Mobile Agent-Based Cross-Layer Anomaly Detection in Smart Home Sensor Networks Using Fuzzy Logic

Muhammad Usman, Vallipuram Muthukkumarasamy, *Member*, IEEE, and Xin-Wen Wu, *Senior Member*, IEEE

Abstract — *Despite the rapid advancements in consumer electronics, the data transmitted by sensing devices in a smart home environment are still vulnerable to anomalies due to node faults, transmission errors, or attacks. This affects the reliability of the received sensed data and may lead to the incorrect decision making at both local (i.e., smart home) and global (i.e., smart city) levels. This study introduces a novel mobile agent-based cross-layer anomaly detection scheme, which takes into account stochastic variability in cross-layer data obtained from received data packets, and defines fuzzy logic-based soft boundaries to characterize behavior of sensor nodes. This cross-layer design approach empowers the proposed scheme to detect both node and link anomalies, and also effectively transmits mobile agents by considering the communication link-state before transmission of the mobile agent. The proposed scheme is implemented on a real testbed and a modular application software is developed to manage the anomaly detection system in the smart home. The experimental results show that the proposed scheme detects cross-layer anomalies with high accuracy and considerably reduces the energy consumption caused by the mobile agent transmission in the poor communication link-state situations*[1].

Index Terms — Smart Home Sensor Networks, Mobile Agent, Anomaly Detection, Fuzzy Logic, Cross-Layer Design.

I. Introduction

The recent advancements in microsensor technology [1], [2], have realized the concept of the *smart home* envisaged in the last century [3]. The underlying device interconnection paradigm, namely, Wireless Sensor Network (or Smart Home Sensor Network in this case) connects sensing devices to set up a smart home [4], [5]. Typically, in a smart home sensor network, the sensor nodes sense their ambient environment or target objects and then transmit the readings to a central node managed by a user through custom-built application software. The sensor nodes and their transmitted data are, however, susceptible to *in situ* and *in transit* anomalies. A software mobile agent-based anomaly detection scheme in such situations not only detects anomalies in a smart home sensor network, but also offers an automated service to verify the source of anomalies, before notifying a user about the anomalies [6], [7]. The in situ verification of a sensor node, which is believed to be a malicious node after receiving an anomalous observation, is carried out by the mobile agent by comparing the values of the received data with the stored values on the node. Over the course of the years, the research community has also investigated the roles of mobile agents for the random sampling of sensed data over the network, and sharing of the network control and anomaly information in networks [8], [9]. In the case of random sampling, mobile agents are randomly dispatched for the inspection of nodes. On the other hand, mobile agents are programmed to perform the collaborative exchange of anomaly and network related information in the latter case. Previous studies [6]-[9], however, do not consider the link-state between the communicating nodes before the transmission of the mobile agents. A poor communication link-state may cause errors in the data (or code) of the mobile agent during transmission, which may ultimately affect its designated functionality.

The previous mobile agent-based anomaly detection schemes have defined crisp boundaries on sensed data in order to characterize the behavior of sensor nodes [6]-[9]. The use of crisp logic for anomaly detection may result in unnecessary generation of alarms in situations, when the values of the received data lie close to the normal profile bounds. Consider, for example, a smart home scenario, where a sensor node is designated to measure and report the room temperature, with normal behavior bounded by the closed interval [15ºC, 20ºC]. In this case, a value close to 15ºC or 20ºC, such as 14.9ºC or 20.3ºC, will generate an unnecessary alarm. In such situations, fuzzy logic can be beneficial to define soft boundaries for decision making [10]. However, an anomaly detection scheme which characterizes the behavior of sensor nodes using only fuzzy logic is unable to consider the stochastic variability of the data to build the normal profile. Furthermore, the previous mobile agent-based anomaly detection schemes have not considered the communication link-state for anomaly detection and mobile agent transmission [6]-[9].

To address the above-stated limitations, this study has introduced a mobile agent-based cross-layer anomaly detection scheme. The proposed scheme employs statistical

[1]M. Usman is with the School of Information and Communication Technology, Griffith University, Gold Coast, QLD 4222, Australia. (e-mail: muhammad.usman3@griffithuni.edu.au).

