Security Enhancement of First Mile Wireless Access with Optimal QoS through Anonymity in Convergence Network

Shankar R and Dananjayan P[*],

[*]Department of Electronics and Communication Engineering,
Pondicherry Engineering College,
Pondicherry.
pdananjayan@rediffmail.com

***Abstract*— The first mile wireless access of convergence networks is prone to various attacks due to the broadcasting feature of transmission. Most of the existing security measures focus on protection of message contents, leaving the header part in plain text. Therefore, the header part that has source and destination address, when transmitted in plain text, can reveal the information that can be useful for traffic analysis attack. In this paper, a master node generates pseudonym sets for slave nodes by using Unique Pair Sequence (UPS), assigns each set to a node and to itself. The assigned set and master set are transmitted to all the nodes. Each node uses an address randomly picked from the set as source and destination address while transmitting a message, for avoiding traffic analysis attack. For message integrity instead of allocating separate bits for Message Authentication Code (MAC) in a frame, an address selected from the address set, depending on the value generated by hash-keyed function, called Address Embedded Message Authentication Code (AMAC) is followed. Since AMAC does the function of the MAC and includes address, it can be used to transmit a message anonymously as well as to authenticate the message. AMAC, an anonymous scheme where additional bits are avoided is compared with MAC scheme using results obtained from simulating the scenarios for ZIGBEE and WiMax in OPNET 14.5. The simulation results show that AMAC performs better than MAC in terms of QoS metrics also.**

***Keywords*— Security, QoS, traffic analysis attack, Convergence networks.**

I. Introduction

Convergence networks or next generation networks are expected to be the evolution toward the integration of existing networks; security will be a weak point in the future as more networks are integrated. Especially the first or last mile, which is the final leg of delivering connectivity from a communication provider to a customer, wireless access to convergence networks such as WLAN or WPAN is inherently vulnerable to various security attacks due to the broadcasting features of wireless medium. To achieve the security goals, i.e., confidentiality and authenticity there have been proposed many countermeasures, which encrypt the contents of a message and attaches a message authentication code (MAC). All of these security countermeasures, however, are focusing on the protection of message contents except for the header part of the message which may cause a fundamental security problem. The address information in a message header is very useful to potential attackers who use it for a traffic analysis attack, which will deduce information from patterns in communication. That significantly threatens the communicator's privacy.

As a solution for this problem, anonymity in communication has been investigated, but most of the solutions are not practical. Many research efforts on anonymous communication have focused on the anonymity of the source and destination by hiding multi-hop routing information and preventing an attacker from tracking the source in multi-hop communications. In the future convergence networks, the last mile wireless access technology will be more essential since the cell coverage becomes smaller, that is less than few hundred meters. In this kind of single-hop wireless communication environment, the existing research results may not be applicable to guaranteeing the anonymity of the sender and the receiver.

In this paper, a novel scheme to guarantee the anonymity of communication parties in the last mile wireless access is proposed that exploits the concept of pseudonymity which is defined as the use of pseudonyms as IDs. Pseudonym will work as an identifier of a subject to be protected. It can be associated with a sender, a receiver, demanding protection. The main idea of the proposed scheme is to give a set of pseudonyms to each node in a network and randomly select one of the pseudonyms in the assigned set to use it as ID. In order to generate the pseudonym sets UPS [1] is used which is a randomly-generated sequence of numbers in which any two subsequent elements appear only once. The proposed scheme can provide two communication parties with the anonymity because each node uses one of unique pseudonyms to exchange messages.

The rest of the paper is organised as follows. Section II describes related work. Under section III, in the discussion of anonymous communication, Network model is presented, Unique Pair Sequence is illustrated and the proposed pseudonym assignment scheme is explained. Then it is shown how the proposed scheme works using an example. In section IV, two situations in which anonymity could be broken off is considered and the countermeasures are proposed. Section V describes the proposed AMAC scheme for authentication of message along with anonymity. Section VI elaborates the two, ZIGBEE and WiMAX, scenarios simulated in OPTNET 14.5 for comparing MAC and AMAC. Finally this work is concluded in section VII.

II. RELATED WORK

Many research results have been reported on anonymous communication. The Mix-net provides source and the destination in the Internet environment with anonymity and unlinkability by encrypting the routing information through public-private key scheme. It was followed by various researches on anonymous communication in wireless multi-hop network [2-9] as well as in the Internet [10]. However, all of them have focused on how to hide multihop routing [11] information for the anonymity between the source and the destination. Only [12] focused on the anonymity of a transmitter and a receiver in single-hop communications. The authors addressed the loss of anonymity in the link layer in contrast to existing researches that focused on the network layer anonymity. They proposed the encryption of address fields by using symmetric key scheme. Although this scheme can provide the anonymity between a transmitter and a receiver, a receiver may suffer from much decryption overhead since it happens that a receiver cannot easily know who the source is when packet transmission error incurs out-of-sync between the sender and the receiver.

III. ANANYMOUS COMMUNICATION

In order to prohibit the traffic analysis attack, the concept of pseudonymity is introduced. The main idea of this scheme is to give a set of pseudonyms to each node in a network as ID. A pseudonym set consists of a sufficient number of unique pseudonyms. If an address is picked randomly from the set and used each time a packet is transmitted, the attacker cannot guess the links between the messages. So traffic analysis attack can be avoided. The need for UPS and generation method of pseudonym sets is discussed in this section.

Let A be the number of bits allocated for address field in a packet and P = A/2 be the number of bits allocated for a node. So there are 2^P different combinations. These are randomly selected and divided among the nodes and have to be transmitted to them. If the addresses are chosen such that two consecutive addresses have half of the bits common in them then they can be grouped and transmitted to nodes. This reduces the memory required to store as well as addresses can be transmitted quickly.

First address 0 1 1 0 **1 0 1 1**

Second address **1 0 1 1** 1 1 0 1

They can be combined while sending and sent as a sequence as shown below

After combining 0 1 1 0 **1 0 1 1** 1 1 0 1

For the purpose of understanding and simplicity ‘P’ is split into two parts, and ordered pair is formed as shown below.

Address = 0 1 1 0 1 0 1 1 = (0110, 1011) = (6, 11) So (6, 11) (11, 13) (13, 7) are combined as (6, 11, 13, 7) to form UPS.

The ordered pairs are chosen such that second element of first ordered pair becomes first element of second ordered pair and so on. Then they are combined to form a sequence of numbers. Any two consecutive numbers can be used to get address. As every address is unique, any two consecutive numbers of the sequence must be unique and the sequence is called UPS.

A. Network Model

The message communication is considered in the first mile wireless access, where the network formation is the star topology with one master and many slaves. The slave nodes cannot communicate with each other directly and all the communication is through master node as shown in Fig.1. For example, in IEEE 802.11 WLAN network, several stations communicate with an access point. IEEE 802.15.4 LR-WPANs is composed of a PAN coordinator and devices in a star topology [13]. A master assigns a unique logical ID to each slave node. This logical ID becomes the node’s identifier in the network.

