} = h_i / h_{CH}$$

where

h_i = Hops between the intermediate filtering cluster head CH_i and the BS, and

h_{CH} = Hops between the BS and the CH that prepares event reports.

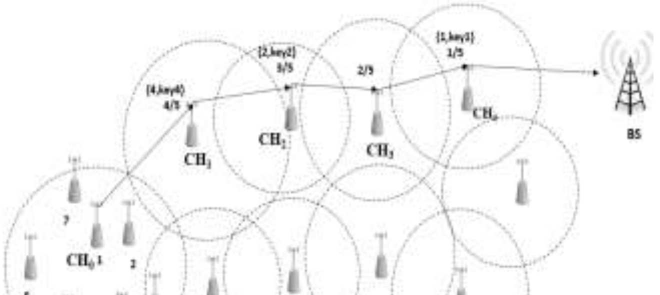

Figure 2: Filtering node selection in PVFS

Figure 2 shows selection of the filtering nodes and assignment of verification keys to them in PVFS. The number of filtering nodes on each routing path is same and equals the nodes constituting a cluster. Assuming that there is no node compromised, the keys are sent either directly or using pairwise key establishment protocols. Each filtering node also shares a symmetric key with the BS to construct the signatures of their verifications.

5.2 Report Preparation

Whenever an event occurs, the clusters closer to the event area contest with each other for generating and forwarding the report and the wining cluster prepares the report. The CH, after generating the report, broadcasts it in the cluster. The member nodes compare the report with their observed values to decide if the report is consistent with their observations. If so, they authenticate the report by casting their votes using their generation keys and send them to the CH. Each vote is of the form: *Vote*: $(i, EK_i\ (H(Report)))$.

After receiving all the votes/MACs, the CH arbitrarily chooses $s \le L$ votes, its own MAC inclusive, attaches them to the prepared report. PVFS uses the least hop count distance for the delivery of reports to the BS. However, in the case of node failure, it choose the second shortest path among the discovered routing paths. Each of the member nodes is notified by the CH to share their keys with one of the chosen filtering nodes.

PVFS does not use variable value of *s* and it is fixed whether the frequency of the attacks is low or high, or the energy of the nodes on the routing path is whether sufficient enough to forward larger packets. The greater the value of *s* (the number of MACs) the greater will be the size of the packets and more nodes will be involved in verifying the report. On the contrary, the smaller *s* is, the greater the likelihood of valid reports being filtered because of false MAC injection attack which leads to an increased filtering of reports [5], [13]. Therefore, it is absolutely desirable that the solution should be able to decide on the suitable sum of MACs to authenticate reports depending upon the values of the input variables.

6. Fuzzy Adaptive Dynamic Selection of MACs

In reality, the CH which prepares the report, always affixes fixed number of MACs to the data report. The energy of the intermediate nodes is limited and making them to either forward reports with more MACs or verify reports all the time, irrespective of the attack intensity, is unreasonable and suffers expense of the energy. Therefore, we propose that the CH must be capable to adaptively decide the value of *s* depending upon the values of the inputs such as attack situation, energy of the intermediate nodes and the distance between the CH and the BS. Logically, if the attacks are low as well as the energy of the filtering nodes, the MACs to include in the report should also be kept low.

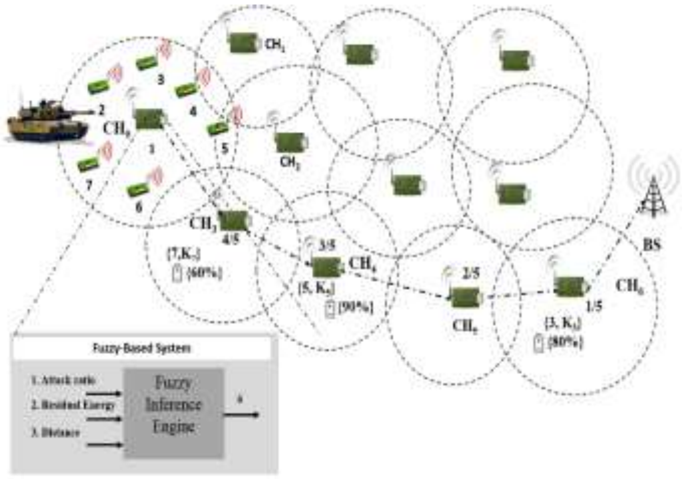

Figure 3: Fuzzy adaptive schematic

Figure 3 is the schematic diagram of our proposed scheme and shows that the CH takes as input the attack ratio, the distance between CH and the BS, and the energy of the intermediate verification nodes and produces an output value of the network parameter, δ.