V. Muthukkumarasamy is with the School of Information and Communication Technology, Griffith University, Gold Coast, QLD 4222, Australia. (e-mail: v.muthu@griffith.edu.au).

X.-W. Wu is with the School of Information and Communication Technology, Griffith University, Gold Coast, QLD 4222, Australia. (e-mail: x.wu@griffith.edu.au).

Contributed Paper
Manuscript received 03/31/15
Current version published 06/29/15
Electronic version published 06/29/15.

0098 3063/15/$20.00 © 2015 IEEE

procedures which consider the stochastic variability in the cross-layer data to define three regions, namely, normal, tolerance, and anomalous, over the cross-layer feature space. The normal region defines the normal behavior of a sensor node. The tolerance region is defined to account for those observations which lie close to the normal region. The proposed method decrements the trust value of a node if an observation from that node falls within the tolerance region, and the user will only be notified when the trust level falls below a predefined threshold. The mobile agent is transmitted for the in situ verification of the sensor node to verify the source of anomalies [6], [7], only if an observation falls in the anomalous region or the trust value reaches the lower bound. The soft boundaries between the tolerance and anomalous regions and the fuzzy logic-based rule-base are designed to detect cross-layer anomalies and to effectively transmit mobile agents. The proposed scheme is implemented on a real testbed and results indicate its ability to detect cross-layer anomalies with high accuracy, and to increase the network longevity.

The main contributions of this study are following: (i) a regions computation method, based on statistical procedures, is proposed to define different regions for decision making about anomaly detection and mobile agent transmission, (ii) a fuzzy logic-based cross-layer rule-base is designed and a corresponding algorithm is presented to detect cross-layer anomalies and transmit a mobile agent after due consideration of the communication link-state, and (iii) the proposed methods are implemented on a real testbed and an application software is developed to manage the proposed anomaly detection system in smart homes.

This paper is organized as follows: Section II elucidates the network model and architecture of the proposed cross-layer anomaly detection module. The details of the proposed scheme are described in Section III. The corresponding algorithm is presented in Section IV. The experiment set up, details of the application software, and results are discussed in Section V. Finally, conclusions are drawn in Section VI.

II. Network Model and Proposed Anomaly Detection Architecture

A. Network Model

The WSN is assumed as a digraph, which can be formally defined as **G** = (**V**, **E**), where **V** represents the vertices (i.e., the sensor nodes) and E denotes the edges (i.e., the communication links) in a smart home sensor network. The nodes $\mathbf{V} = \bigcup_{i=1}^{3} V_i$ create the smart home sensor network, where V_1 is a laptop node which works as a central network authority and is connected with *m* number of resource rich cluster head nodes, i.e., $V_2 = \{v_1, v_2, ..., v_m\}$. The nodes $V_3 = \bigcup_{j=1}^{m} V_{3j}$ form *m* number of clusters, where $V_{3j} = \{v_j, s_{j1}, s_{j1}, ..., s_{jk}\}$. The notation V_{3j} denotes the *j*th cluster in the network, v_j represents the cluster head node in that cluster and, k is the number of member sensor nodes in the cluster. The cardinality of the node sets must hold the relation $|V_1| \le |V_2| \le |V_3|$ to create a hierarchical

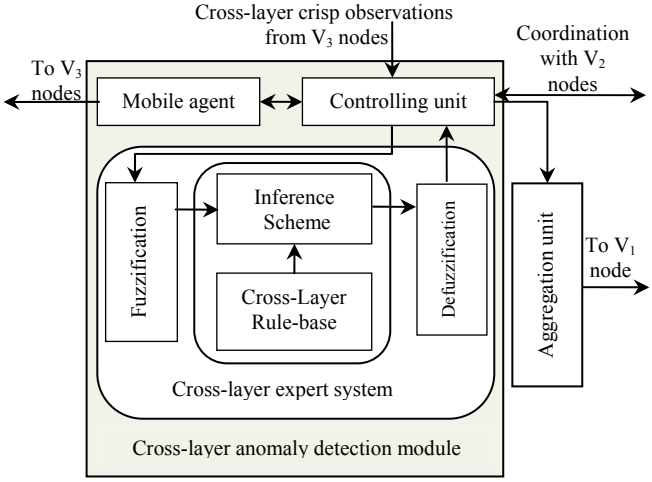

Fig. 1. Architecture of the cross-layer anomaly detection module.

smart home sensor network, where V_1, V_2, and V_3 are the top, intermediate, and leaf level nodes, respectively.