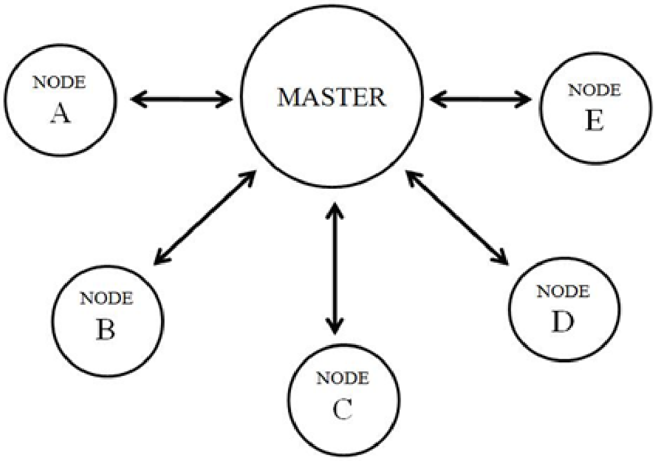


Fig.1 Star topology of a network

B. Unique Pair Sequence

Consider a sequence $\{u_1, u_2, u_3, \ldots\ldots\ldots, u_i, u_{i+1}, \ldots., u_m\}$

The above sequence is said to be a UPS if, $0 \le u_i \le S - 1$, $(u_i, u_{i+1}) \neq (u_j, u_{j+1})$ $\forall\, i \neq j$, $0 \le i, j \le m - 1$,

where,

S is a positive integer.

m is the length of the sequence generated

UPS obeys unique pair property because two subsequent elements which forms a unique pair occurs only once in a sequence.

1) UPS generation: Consider a positive number S. Then there exists S^2 unique pairs ranging from 0 to S-1 as shown in the Fig.2.

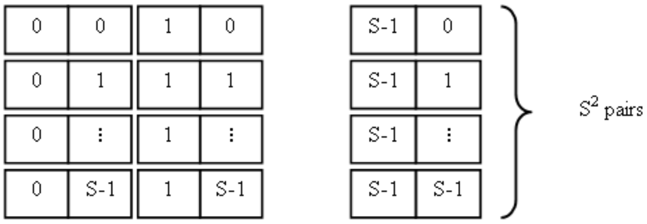


Fig.2 Set of possible pairs using S integers

Now randomly pick one pair from the above set. The second pair is selected such that the last element of the first pair matches with the first element of the second pair. This procedure continues until all possible pairs in the above set gets exhausted such that the sequence satisfies the unique pair property.

Consider an example as shown in the Fig.3, Here the pair (0,1) is randomly selected from S^2 pairs. Now the second pair is randomly selected from all the possible pairs starting with ‘1’ i.e. the pairs (1,1),(1,2),(1,3),……,(1,S-1),(1,S). In this case it is (1, 3). This procedure continues until there is no other pair to select.

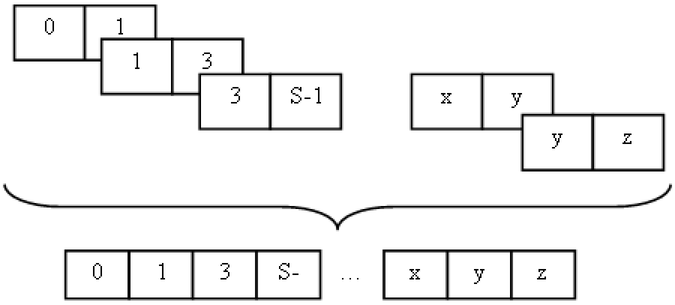


Fig.3 Generation of UPS

The last element of the preceding pair is grouped with the first element of the current pair i.e. the pairs (0, 1) & (1, 3) are grouped and written as {0, 1, 3}. This procedure is done for all other pairs. The sequence, as shown in Fig. 3, is an UPS.

2) Length of the UPS: Length of the UPS varies randomly according to the selection of pairs. The length of the UPS cannot be exactly known as it is a randomly-generated sequence. But the boundary of length of the UPS can be defined.

The best possible case is getting the maximum length and this happens when all the S^2 possible pairs are used up. So the maximum length of the UPS for a given value of S becomes S^2+1.The worst case exists when all the elements of a single row or a single column are used. So the minimum length of the UPS for a given value of S is 2S.

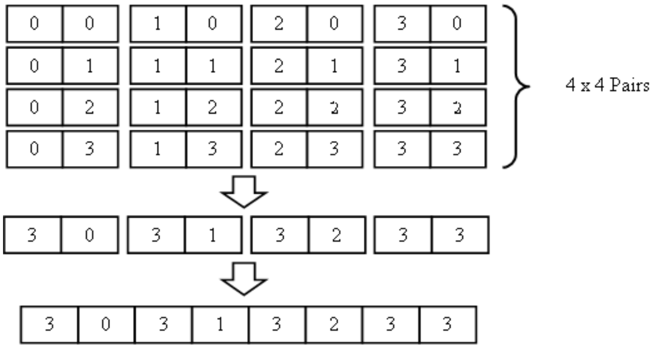


Fig.4 Example for UPS generation

Consider the value of S as 4 as shown in the Fig.4 All the elements are selected by randomly picking them as {1,2,0,3,2,1,0,2,2,3,0,0,1,3,3,1,1}, then the so formed sequence has the maximum length of 17 (S^2+1 = 4^2+1). The minimum length exists when only one single row or a single column is used. Thus, considering the column 4 the UPS generated is {3,0,3,1,3,2,3,3} which mean that {(3,0),(0,3),(3,1),(1,3),(3,2),(2,3),(3,3)} are picked. Hence the minimum length of sequence 8 (2xS = 2x4), is obtained.

3) Properties of UPS generated using MATLAB: An algorithm has been developed for generation of UPS and MATLAB code is developed to implement the algorithm. The code can be used to generate UPS for a given value of S. Then the properties of UPS were studied.

The variation of length of UPS for different trails for a value of S has been studied. The length varies from a minimum of 2S to a maximum of S^2+1. The plot in Fig.5 show the randomness in the length of UPS generated at different trials.

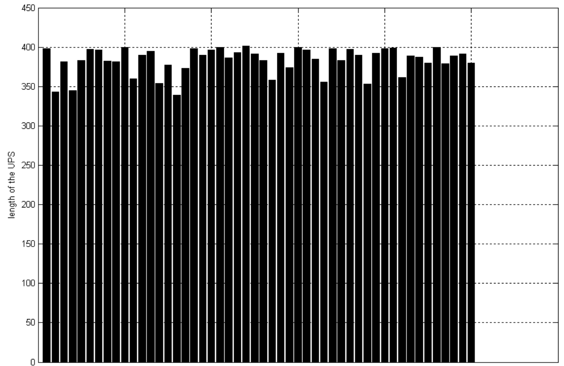


Fig.5 Randomness in the length of the UPS with every trial

This shows the random nature of the UPS. Since it is the sequence of integers following the condition that no two adjacent pair is repeated in the sequence, it is highly random. Not only the length of UPS varies with every trial, but also the sequence itself varies. Hence with every generation of UPS the set of addresses generated will vary. The addresses will be circulated among the address sets of each node. Thus the node will not get the same address set every time it enters the network.

The average length of UPS for different values of S and different trials is calculated. To show how far the average length is from the maximum possible length of UPS, the fraction of average simulated length to the maximum length (S^2+1) of UPS for that S value is considered. This fraction is plotted against the value of S as shown in Fig.6. From the plot it can be observed that as the value of S increases, the fraction tends closer to the value of unity. This means the average length of UPS simulated is closer to the maximum length possible for the given S value. Hence it has been inferred that as the value of S increases, the average length of the obtained UPS approaches the maximum length. Hence in such case more number of addresses can be generated for each node. With more number of addresses for each node the anonymity can be increased. It becomes more difficult for the attacker to guess the address of the nodes. Hence the choice of the value S which depends on the size of the address field can be selected such that it produces the UPS of length closer to the maximum length for the given S.