The value of the output parameter δ is used in algorithm 1 to compute the value of *s* and CH attaches the calculated number of MACs.

Algorithm 1.

```
$L$:  Number of member cluster nodes
$s$:  Number of MACs to authenticate a report
$CH$:   report generating cluster head
$\{Vote_i\}$= set of votes by all the nodes in the cluster
$\delta$:  Output of the fuzzy inference system
$V_t$: Threshold value for true MACs
$V_f$: Threshold value for false MACs
$\delta = CH.fuzzy(Attack\ ratio, Residual\_energy, Distance)$;
set $V_f = 2$;
if $\lfloor \delta \rfloor \le (1/3)$ then
    set $s := V_t = floor(1/3 \times L)$ and $V_f = 1$;
end if
else if $\lfloor \delta \rfloor \ge (2/3)$ then
    set $s := \lfloor 2/3 \times L \rfloor$ , $V_t = \lceil s/2 \rceil$ and $V_f = 3$;
end else if
else
    set $s := \lfloor \delta \times L \rfloor$ and $V_t = s - 1$;
end else
Select $s$ votes from $\{Vote_i\}$ and affix to report;
Forward report, exit;
```

If frequency of attacks as well as the energy of the filtering nodes are high, the value of *s* should be essentially high to permit more en-route verifications. The fuzzy inference system takes the following as input:

1. *Attack ratio*: This input shows the frequency of the attacks made by the adversary by exploiting compromised node. Attack ratio can be calculate as

$$Attack\ ratio = \left(\frac{R_{total} - R_{received}}{R_{total}}\right).100(\%)$$

where

R_{total} = total reports generated and
$R_{received}$ = total reports received at the BS.

2. *Residual Energy*: This values shows the average remaining energy of filtering nodes between the CH and the BS.

$$Residual\ energy = \sum_{i=1}^{L}\left(\frac{energy_i}{L}\right).100(\%)$$

3. *Distance*: This input shows the distance between the CH and the BS.

The decision making system at the CH makes use of fuzzy inferencing to understand the blend of the above three input factors and produces an output value $0 \le \delta \le 1$ which is used by algorithm 1 to calculate *s*.

As we can we see in the algorithm that the minimum value of *s* is 1/3×*L*. The value of $V_f \geq 2$ ensures to counter both the attacks, i.e. FRI and FMI attack. $V_f = 1$ means that only FRI attack is being tackled. In order to save more energy, V_t can be selected to be less than $\lceil s/2 \rceil$, when s = $\lfloor \delta \times L \rfloor$ is greater; because with larger cluster size *L*, $\lceil s/2 \rceil$ also gets unreasonably high.

We can also see that in case of more nodes are compromised in the cluster, there will be more nodes attaching false votes to the report. Therefore, the value of V_f can also be varied beginning with an integer value 2 to desired value. Smaller value of V_f helps save more energy however it also increases the risk of FMI attack. The higher value of V_f helps achieve greater security against the FMI attack, however a false report may incur greater energy expense before being dropped by a filtering node.

The fuzzified set of values for the fuzzy membership functions have been selected consequent upon comprehensive study of the WSN behavior. These sets of values explain the ideal energy conserving requirements of WSNs very well. The tags of the fuzzified set of values for the fuzzy membership functions are as under:

- Attack ratio: {low, moderate, high}
- Residual energy: {little, enough, high}
- Distance: {small, medium, big}
- δ: {very few, few, normal, many}

There are total 27(3×3×3) fuzzy *if-then* rules that transform the arrangement of the input parameters to an output. Table 1 shows some of these rules.