The sensor nodes which belong to the V_2 and V_3 types are IEEE 802.15.4-compliant motes. The V_2 type nodes are resource rich, as they are equipped with additional memory and continuous power supply. On the other hand, the V_3 type nodes have limited memory and battery resources. The V_3 type nodes are deployed to sense their environment, store sensed data and battery status in their memories for the in situ verification process, and then transmit those measurements to the corresponding V_2 type node. The V_2 type nodes detect cross-layer anomalies on the received data packets and transmit the mobile agents for the in situ verification after due consideration of the communication link-state.

B. Architecture of Cross-Layer Anomaly Detection Module

Each V_2 type node is equipped with a cross-layer anomaly detection module, which detects cross-layer anomalies and also performs the function of mobile agent transmission after considering the communication link-state. The cross-layer anomaly detection module consists of three logical components, namely, Controlling Unit, Mobile Agent, and Cross-Layer Expert System, as depicted in Fig. 1.

1) Controlling Unit

The controlling unit acts as a coordinator among the internal logical components of the cross-layer anomaly detection module and other entities of the network such as peer V_2 type nodes and the V_1 node to coordinate the anomaly detection and mobile agent transmission processes.

The controlling unit receives data packets from member V_3 type nodes and passes them to the cross-layer expert system, which performs the tasks of the cross-layer anomaly detection and mobile agent transmission, and sends back the result to the controlling unit.

The normal sensor reading is forwarded to the aggregation unit, which stores it for a predefined period of time and then transmits it to the V_1 node for further processing. In the case of an anomalous observation, the controlling unit triggers a mobile agent in order to carry out the in situ verification of the V_3 type node to identify the source of the anomalies.

2) Mobile Agent

The mobile agent uses the values obtained from the previous data packets to carry out the task of the in situ verification on the V_3 type node. The mobile agent performs a match between both the stored values of the battery status and sensor readings with the values of the battery status and sensor readings which are brought by the mobile agent to perform the in situ verification. If the values are matched, then the V_3 type node is considered to be normal. Otherwise, the anomalous status of the node is reported to the corresponding V_2 type node. For further details of the in situ verification process, readers are referred to the previous studies [6], [7].

3) Cross-Layer Expert System

A general fuzzy expert system fuzzifies crisp input data into fuzzy data and process them using a set of rules to obtain fuzzy output data [12]. Fuzzy output data is then defuzzified to obtain a crisp output value which causes the execution of the predefined corresponding action. The cross-layer expert system receives crisp values of the cross-layer features from the controlling unit and fuzzifies them using the membership functions presented in Section III-B. Then fuzzified input is processed for the cross-layer anomaly detection and mobile agent transmission decision making by the fuzzy logic cross-layer rule-base described in Section III-C. Finally, the defuzzification unit defuzzifies the fuzzy output by employing the *maximum* method, i.e., by selecting the value which has the maximum fuzzy membership function value.

III. The Proposed Scheme

A. Cross-Layer Feature Set

The behavior of the V_3 type (i.e., IEEE 802.15.4-compliant sensor) nodes is characterized by node and communication link features. The node features are those features whose values are transmitted by V_3 type nodes to their corresponding V_2 type nodes. The node features include Sensor Reading (SR) and Battery Status (BS). The sensor readings may include, but are not limited to the measurements of temperature, pressure, and motion detection.