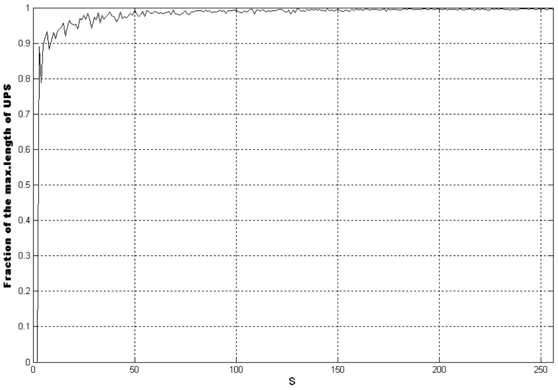


Fig.6 Fractional value of the average length of UPS simulated for different values of S in 50 trials

C. Pseudonym Assignment

The use of fixed address every time is the main cause for traffic analysis attack [14]. If a set of addresses are given to each node, then a node can use different address each time thus the traffic analysis attack can be avoided. These addresses are termed as secured addresses. The idea of the proposed scheme is to give a set of secured addresses as IDs to each node in a network. A secured address set consists of a sufficient number of addresses.

1) Pseudonym set pool: The master node generates a UPS for a value of S. Suppose the number of nodes in the network is N. Then the UPS is arranged in the form of N × K matrix, where K is an integer such that product of N and K is less than the length of UPS. The values of S, N and K are determined according to the network conditions such as the size of address fields in a message and the number of slave nodes which a master should be able to accommodate. Thus the matrix obtained is called Pseudonym Set Pool (PS Pool) and each row of PS Pool is called Pseudonym Set (PS) and is given below.

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_{K-1} & u_K \\ u_{K+1} & u_{K+2} & \cdots & u_{2K-1} & u_{2K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{(N-1)K+1} & u_{(N-1)K+2} & \cdots & u_{NK-1} & u_{NK} \end{bmatrix}$$

All the nodes including a master should have logical address as their own identity in the network. A master node assigns the logical address to each slave node entering the network. In the proposed scheme, each node gets a PS as a logical address. The master randomly picks a PS from PSPool and uses the PS as its own logical address. In the same manner, the master assigns a randomly-selected PS to each slave node, which corresponds to the assignment of a logical address. The master's PS and the PS corresponding to each node are sent to all the nodes by the master. However, the PS itself is precisely not a logical address. Each node uses one pseudonym of the PS as its logical address since all two subsequent elements of each PS can be unique pseudonyms for the corresponding node. If a node wants to transmit a message to a master, the node fills the sender's address of the message with the pseudonym which is randomly selected from its own PS, the receiver's address with the pseudonym randomly selected from the master's PS. Since pseudonyms in a message are changed randomly, an attacker cannot get any information from the pseudonyms in the eavesdropped message.

Since an UPS holds the unique pair property, a PSPool and each PS also holds the same property. For the proposed scheme, each node randomly picks one of assigned secured addresses and uses it as its own logical address

2) Example: Consider the value of S as 5 and number of nodes in the network as 5, the master generates UPS of length 25. Then PSPool is formed by arranging the UPS as a 5 x 5 matrix. Each row forms a PS and is assigned to one node each as shown in the Fig. 7. After assignment, each node has two PS, its own and the master's PS. Each pair in the PS is a secured address that is used as address.

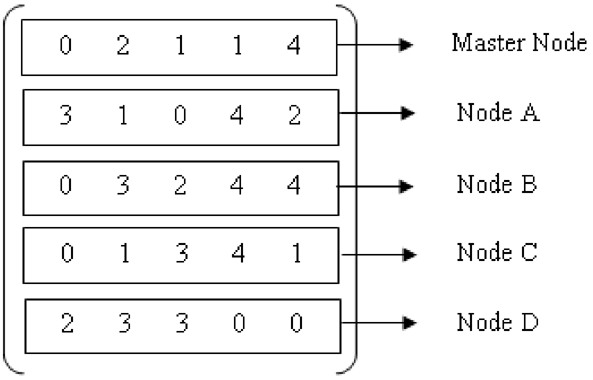


Fig.7 Example of the PS assignment

Suppose node B transmits a message to the master. It randomly selects one secured address from its PS and uses it as source address. Then it chooses one secured address from the master's PS and uses it as the destination address for the message it wants to transmit. For example, let it pick (3, 2) from its PS and pick (1, 4) from master's PS. Thus the source and destination address of the message would be (3, 2) and (1, 4) respectively. Once the message is transmitted, the other nodes check the secured address in the destination. None of the nodes, except the intended receiver, would find the secured address (1, 4) in their PS and thus they discard the message. The intended receiver (the master) checks the source address and looks for the secured address (3, 2) in the PSs of other nodes to find the sender. Even though an attacker looks at the header of the message, attacker does not know the nodes that are communicating as PS is not known.

D. Application to IEEE 802.15.4 MAC layer

In this subsection, the scheme is mapped to the IEEE 802.15.4 MAC layer and its feasibility is shown. The IEEE 802.15.4 networks define the star topology, in which there is a master and several slave devices. The master node, a PAN coordinator assigns a 16 bit logical address to each device. The proposed scheme is applied for address allocation. The network conditions are the size of address field A bits in a message and the number of slave nodes N that a master should be able to accommodate.

The source and destination address field occupies A/2 bits each. In the proposed scheme, each secured address consists of a pair of non-negative integers, each of which is smaller than S. Thus, the value of S determined from 2 $\log_2$ S is A/2.

Since A = 32, for the network considered, S becomes equal to 28. The length of the UPS can be obtained from the expression 0.9x (S^2+1). The value approximately becomes 58983. If the number of nodes N=256, the longest length of the PS is calculated as follows:

The length of PS =   230

Thus, the memory size of each PS is 230 bytes (the length of the PS x $\log_2$ S is 230 x $\log_2$ 28). As each slave node keeps its own PS and the master's PS, it requires only 460 bytes of memory. The master requires memory size of just 58 Kbytes.

IV. Consideration of Anonymity

Since the pseudonyms in a message are meaningless to those nodes which do not have the corresponding PS, this scheme can guarantee the anonymity of the sender and the receiver. The node which is intended to receive the message only can know the address as it possesses the senders address, in this case the master's PS. Hence even if the attacker can read the address of source and destination, it is difficult to point out the node which corresponds to this address. But the attacker might guess the source and destination address with some probability even though the exact source and destination could not be found.

A. Exposure of PSPool

The master generates the UPS once and assigns the fixed PSs to the nodes that enter. If a node leaves and joins a network frequently, for a long time, it gets to know the whole PSPool used in the network.