Table 1: Fuzzy if-then rules

No.	Input			Output
	Attack ratio	Residual energy	Distance	δ
01	low	little	small	very few
06	low	enough	big	few
14	moderate	enough	medium	normal
18	moderate	high	big	many
24	high	enough	big	many

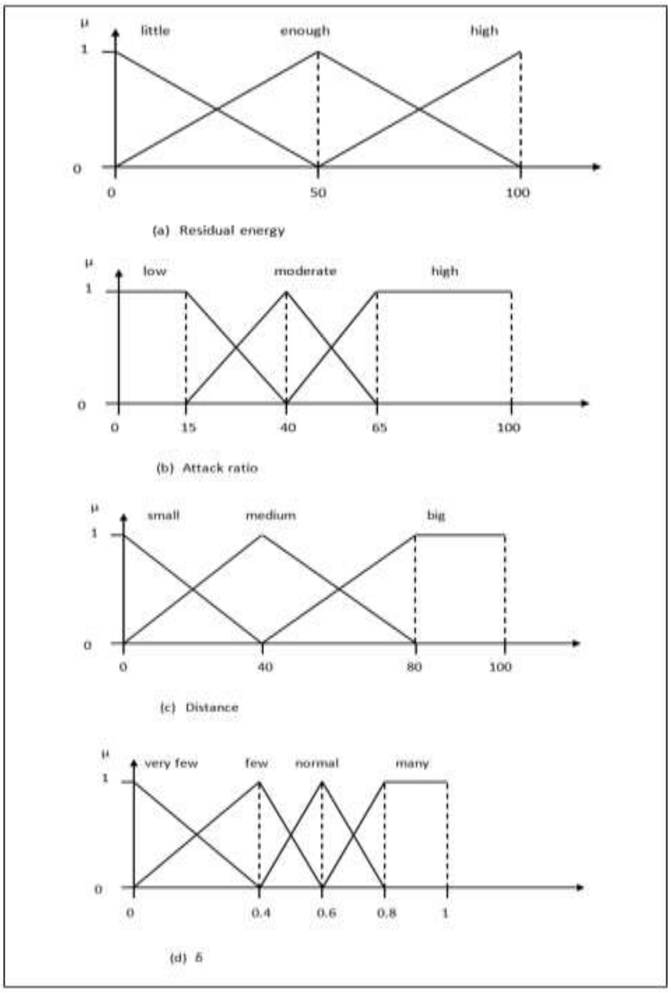

Figure 4: Fuzzy membership functions

Figure 4 shows the fuzzy membership functions of the input and output variables. The fuzzy membership functions have been cautiously picked and adjusted based on the simulation results to attain the best results.

7. Simulation Results

The efficacy and energy saving behavior of our proposed scheme have been validated through simulations. Simulations were done in a custom made simulator made in MS visual C++ 2012. The network comprises of 3500 nodes organized into clusters. The number of CHs is 350 and each cluster comprises of 10 nodes. The network size is 1000×700 m^2. The size of a report is 36 bytes and MAC is 4 bytes. It consumes 15μJ to generate a MAC, 75μJ to verify a report, 16.25μJ/byte for transmission and 12.5μJ/byte for receiving. Table 3 displays the simulation parameters.

Table 2: Simulation setup parameters

Parameter	Value
Total Nodes	3500
Network area	1000 x 700 m^2
Number of clusters and CHs	350
Cluster member nodes including CH	10
Threshold value of V_f	1~3
Report size	36 bytes
MAC size	4 bytes
Energy Consumed to:	
Generate a MAC	15μJ
Verify a report	75 μJ
Transmit/receive a byte	16.25/12.5 μJ

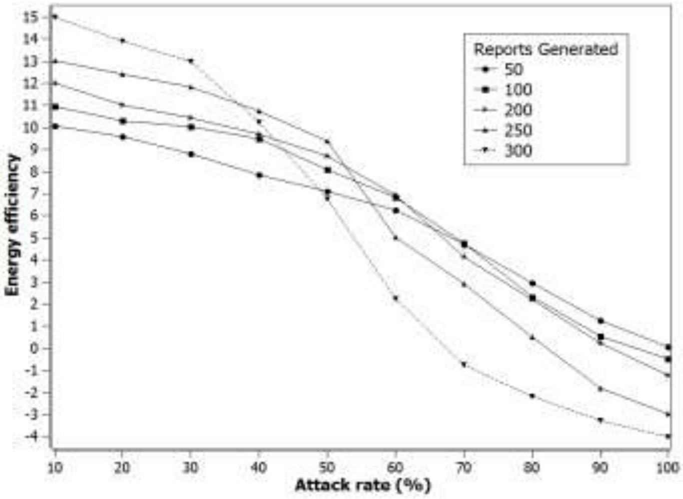

Figure 5: Energy Efficiency graph

Figure 5 illustrates the energy efficiency performance of our proposed scheme. The energy efficiency curves are with reference to the energy consumption in original PVFS. We can observer that when the attack rate is low and bigger number of reports are generated, the energy conserving efficiency is greater. However when the attacks are high and since the fuzzy system selects bigger *s* value, the energy efficiency degenerates owing to an increase in the attacks. The larger the attack ratio, the poorer is the energy saving. To shun negative values in the energy gain graph, we can put a cap on the maximum value of *s* because larger value of *s* means more MACs to be forwarded along with the report and higher energy consumption in forwarding the report which decreases the network lifetime.