The communication link-state is characterized by three features, namely, Link Quality Indicator (LQI), Received Signal Strength Indicator (RSSI), and Packet Error Rate (PER) for anomaly detection and mobile agent transmission decision making. The values of the communication link features are extracted by the V_2 type node from the received data traffic of a V_3 type node. The mote has an IEEE 802.15.4/ Zigbee ready Radio Frequency (RF) transceiver chip which has 250 Kbps data rate and an adjustable transmission power [11]. The RF transceiver chip computes RSSI and average correlation (CORR) values of each received packet to determine the LQI value. The value of CORR indicates the raw link information within the closed interval [50, 110], from the worst to the best case values, respectively. This study has considered CORR = LQI to derive LQI values [13]. This shows that the sensor nodes do not need to perform any additional computation to compute the values of the RSSI and LQI features to make anomaly detection and mobile agent transmission decisions.

The values of the PER feature are important for the correct execution of the in situ verification process, as the values of the SR and BS features obtained from the received data packets are later used for the verification process. If the errors in the received data packets are ignored, they may lead to an incorrect result of the in situ verification process. Therefore, only those packets which pass the 16-bit cyclic redundancy check (CRC) are considered as successfully received [13]. The PER is computed as the number of successfully received data packets over the number of total transmitted data packets.

B. Regions Computation

The limited available energy budget of sensor nodes demands careful transmission of mobile agents. The cross-layer anomaly detection module, therefore, partitions the feature space of every cross-layer feature of V_3 type nodes into three regions, namely, normal, tolerance, and anomalous. The normal region defines the normal behavior of sensor nodes. If the values of the cross-layer features of the received data packets do not lie within the normal region, but in its close proximity, then it would not be appropriate to immediately transmit a mobile agent to carry out the in situ verification process due to the energy expensive nature of the communication operation [14]. The cross-layer anomaly detection module considers this region to be a tolerance region and decrements the trust value of the V_3 type node after receiving the data packets with the values in the tolerance region. The mobile agent is only transmitted after the V_2 type node loses trust in the V_3 type node to a certain degree. The data packets having values outside the tolerance region are treated as anomalous and a mobile agent is immediately transmitted to carry out its designated task. The generalized method for regions computation for a single cross-layer feature is described below. Note that the proposed scheme computes these regions independently for all cross-layer features.

Formally, let X be the Universe of Discourse (UoD), representing the feature space of a single cross-layer feature of

a V_3 type node, where $X = \{N, T, A\}$. In the UoD X, the fuzzy numbers N, T, and A denote normal, tolerated, and anomalous regions, respectively. The domains of these fuzzy numbers are defined below.

$$N = [c^*,\ d^*]$$
$$T = [a^*,\ c^*] \cup [d^*,\ f^*]$$
$$A = (-\infty,\ b^*] \cup [e^*,\ +\infty)$$

In the above definitions, $a^* = a \pm s^l$, $b^* = b \pm A_r^l$, $c^* = c \pm (s/\sqrt{n})^l$, $d^* = d \pm (s/\sqrt{n})^r$, $e^* = e \pm A_r^r$, $f^* = f \pm s^r$, and the parameters a^* to f^* must satisfy the relation $a^* \le b^* \le c^* \le d^* \le e^* \le f^*$ in order to define the domains of the fuzzy numbers. The symbol s represents the standard deviation of the n number of sampled observations which are used to compute the regions. The notation A_r denotes the anomalous region bound. Note that the superscripts l and r do not represent the power, instead these are the left and right side values of the parameters along the horizontal or x-axis. The left side value of a parameter is calculated by the subtraction, whereas the right side value is computed by the addition of the statistic value (obtained through a statistical procedure) from/ to the mean value of the feature. The variables a to f are the user defined adjustment variables which are used to adjust the values of corresponding parameters to update the computed regions. The values of the adjustment variables are independent of the values of the parameters which are derived through the statistical procedures. The values of the bounds of the domains, in the above definitions of the fuzzy numbers, are determined through the statistical procedures applied on the n sampled observations. The normal region, defined by the fuzzy number N, is computed through the standard deviation of the mean of n observations, i.e., $s/\sqrt{n}$. Then the left and right sides of the mean ($\bar{x}$) along the x-axis are bounded by the values $c^* = c \pm (s/\sqrt{n})^l$ and $d^* = d \pm (s/\sqrt{n})^r$, respectively, to define the normal behavior of the V_3 type node.