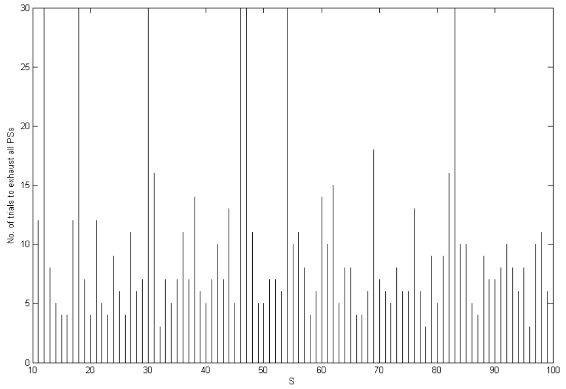


Fig.8 Plot showing the number of entries that a node has to make to gain complete knowledge of PSPool for different values of S

Consider a case where there are four nodes in a network. The master node would generate four PSs and allocate it to each node. Suppose a node leaves the network, its PS will be idle and it would be given to any other node that joins the network. If the node that left joins the network again it would be given a new PS. In this way, if it leaves and enters the network repeatedly, it can gain the knowledge of the PSPool within few entries. The Fig. 8 shows this case, where the y-axis shows the number of entries required for the node to exhaust all the PSs. It is plotted against the different values of S. It can be inferred that this is independent of the value of S. Moreover, in most of the cases, the node requires four or less number of entries to gain complete knowledge of the PSPool.

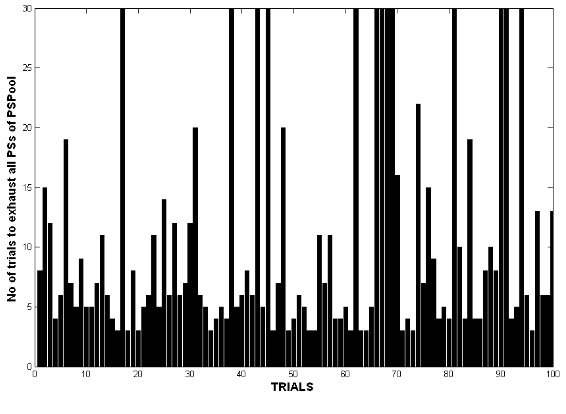


Fig.9 Plot showing the number of entries a node has to make to gain complete knowledge of PSPool in different trials

As in the previous case, the simulation is run for different trials and for the same S value. From Fig.9, it can be observed that in this case also, the node requires only four or less number of entries into the network to gain the complete knowledge of the PSPool. From the results obtained, it can be inferred that if the attacker enters the network repeatedly, the PSs can be known easily. This helps the attacker in identifying the two different addresses that correspond to the same PSs even if it is some other node's PS. This proves to be a threat to privacy and security. Hence a modified PSPool is suggested.

At the network initialisation stage the master generates the e.K – length UPS, where the value of e is the number of extra PSs which are necessary to remove the delay for the additional assignment of PSs for joining nodes. Then the initial PSPool is formed as follows

$$\begin{bmatrix} \overset{1}{V}_1 & \overset{1}{V}_2 & \cdots & \overset{1}{V}_{K-1} & \overset{1}{V}_K \\ \overset{2}{V}_1 & \overset{2}{V}_2 & \cdots & \overset{2}{V}_{K-1} & \overset{2}{V}_K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \overset{e}{V}_1 & \overset{e}{V}_2 & \cdots & \overset{e}{V}_{K-1} & \overset{e}{V}_K \end{bmatrix}$$

When a new node joins, the master assigns the PS from the extra PSs and lengthens the UPS by adding K elements. It generates extra PS from the lengthened UPS and appends it to the PSPool. Suppose n nodes have joined continuously, the UPS would be $\{u_1, u_2, u_3 \ldots u_{nK}, u_{nK+1} \ldots u_{(n+e)K}\}$

Thus the PSPool becomes

$$\begin{bmatrix} \overset{1}{V}_1 & \overset{1}{V}_2 & \cdots & \overset{1}{V}_{K-1} & \overset{1}{V}_K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \overset{n}{V}_1 & \overset{n}{V}_2 & \cdots & \overset{n}{V}_{K-1} & \overset{n}{V}_K \\ \overset{n+1}{V}_1 & \overset{n+1}{V}_2 & \cdots & \overset{n+1}{V}_{K-1} & \overset{n+1}{V}_K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \overset{n+e}{V}_1 & \overset{n+e}{V}_2 & \cdots & \overset{n+e}{V}_{K-1} & \overset{n+e}{V}_K \end{bmatrix}$$

If the node using PS_i leaves, the master removes that particular PS from the PSPool and deletes K-subsequent elements from the UPS. Even though the unique pair property of the UPS is lost at this moment, the PSPool still possess the unique pair property and the master can distinguish each node by its secured address.When the node B, shown in the Fig.10, leaves the network the PS corresponding to the node B is deleted from the sequence and also from the PSPool.

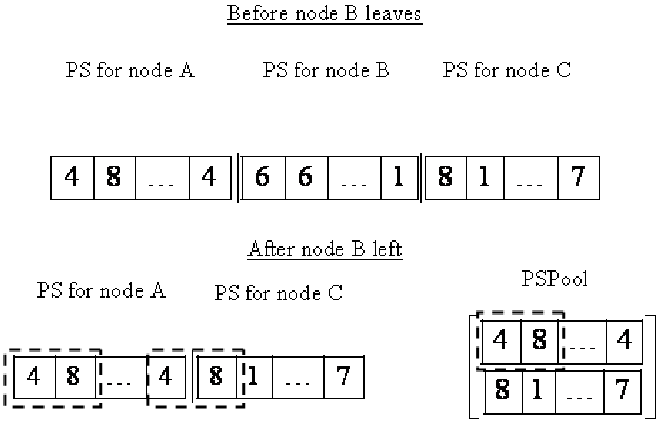


Fig.10 Unique pair property of PSPool

In such case, the pair (4, 8) occurs twice in the sequence as shown below. Even though the unique pair property of the UPS is lost, the unique pair property is satisfied by the PSPool. The pair (4, 8) is unique in the PSPool as each row corresponds to a PS. Thus the property is still held by the PS.

B. Guessing of Links

This is the case where the attacker might guess the links between different pseudonyms and thereby relate them to any particular node. Since each node uses more than one address as its identity, it becomes difficult for an attacker to identify the node exactly. Each node randomly selects from a set of addresses known as PS. It is difficult for the attacker to exactly point out the node that corresponds to any address. If the number of messages in a network is very few at any instant, the attacker can make certain guess to link the addresses used in those messages.

Consider a scenario when one of the nodes has sent a message which can be represented, [1, 2 || 3, 4 || contents], to the master and then the master replied with acknowledgement, [6, 3 || 5, 7 || ack]. When there are no other messages in a network, an attacker can guess that someone uses (1, 2), (5, 7) and another uses (3, 4), (6, 3) as their address. The source address (1, 2) in message and the destination address (5, 7) in the acknowledgement correspond to the same node. Similarly the destination address (3, 4) in the message and the source address (6, 3) in the acknowledgement correspond to another node. Thus one can guess that there is a link between these two secured addresses. But one does not know exactly who sent the message to whom. So anonymity is still valid, although an attacker guessed the links. In the case of a secured network even guessing of these links cannot be allowed. To minimise the probability of this guessing a method should be introduced. To prevent even this guessing between links the number of messages in a network at any instant should not be very low.