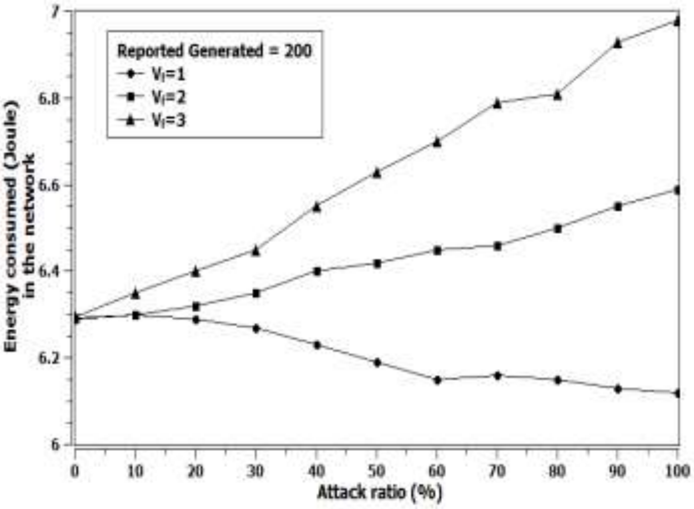

Figure 6: Energy consumption Vs.V_f Graph

Figure 6 show the effect of changing the value of V_f on the energy consumption behavior of the network when false data is forwarded towards the BS. When there is literally no attack, the energy consumption in all cases of V_f *is* same, however as the attack ratio increases, the difference in the energy consumption behavior also increases. False report consumes lesser energy when V_f =1 because they are immediately dropped after a filtering node detects single false MAC.

However, when V_f = 2 or higher, it help to avoid FMI attack and true reports reach the base station with higher probability. For V_f = 1, the filtering scheme only considers FRIA attack and completely ignores FMI attack.

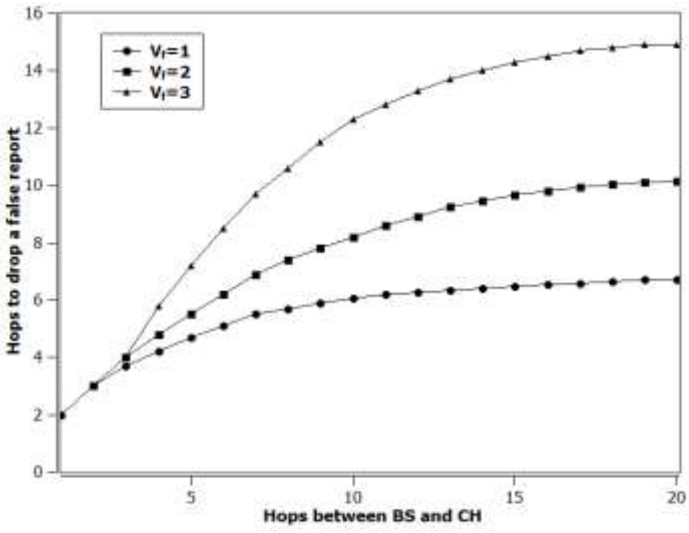

Figure 7: Distance travelled by a dropped false report

Figure7 shows the number of hops travelled by a false report against different of values of V_f. When V_f is smaller, the false report is dropped earlier, whereas when V_f is higher, the false report travels more distance before being filtered out by the filtering node. But V_f only becomes larger when there are more FRI and FMI attacks being launched by the adversaries. It is also noted that V_f =1 only detects false reports and true reports with just one false MACs are dropped which is not desirable. Thus the energy consumption in delivering a report only increases when the attack rate is high because the value of *s* as well as V_f increases and there are more MACs attached to the report. During lesser attack rates, the value of s as well as V_f is small and fewer MACs are attached to the report thus it reduces the overall size of the report.