Similarly, the boundaries of the tolerance region, defined by the fuzzy number T, are computed by the calculation of s value on n observations. Based on this computation, the bounds $[a^* = a \pm s^l,\ c^* = c \pm (s/\sqrt{n})^l]$ and $[d^* = d \pm (s/\sqrt{n})^r, f^* = f \pm s^r]$ define the left and right tolerance regions, respectively. Finally, the anomalous region is derived through the computation of the following formulas.

$$A_r^l = \frac{(s/\sqrt{n})^l + s^l}{2} \tag{1}$$

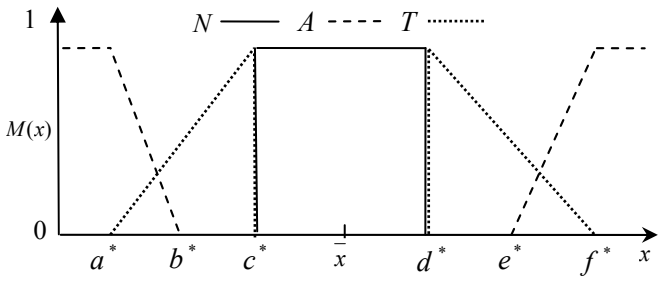

Fig. 2. Illustration of the membership functions.

$$A_r^r = \frac{(s/\sqrt{n})^r + s^r}{2} \tag{2}$$

Equations (1) and (2) define the upper bound of the left and lower bound of the right anomalous regions, respectively.

Based on the above computations, the bounds $(-\infty,\ b^* = b \pm A_r^l]$ and $[e^* = e \pm A_r^r,\ +\infty)$ define the domains of the left and right anomalous regions, respectively, as shown along the x-axis in Fig. 2.

Based on the values of the parameters, the membership function of the fuzzy number N is defined as

$$M_N(x) = \begin{cases} 1, & c^* < x < d^*, \\ 0, & x \le c^*, x \le d^* \end{cases} \tag{3}$$

Next, the membership function of the fuzzy number T is computed as

$$M_T(x) = \begin{cases} 1, & x = c^*, x = d^*, \\ \dfrac{(x - a^*)}{(c^* - a^*)}, & a^* \le x \le c^*, \\ \dfrac{(f^* - x)}{(f^* - d^*)}, & d^* \le x \le f^*, \\ 0, & x < a^*, x > f^*, c^* < x < d^* \end{cases} \tag{4}$$

Finally, the membership function of the fuzzy number A is derived by

$$M_A(x) = \begin{cases} 1, & x < a^* or\ x > f^*, \\ \dfrac{(b^* - x)}{(b^* - a^*)}, & a^* \le x \le b^*, \\ \dfrac{(x - e^*)}{(f^* - e^*)}, & e^* \le x \le f^*, \\ 0, & b^* < x < e^* \end{cases} \tag{5}$$

The realization of the membership functions of the fuzzy numbers N, T, and A is shown in Fig. 2. Note that the fuzzy number N has a crisp-valued membership function, which is a special case of the fuzzy membership functions. This design rationale is chosen to empower the cross-layer expert system to decrement the trust value of the V_3 type node as soon as the values of the observations of cross-layer features start falling in the regions defined by the fuzzy

number T, even if they are very close to the boundary of the fuzzy number N.

Example 1. Consider, for example, a scenario where a V_3 type node in a smart home senses its ambient environment and reports the temperature sensor readings to the corresponding V_2 type node. Let X be the UoD for the temperature sensor readings, n = 50, $\bar{x}$ = 20.10, s = 3.39, and $a = b = c = d = e = f$ = 0. This implies $s/\sqrt{n}$ = 0.48. Thus, the normal region can be defined as $[c^* = c \pm (s/\sqrt{n})^l$ =19.62, $d^* = d \pm (s/\sqrt{n})^r$ = 20.58]. Next, the tolerance regions can be demarked as $[a^* = a \pm s^l$ =16.71, $c^* = c \pm (s/\sqrt{n})^l$ = 19.62] and $[d^* = d \pm (s/\sqrt{n})^r$ = 20.58 $f^* = f \pm s^r$ = 23.49]. Finally, the anomalous regions can be computed as $(-\infty,\ b^* = b \pm A_r^l$ = 18.17] and $[e^* = e \pm A_r^r$ = 22.03, $+\infty)$. Note that the values of the adjustment variables (i.e., a to f) are set as 0 in this example. In practice, however, a user can adjust these values to update the computed regions. Furthermore, the membership values can be assigned using (3) to (5).