Hence the concept of garbage messages is opted. At any instant *t*, if there are less than n_{th} (threshold) messages, then n_g garbage messages are transmitted by the master. This threshold messages are very limited so that the resources are not wasted. The contents of a garbage message have no information and the secured address is randomly chosen from extra PSs available with the master. When minimum number of messages in a network at any instant is high, it is difficult for an attacker to guess the links. In this way the probability of guessing the links is reduced.

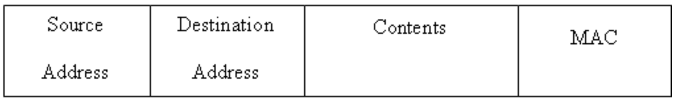
V. AMAC SCHEME

A. Proposed AMAC Scheme

The AMAC scheme can guarantee anonymous communication in wireless access networks. The main idea of this scheme is to integrate the addresses with the MAC. Since the value of the MAC randomly varies according to each sender-receiver pair and message contents, only the sender and intended receiver can obtain the same value of the MAC. However, the MAC, itself, cannot work as an identifier since the value is meaningless and not unique. As a solution for this problem, UPS that makes the value of the MAC meaningful is introduced. An address selected from the address set, according to the value of MAC, is called the Address-embedded Message Authentication Code (AMAC). This code can be substituted for both the addresses and the MAC as in Fig. 11.

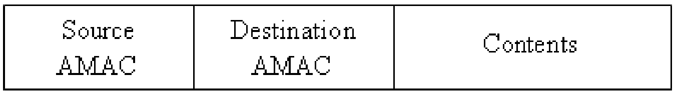
Since AMAC includes the value of the MAC and an implicit unique address, it can be used to transmit a message anonymously as well as to authenticate the message.

Legacy message format



Source Address	Destination Address	Contents	MAC

New message format using proposed AMAC



Source AMAC	Destination AMAC	Contents

Fig. 11 Comparison of MAC scheme with AMAC scheme

1) Properties of AMAC: The objective of the scheme is to guarantee the privacy of a sender and a receiver in wireless access networks through an anonymised address. In order to achieve this goal, the anonymised address must be generated to guarantee anonymity under the following requirements:

- Randomness: The address should be generated in random manner being meaningless to unintended receivers.
- Uniqueness: The anonymised address must indicate each communicator uniquely. Otherwise, the address cannot work as an identifier.
- Variableness: A constant address for each node may provide attackers with useful clues for some kinds of attacks. Thus, the address should vary packet by packet.
- Low computational overhead: Each communicator should be able to compute the real value of the address with low computational cost so that it can be applicable to every message.
- Low decryption overhead: The address encryption may cause a high number of decryptions at the receiver side. These large numbers of decryptions can bring some drawbacks such as a quick exhaustion of the device power or long decryption delay. Therefore, all receivers should be able to determine the sender and the intended receiver with the fewest number of decryptions.

2) *Use of AMAC for authentication:* In AMAC pool generation, a network master generates a UPS with a certain value of S. It can easily make an N × K matrix by extracting K subsequent elements from the UPS N times, such that N × K = *a* × Maximum Length. In previous section, the simulation results show the length of the UPS with S larger than 128 is longer than 0.9×Maximum Length. Therefore, the value of *a* assumed is 0.9. As for the values of S, N, and K, it is described how these values affect the security level and the network capacity in the next section. This N×K matrix is defined as the Address-embedded Message Authentication Code Pool or in simply the AMAC Pool This AMAC Pool is made from the UPS $\{u_1, u_2, u_3, \cdots, ,\cdots ,u_a$*Maximum Length$\}$, similar to the PSPool illustrated in the previous section.

ACS assignment

Let all nodes including a master should have a logical address as their own identity in the network. A master node assigns a logical address to each slave node entering the network. The master uses a randomly-selected ACS from the AMAC Pool as its own logical address. In the same manner, the master assigns a randomly-selected ACS to each slave node, corresponding to the assignment of a logical address. The master sends its ACS as well as each node's ACS to each slave node to share them.

AMAC generation at sender

After the ACS assignment, each slave node has two ACSs, one for the master and the other for itself. It is assumed that each node has a key_{hash} which is a pre-shared hashing key between the master and the slave node. All nodes have three kinds of functions

- H (key, contents) : A keyed-hash function such as MD5 or SHA-1. The contents are the contents of a message. The output is an integer from 1 to K
- E (ACS_i, j) : An AMAC extraction function from ACS.
- F (ACS_i , (α,β)) : Source identification function

The output of E(ACS_i, j) is Address-embedded Message Authentication Code (AMAC). The following three steps are a sender's procedure when the sender using ACS_s sends a message to a receiver using ACS_d.

STEP1: Make the receiver's AMAC, $AMAC_D$.

$$AMAC_D = E(ACS_d , H(key, contents))$$

STEP2: Marks the sender's AMAC

$$AMAC_S = E(ACS_s , H(key, contents))$$

STEP3: Places $AMAC_D$ and $AMAC_S$ into address field.

From the above three steps, the format of the message to be sent can be shown as [$AMAC_D$ || $AMAC_S$ || Contents]. Note that there are no explicit addresses. $AMAC_D$ and $AMAC_S$ belong to only the receiver's and sender's ACS, respectively. Therefore, all nodes receiving the message can distinguish the message's destination and source. Also, the AMACs can be used for authenticating the message because the AMAC generation process includes a conventional HMAC generation process. Thus the AMAC works as an address and a conventional MAC at the same time. It can provide anonymous communication between a transmitter and a receiver. Any nodes, which do not have corresponding ACSs, cannot know who is communicating with whom.

Verification of message at receiver

On receiving the message in the air, all nodes, including the intended receiver, would check whether or not their own ACS has the value of $AMAC_D$ to decide if the destination of the message is itself. If a node's ACS contains the value of $AMAC_D$, the node is the destination of the message. Due to the property that the value of the $AMAC_D$ is unique in the

AMAC Pool, every node can distinguish its own message from others. While other nodes discard the message, only the intended receiver tries to authenticate the message [15]. The authentication procedure at the intended receiver is:

Step 1: Source Identification

Find ACS_i such that $F\ (ACS_i, AMAC_S) = 1$

Suppose the node using the ACS_i to be the source.

Step 2: Message Authentication

Re-generate two AMACs.

$AMAC'_D = E\ (ACS_{oneself}, H(key, contents))$

$AMAC'_S = E\ (ACS_{supposed}, H(key, contents))$

Compare $AMAC'_D$, $AMAC_D$ and $AMAC'_S$, $AMAC_S$

$ACS_{oneself}$ and $ACS_{supposed}$ are the ACSs of the receiver and the supposed source, respectively. If the generated AMACs are the same as the received AMACs, the receiver can be convinced that this message is intact and from an authentic node. By using the AMACs, the anonymity of the source and destination nodes are also guaranteed since there are no explicit addresses although a message is eavesdropped in wireless networks.

False positive ratio

False positive ratio (Fpr) is the probability that a receiver misjudges an illegal message as a valid one. The AMAC Pool should be constructed while satisfying the given network requirements because Fpr depends on how the AMAC Pool is formed. A network should be able to support up to N nodes and the acceptable Fpr for a message with AMAC is F. A possible attack scenario is that an attacker eavesdrops a message in the air, extracts AMACs from the message, and then broadcasts an illegal message that consists of the eavesdropped AMACs and a malicious payload. Then Fpr = 1/ Number of unique pairs in an ACS.