8. Conclusion

In WSNs, Adversaries compromise sensor nodes and launch FRI attack to drain the energy resource of the network and FMI attack to cause legitimate reports be filtered out en-route. PVFS is the only scheme that offers to tackle both the threats at the same time. However, in PVFS, the number of MACs required to authenticate a report is decided before the network deployment. The fuzzy adaptive choice of attaching variable number of MACs to the report in PVFS can improve energy efficiency unless the attacks are frequent and high. Our scheme sustains the same security as offered by the PVFS against FMI and FRI attack. The energy statuses of the intermediate nodes, attack scenario and the pat-distance are important factors to consider while choosing suitable number of MACs in the network. In future, we aim to propose the fuzzy adaptive selection of routing paths to improve the network lifetime.

9. Acknowledgment

This research was supported by the Basic Science Research Program through National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. NRF-2015R1D1A1A01059484)

References

[1] Z. Yu, and Y. Guan, "A dynamic en-route scheme for filtering false data injection in wireless sensor networks," SenSys, vol. 5, pp. 294-295, November 2005.

[2] X. Han , X. Cao , E.L. Lloyd , C.C. Shen , Fault-tolerant relay node placement in heterogeneous wireless sensor networks, IEEE Trans. 9 (5) (2010) 643–656 .

[3] F. Ye, H. Luo, and S. Lu, "Statistical en-route filtering of injected false data in sensor networks," IEEE J. Selected Areas Commun., vol. 23, issue 4, pp. 839-850, April 2005.

[4] J. Xu , X. Zhou , J. Han , F. Li , F. Zhou , Data authentication model based on reed–solomon error-correcting codes in wireless sensor networks, IETE Tech. Rev. 30 (3) (2013) 191–199 .

[5] PVFS: A probabilistic voting-based filtering scheme in wireless sensor networks," Int. J. Security Netw., vol. 3, issue 3, pp. 173-182, January 2008.

[6] S.J. Lee, T.H. Cho, An en-route filtering scheme based on priority as determined by the fuzzy rule-based system, Int. J. Comput. Sci. Netw. Security 9 (7) (2009) 46–50.

[7] H.Y. Lee, T.H. Cho, Fuzzy adaptive selection of filtering schemes for energy saving in sensor networks, IEICE Trans. Commun. 90 (12) (2007) 3346–3353.

[8] A. S. Uluagac, R. A. Beyah, and J. A. Copeland, "Time-based dynamic keying and en-route filtering (TICK) for wireless sensor networks,"IEEE Global Telecommunications Conference (GLOBECOM 2010), pp. 1-6, December 2010.

[9] C. Kraub, et al. "STEF: A secure ticket-based en-route filtering scheme for wireless sensor networks," The second international conference on Availability, reliability and security, IEEE ARES, pp. 310-317, April 2007.

[10] H. Yang, and L. Songwu, "Commutative cipher based en-route filtering in wireless sensor networks," Vehicular Technology Conference, VTC2004, Vol. 2.pp. 1223-1227, September 2004.

[11] S. Zhu, S. Setia, S. Jajodia, and P. Ning, "An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks," In Proceedings of IEEE symposium on Security and privacy, pp. 259-271, May 2004.

[12] M.P. Sousa , M.T.A. Barros , W.J.L. Queiroz , W.T.A. Lopes , M.S. Alencar ,On the improvement of wireless sensor networks using modulation diversity and fuzzy clustering, International Workshop on Telecommunications (IWT11) Published in CD, 2011.

[13] M. Akram, and T.H. Cho, "Energy efficient fuzzy adaptive selection of verification nodes," Ad Hoc Networks, vol. 47, issue 1, pp. 16-25, September 2016.

Authors Profiles

Muhammad Akram received an M.S. degree in Software Engineering from Hamdard University, Karachi, Pakistan in 2012 and a B.S. degree in Computer Engineering from BUITEMS, Pakistan. He is currently a Ph.D. scholar in the College of Information and Communication Engineering at Sungkyunkwan University, Korea. His research interests include wireless sensor networks, genetic algorithms, context-aware computing and distributed systems.

Tae Ho Cho received a Ph.D. degree in Electrical and Computer Engineering from the University of Arizona, USA, in 1993, and B.S. and M.S. degrees in Electrical Engineering from Sungkyunkwan University, Korea and the University of Alabama, USA, respectively. He is currently a Professor in the College of Software at Sungkyunkwan University, Korea. His research interests include wireless sensor networks, intelligent systems, modeling and simulation, and enterprise resource planning.