C. The Cross-Layer Rule-Base

The cross-layer expert system is instrumented with the cross-layer rule-base which processes the received data traffic to make decisions about anomaly detection and mobile agent transmission. The cross-layer rule-base has *IF antecedent(s), THEN consequent(s)* rules, where antecedents have five input linguistic variables (i.e., cross-layer features), namely, Sensor Reading (SR), Battery Status (BS), Link Quality Indicator (LQI), Received Signal Strength Indicator (RSSI), and Packet Error Rate (PER). These input linguistic variables are connected through the AND logical operator. The term-set of each input linguistic variable has three values: N, T, and A, as defined as fuzzy numbers in Section III-B. It is pertinent to mention that the granularity of the term-set can be increased or decreased as per the discretion of the user in order to tune the performance of the system.

The consequent (i.e. output linguistic variable denoted by D) has three decision values, namely, D_1, D_2, and D_3, where D_1 denotes the decision of the aggregation of the sensed data for the case when the received data is found normal, D_2 causes decrement in the trust value, and D_3 triggers the mobile agent to perform the task of the in situ verification. The D_1, D_2, and D_3 types of decisions have the triangular-shaped membership functions specified by the three parameters (t_l, t_m, t_r), where t_l, t_m, and t_r are the left, middle, and right values along the x-axis. The parameters of the decision variables take the following values: D_1 = (0, 0.2, 0.4), D_2 = (0.3, 0.5, 0.7), and D_3 = (0.6, 0.8, 1). An important design characteristic of the rule-base is that the rules execute action D_3 even if the value of only one feature is anomalous. The formal syntax of the first rule in the rule-base, as an illustration, is given in the below equation.

$$(SR = N_{SR}) \wedge (BS = N_{BS}) \wedge (LQI = N_{LQI}) \wedge (RSSI = N_{RSSI}) \wedge (PER = N_{PER}) \rightarrow D = D_1 \quad (6)$$

Semantically, in the antecedent part of the above rule, the N_{SR}, N_{BS}, N_{LQI}, N_{RSSI}, and N_{PER} are the normal values taken by the input linguistic variables SR, BS, LQI, RSSI, and PER, respectively. The consequent part causes the aggregation of the sensed data. The number of the input linguistic variables is 5 in the proposed method and each variable can take 3 values. Thus, the total number of rules, with all possible combinations, is 243. The general structure of the complete rule-base is shown in TABLE I.

TABLE I
CROSS-LAYER RULE-BASE

Rule No.	SR	BS	LQI	RSSI	PER	D
1	N_{SR}	N_{BS}	N_{LQI}	N_{RSSI}	N_{PER}	D_1
2	T_{SR}	N_{BS}	N_{LQI}	N_{RSSI}	N_{PER}	D_2
3	A_{SR}	N_{BS}	N_{LQI}	N_{RSSI}	N_{PER}	D_3
.	.	.	.	.	.	.
242	T_{SR}	A_{BS}	A_{LQI}	A_{RSSI}	A_{PER}	D_3
243	A_{SR}	A_{BS}	A_{LQI}	A_{RSSI}	A_{PER}	D_3

IV. THE ALGORITHM AND ANALYSIS

A. The Algorithm

The proposed algorithm runs on the resource rich V_2 type nodes after receiving the data traffic from the member V_3 type nodes. The algorithm has two phases, namely, Initialization Procedure and Main Procedure. The Initialization Procedure is responsible for the computation of the regions. It is first executed at the time of the system deployment and then only executes if the user wishes to recompute the regions and update the rule-base. The Initialization Procedure takes the values of the n number of sampled observations of the cross-layer features along with the value of the n itself as input to compute the parameters $\bar{x}$, $(s/\sqrt{n})^l$, $(s/\sqrt{n})^r$, s^l, s^r, A_r^l, and A_r^r. The regions are then computed using these values and the membership functions are defined by employing (3) to (5). The user-defined heuristic rule-base is generated after the execution of the first phase.