Consider there are K elements in the ACS. The probability that a malicious message yields the same address pair would be 1/ $(K-1)^2$. In order to meet the network requirement, the condition is

$$\frac{1}{(K-1)^2} \le F \qquad (1)$$

The value of K should be selected such that

$$1+\sqrt{\frac{1}{F}} \le K \qquad (2)$$

Also a master should have at least N ACSs because the network should support N nodes. Therefore, there needs to be a UPS where the length is at least N.K to make an N x K matrix such that

$$N.K \le a.\text{MaximumLength} = a(S^2+1) \qquad (3)$$

The approximate value of S can be obtained from

$$N.\left(1+\sqrt{\frac{1}{F}}\right) \le a\left(S^2+1\right), \qquad (4)$$

$$\left(\frac{1}{a}.N.\left(1+\sqrt{\frac{1}{F}}\right)-1\right)^{\frac{1}{2}} \le S \qquad (5)$$

To satisfy the network requirements (2) to (5), in this paper, the value of the effective length coefficient (a) is assumed to be 0.9. The AMAC field size of the two AMACs ($AMAC_D$ || $AMAC_S$) can be calculated. Each should be able to represent two non-negative integers smaller than S. Thus, the AMAC field size becomes $2(2.\log_2 S)$ bits. Fig.12 shows the field size in a message when the number of supportable nodes, N and acceptable fpr F are given. For example, in the case that N is 512 and F is 10^{-4}, at least 20 bits for the AMAC field should be allocated. In other words, it can authenticate a message with 10^{-4} fpr while supporting 512 nodes by using only 20-bit address fields without an additional MAC.

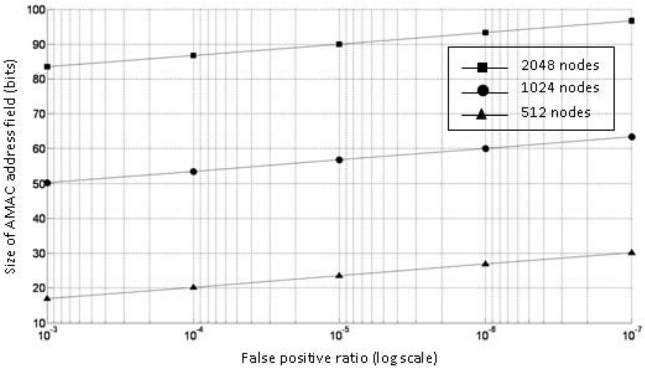


Fig.12 Size of AMAC address field (bits) varying with fpr values for different number of nodes

In a similar way the fpr value can be related to the value of S chosen for generating UPS. Fig.13 shows the plot between fpr and the S value. It can be seen that the maximum value of S to be chosen increases exponentially with the decrease in the value of fpr. From this it is inferred that for better fpr values the higher S value should be selected. The variation in the required S value can also be seen with the change in the number of nodes in a network.

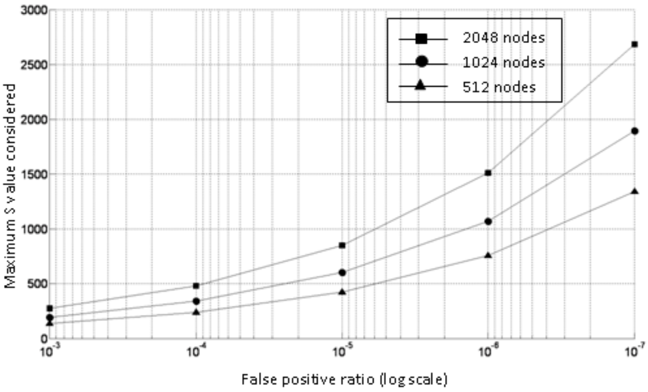


Fig.13 Maximum S value considered varying with fpr values for different number of nodes

VI. COMPARISON OF AMAC WITH MAC SCENARIOS IN OPNET

As mentioned before, AMAC scheme avoids the use of extra bits which are attached to the message in case of MAC scheme. In order to understand the nuances in both the schemes, two scenarios are simulated in OPNET 14.5; the QoS parameters are compared in both the cases. The networks like Zigbee and Wimax which have star topology have been considered for simulation of the above mentioned scenarios. The plots shown below compare QoS parameters like delay, load, etc. for both the cases.

A) Zigbee Network Model

Zigbee network is simulated as shown in Fig.14. The network is formed with a coordinator connected to two end devices.

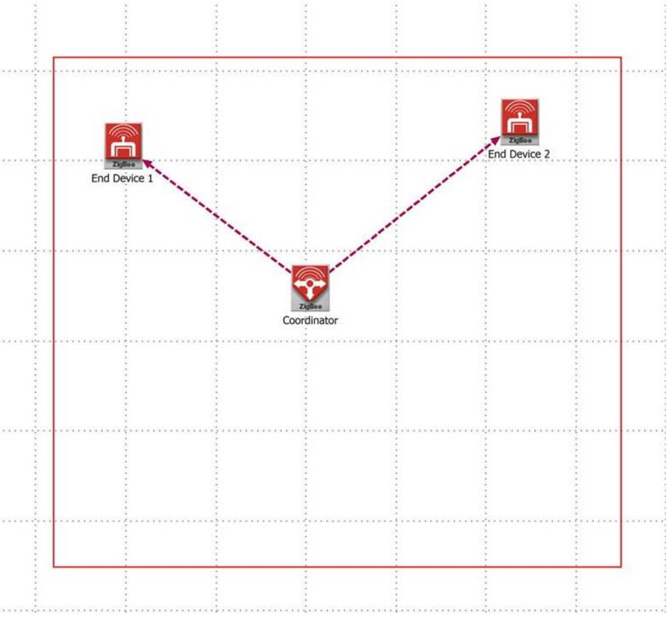


Fig.14 Zigbee network model in OPNET

The coordinator acts as the master node and the end devices act as the slave nodes. Here, three scenarios were simulated. One is the case of AMAC in which the message is transmitted without attaching any additional bits.

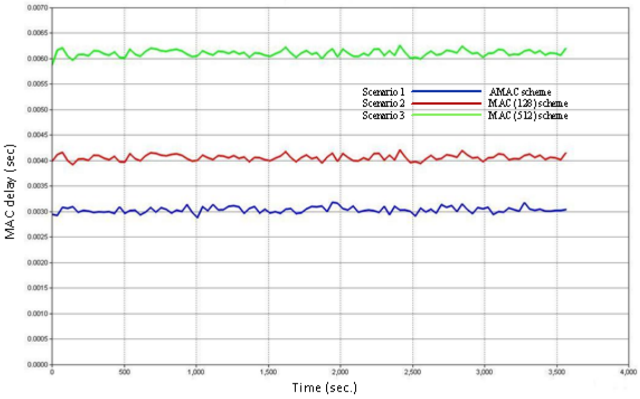


Fig.15 MAC delay comparison for the three scenarios

The remaining two are MAC cases where in one scenario 64 bits are attached and in the other one 128 bits are attached. The MAC delay represents the end to end delay of all the packets received by the 802.15.4 MACs of all WPAN nodes in the network and forwarded to the higher layer. The load represents the total load (in bits/sec) submitted to 802.15.4 MAC by all higher layers in all WPAN nodes of the network.