The Main Procedure performs the functions of the anomaly detection and mobile agent transmission by employing the rule-base. This phase executes after receiving every data packet from the member V_3 type nodes. In this phase, the cross-layer anomaly detection module extracts the crisp values

of the cross-layer features, namely, SR, BS, LQI, and RSSI from the received data packets and also computes the value of PER. These values are then fuzzified using the membership functions defined in (3) to (5) and processed by the cross-layer rule-base. This is followed by the defuzzification of the decision variable, using the *maximum* method (as discussed in Section II-B-3), to execute actions, namely, aggregation of the sensor reading, decrement in the trust count of V_3 type node, or transmission of the mobile agent. The pseudocode for the proposed methods is given in Algorithm 1.

Algorithm 1: *Cross-Layer Anomaly Detection and Mobile Agent Transmission*

Phase 1: Initialization Procedure

Input n sampled observations and value of n

for *SR, BS, LQI, RSSI, PER* **do**

Step 1: Compute: Eset = {*E*(1), *E*(2), *E*(3), *E*(4), *E*(5), *E*(6), *E*(7)} // where $E(1)=\bar{x}$, $E(2)=s^l$, $E(3)=s^r$, $E(4)=(s/\sqrt{n})^l$, $E(5)=(s/\sqrt{n})^r$, $E(6)=A_r^l$, $E(7)=A_r^r$

Step 2: *EstReg* (Eset) // estimate regions for all features

Step 3: *ConstructMemb* $M_N(x), M_T(x), M_A(x)$ // construct membership functions for all features

end for

Output Membership functions

Phase 2: Main Procedure

Input *DatPkt* // data packet
RlBs // cross-layer rule-base

for each *DatPkt* **do**

Step 1: *GetVal* (*SR, BS, LQI, RSSI, PER*)

Step 2: Fuzzify: Fuzzset={*fuzz*(*SR*), *fuzz*(*BS*), *fuzz*(*LQI*), *fuzz*(*RSSI*), *fuzz*(*PER*)} // using equation (3) to (5) for every feature

for Fuzzset={*fuzz*(*SR*), *fuzz*(*BS*), *fuzz*(*LQI*), *fuzz*(*RSSI*), *fuzz*(*PER*)} **do**

Step 3: *EvalRlBs*(Fuzzset) // evaluate rule-base

end for

Step 4: *DefuzzDes*(D_1, D_2, D_3) // defuzzify decision

Step 5: **if** *D* = = *D1* **then** // checking decision

Step 6: *Agg*(*SR*) □ *Str* (*SR, BS, LQI, RSSI, PER*) // aggregate SR and store values of all features

Step 7: **else if** *D* = = *D2* **then** // check decision

Step 8: *DecrTrst*(*Tr*) // decrement trust value

Step 9: **else** *Trnsmt MA* // transmit mobile agent

end if

end for

Output Aggregate *SR* and store *SR, BS, LQI, RSSI, PER*, *NewTrust*, or transmit *MA*

B. Computation Complexity

Proposition 1. (i) The computation cost of the Phase 1 of the proposed algorithm is $O(n)$ and (ii) Phase 2 is $O(n^2)$.

Proof. (i) The processes of the computation of the values of the statistical parameters, regions estimation, and construction of the membership functions take constant time for each process for n number of cross-layer features. Thus, considering $n(1+1+1)$ as the total complexity, the computation cost of the Phase 1 is $O(n)$.

(ii) Phase 2 takes constant time to perform each of the processes, namely, obtaining values of the cross-layer features from the received data packets, fuzzification, defuzzification, and decision making processes on n features. Next, n time is taken by Phase 2 to process n number of rules. Thus, the computation complexity of Phase 2 is $O(n^2)$ ■

Note that the above proofs have the relation $O(n^2) > O(n)$, because of the fact that Phase 2 involves the processing of the cross-layer rule-base, which is a computationally expensive operation as compared to the rest of the tasks in the algorithm.