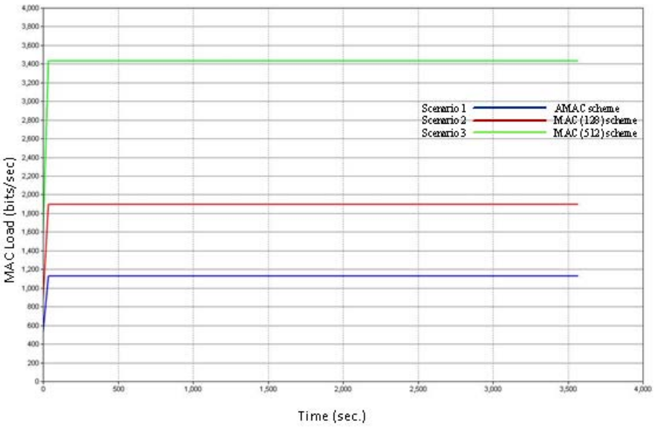


Fig.16 MAC load comparison for the three scenarios

B) WiMaX

A Wimax network model is simulated as shown in Fig.17. The base station at the center acts as the master node. It is surrounded by five mobile stations which act as slave nodes. In this case two scenarios are simulated. One is of the AMAC scheme while the other one is of MAC scheme with 128 additional bits.

Time (sec.)

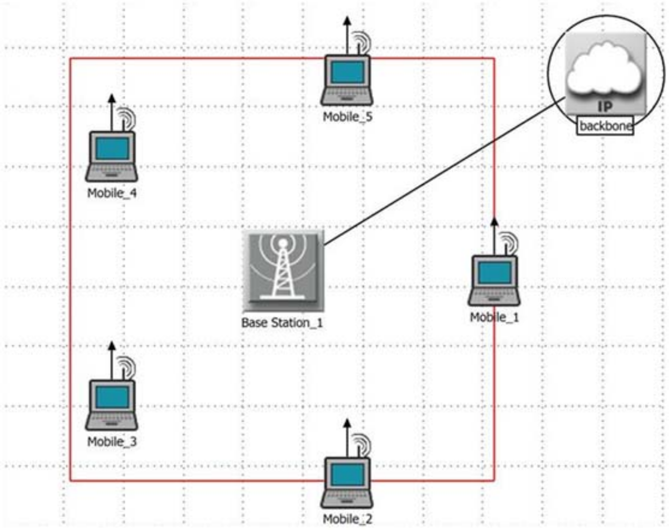


Fig.17 WiMaX network model in OPNET

The plot in Fig.18 shows the delay in the two scenarios. The delay, here, represents the end-to-end delay of all the packets received by the WiMAX MACs of all WiMAX nodes in the network and forwarded to the higher layer. The plot in the Fig.19 compares the average traffic received by the nodes in the two scenarios. Traffic received is the data traffic successfully received by the WiMAX MAC from the physical layer in packets/sec. While computing the size of the received packets for this statistic, the physical layer and MAC headers of the packet are also included.

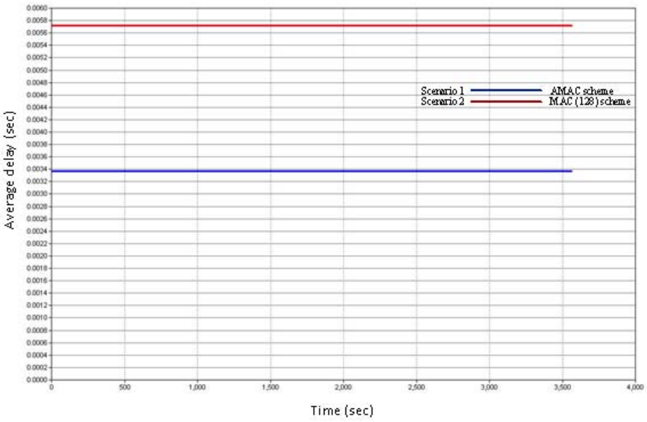


Fig.18 Delay comparison for the two scenarios

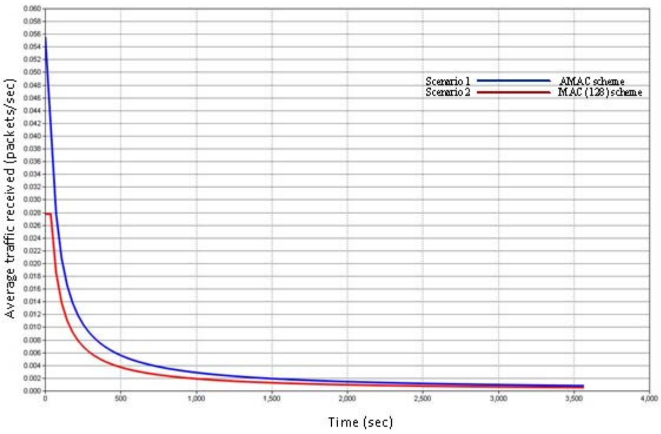


Fig.19 Average traffic received in terms of packets/sec compared for the two scenarios

From Fig.18 it can be observed that the MAC scheme produces more delay when compared to the AMAC case. In Fig.19, the average traffic received, in terms of packets/sec, by the nodes is more for AMAC case when compared to the MAC case. This is because the packet size is small for AMAC case when compared to the MAC scheme. Hence more packets can be transferred in case of AMAC. This increases the throughput for AMAC scheme.

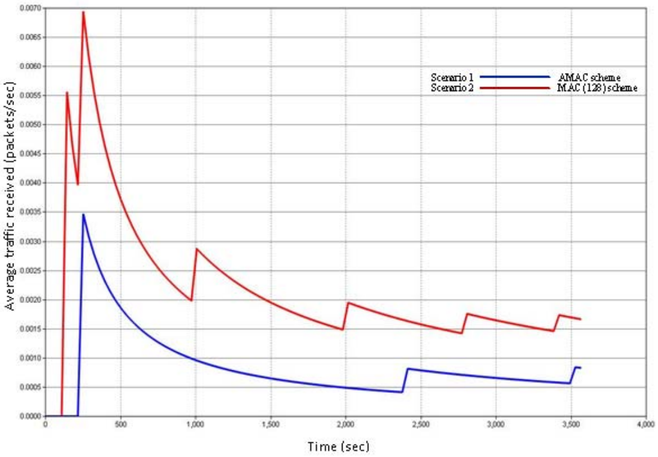


Fig.20 Average uplink packets dropped is compared for the two scenarios

The plot in Fig.20 compares the number of packets dropped per second for the two scenarios. This statistic records the uplink packets dropped (in packets/second) due to physical layer impairments. For SS nodes, this statistic represents the packet drops measured at a BS for all packets arriving from a particular SS. For BS nodes, this statistic represents the packet drops measured at the BS for all packets arriving from all SS nodes in the cell/sector.

The comparisons in all the above results show that the QoS parameters like the delay, load and the traffic received and dropped, etc. for AMAC scheme are better when compared to those of the MAC scheme.

VII. CONCLUSION

Pseudonym assignment scheme in first mile wireless networks enhances the security by providing anonymous addressing, which therefore avoids the traffic analysis attacks.

Each node is given with a set of unique pseudonyms (addresses). So, while transmitting a message, a node chooses one of the addresses from its PS and uses it as source address and chooses one from the receivers PS and uses it as destination address. The node which possesses the PS pool containing the destination address can only receive the message. This makes it difficult for the attacker to trace the node that is sending a message by mere looking at the address field. Thus the anonymity of the nodes prevents the possible attack. In order to provide this set of unique addresses, UPS was generated. Then the master node generates PSPool and assigns PSs to all the nodes. A master node contains the PSs of all the nodes, where as slave node contains PS of master and its own.

There are cases where this anonymity could be broken namely exposure of PSPool and guessing of links. The exposure of PSPool occurs when a node continuously joins and leaves a network. Every time the node enters the network it is given a PS different from the one it used before. This can be avoided by deleting the PS of the node as soon as when the node leaves the network and by adding extra PS to the PSPool. The second case where the anonymity is broken is the guessing of links. This occurs when there is no communication in the network except between two nodes. If only two packets are present in a network, an attacker can easily guess the addresses of the two nodes. In order to prevent this garbage messages are introduced in the network. Though these garbage messages do not carry any information, they use the pseudonyms from the extra PSs. So, the attacker thinks that many nodes are communicating and hence cannot guess the links from the messages in the network.

This pseudonym scheme also provides authentication without using the extra MAC bits, by using AMAC. This AMAC schemes provides the same function as MAC bits. A node instead of randomly selecting the address from the PS uses hash function. This hash function generates a value based on the contents of the message, which is used to select the address from the set. At the transmitting side the node selects the source and destination addresses from the respective PSs by using hash function and transmits the message. If both the addresses of the received AMAC and that of calculated AMAC match then it can be inferred that the message is sent to the intended receiver and also that there are no errors in the message. it is obvious that the message is intended to it and the message integrity is maintained. This hash function which is used for selecting the addresses depends on the contents of the message. In the network all the nodes check for whether the destination address is present in their PSs .If the destination address is present then it accepts the message or else discards it. Hence the integrity of the message is thus achieved by using this AMAC scheme

The performance of this AMAC scheme is calculated by using fpr. This fpr in turn depends on the value of S which is used for the generation of UPS. Then the AMAC scheme is compared with the MAC scheme for QoS parameters. The scenarios are simulated for Zigbee and Wimax in OPNET. From the results it can be inferred that the QoS parameters for the AMAC scheme are better when compared to those of the MAC scheme.

The proposed pseudonym scheme enhances the security in the first mile wireless access through anonymity. When the scheme is used along with hash function, the resulting AMAC scheme provides authentication apart from the anonymity. Since this AMAC scheme avoids the use of extra MAC bits, it provides security and authentication with optimal QoS.

REFERENCES

[1] M.-H. Park and S.-W. Seo, “A pseudonym assignment for the last mile wireless access to 4G networks”, *Proceedings of the IEEE Global Telecommunications Conference*, Washington, DC, USA. November 26-30, 2007.

[2] J. Kong and X. Hong, “ANODR: Anonymous on demand routing with untraceable routes for mobile ad-hoc networks”, *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Annapolis, Maryland, USA, pp.291-302, June 1-3, 2003.

[3] Azzedine Boukerche, Khalil El-Khatib, Li Xu and Larry Korba, “SDAR: A secure distributed anonymous routing protocol for wireless and mobile ad hoc networks”, *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, Tampa, FL, USA, pp.618-624 ,16-18 November 2004.

[4] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, “Enhancing source-location privacy in sensor network routing”, *Proceedings of 25th IEEE International Conference on Distributed Computing Systems* (25th ICDCS’2005), Columbus, Ohio, USA, pp.599–608, June 6-10, 2005.

[5] Khalil El-Khatib, Larry Korba, Ronggong Song and George Yee, “Secure dynamic distributed routing algorithm for ad hoc wireless networks”, *Proceedings of the 32nd International Conference on Parallel Processing Workshops*, Kaohsiung, Taiwan , pp.359-366, 6-9 October 2003.

[6] Ronggong Song, Larry Korba, and George Yee, “AnonDSR: Efficient Anonymous Dynamic Source Routing for Mobile Ad hoc Networks”, *Proceedings of the 2005 ACM Workshop on Security of Ad Hoc and Sensor Networks in conjunction with the 12th ACM Conference on Computer and Communications Security*, Alexandria, Virginia, USA. pp.32-42. November 7-11, 2005.

[7] S. Basagni, K. Herrin, D. Bruschi, and E. Rosti, “Secure pebblenets”, *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing*, New York, USA, pp.156–163, October 2001.

[8] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, “Enhancing source-location privacy in sensor network routing”, *Proceedings of 25th IEEE International Conference on Distributed Computing Systems* (25th ICDCS’2005), Columbus, Ohio, USA, pp.599–608, June 6-10, 2005.

[9] C. Shields and B. N. Levine, “A protocol for anonymous communication over the internet”, *Proceedings of the 7th ACM Conference on Computer and Communications Security*, Athens, Greece, pp.33–42, November 1-4, 2000.

[10] O. Berthold, H. Federrath, and M. Kohntopp, “Project Anonymity and Unobservability in the Internet”, *Proceedings of the tenth conference on Computers, freedom and privacy*, New York, USA, pp.57–65, 2000.

[11] Ulrich Herberg and Thomas Clausen, “Security issues in the optimized link state routing protocol version 2 (olsrv2)”, International Journal of Network Security & Its Applications , Volume 2, Number 2, April 2010.

[12] Frederik Armknecht, Joao Girao, Alfredo Matos, and Rui L. Aguiar, “Who said that? privacy at link layer”, *26th Annual IEEE Conference on Computer Communications*, Anchorage, Alaska, USA, May 2007.

[13] Wireless medium access control and physical layer specifications for low-rate wireless personal area networks. IEEE Standard, 802.15.4-2003, May 2003.

[14] Zulfa Shaikh, Tonu Sojatia, Pushpa Pathak and Shilpa Bhalerao, “Analysis of data and traffic management during privacy preservation in secure multiparty computation”, International Journal of Computer Science and Information Technologies**,** pp.1-5, Vol. 1 (1), 2010.

[15] Kin Choong Yow and Amol Dabholkar, “A light-weight mutual authentication and key-exchange protocol based On elliptical curve

cryptogaphy for energy-constrained devices", International Journal of Network Security & Its Applications (IJNSA), Volume 2, Number 2, April 2010.

R. Shankar received Bachelor of Engineering in 2001 and Master of Technology in 2006 in Electronics and Communication Engineering (ECE) from Bharathidasan University, Trichy and Pondicherry Engineering College, Pondicherry, India respectively. He is pursuing his Ph.D. programme in the Department of ECE, Pondicherry University. He is currently working as Assistant Professor in the Department of ECE at Sri Manakula Vinayagar Engineering College, Pondicherry, India. His research interests include wireless communication, computer networks and convergence network.

P. Dananjayan received Bachelor of Science from University of Madras in 1979, Bachelor of Technology in 1982 and Master of Engineering in 1984 from the Madras Institute of Technology, Chennai and Ph.D. degree from Anna University, Chennai in 1998. He is working as a Professor, Department of ECE, Pondicherry Engineering College, India. He is also visiting Professor to Asian Institute of Technology, Thailand. He has more than 60 publications in National and International Journals. He has presented more than 130 papers in National and International conferences. He has guided 9 Ph.D candidates and is currently guiding 6 Ph.D students. His areas of interest include spread spectrum techniques and wireless communication, wireless adhoc and sensor networks.