V. PERFORMANCE EVALUATION

The performance of the proposed scheme is examined in terms of the detection accuracy, energy and memory consumptions, and processing time estimation.

As a proof of the concept, the proposed scheme is implemented on a real testbed based on the two sensor nodes and a laptop node, representing a minimal working smart home sensor network. One mote was deployed as a V_3 type node, which was responsible for sensing its ambient environment and reporting the temperature readings to the resource rich V_2 type node. The V_2 type node was responsible for anomaly detection, aggregation and then transmission of the aggregated sensed data to the laptop node, and mobile agent transmission to the V_3 type node for the in situ verification process. A software application was developed and deployed on the testbed to manage the anomaly detection system. The TinyOS, object oriented programming language, and relational database management system were used to build the complete application software package.

The developed software is made up of five functional layers. The lower layer (i.e. layer 5) performs the core functions such as sensing the ambient temperature and performing the in situ verification on the V_3 type node, the cross-layer anomaly detection and mobile agent transmission on the V_2 type node, and the regions computation and update on the V_1 type node. Layer 4 handles the storage functionality across the network. On the V_3 type node, it stores the sensed data and battery status, which are later used by the mobile agent for the in situ verification process. Layer 4 stores the aggregated data, the cross-layer rule-base, and the trust value on the V_2 type node. Finally, on the V_1 type node, the anomaly detection reports which are transmitted by the V_2 node and information about the identities of the nodes in the network are stored by layer 4.

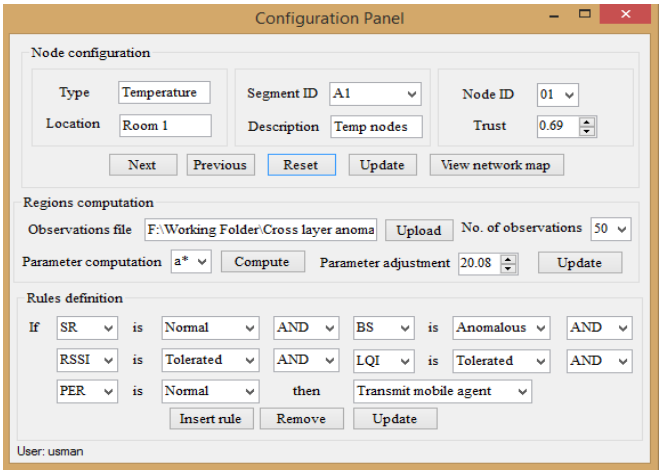

Fig. 3. GUI of the configuration panel.

Layer 3 defines the communication interfaces on all types of nodes and performs the following key functionalities: (i) the transmission of the sensed data from the V_3 to V_2 type node, (ii) the anomaly detection reports and the aggregated data transmission from the V_2 to V_1 type node, (iii) the transmission of the mobile agent for the in situ verification from the V_2 to V_3 type node, and (iv) the transmission of the in situ verification result from the V_3 to V_2 type node. The next layer (i.e., layer 2) extracts the information from the received data packets on the receiver side and hands them over to layer 5 to perform its functionalities. On the other hand, on the transmission side, layer 2 builds the packets and passes them to layer 3 for transmission. Finally, layer 1 provides the Graphical User Interfaces (GUIs) to manage and control the anomaly detection system in the smart home.

The application software is composed of two modules, namely, Configuration Panel and Report Panel. Note that due to the modular approach, the available options on the GUIs of the modules can be modified or even new modules can be included as per the preferences of the user. The configuration panel window has three components, namely, Node Configuration, Regions Computation, and Rules Definition. The Node Configuration component is responsible for defining types of nodes such as temperature, pressure, and motion sensors, and also defining identities of nodes and location of nodes with respect to the rooms and network segments within the smart home. The Node Configuration component also has an option to increase or decrease the trust level of a chosen node. It also enables the user to view a preloaded network map to determine the location of sensor nodes within the network. The Regions Computation component can be used to compute or adjust the values of the parameters in order to define regions. The third component, namely, Rules Definition offers a service to define fuzzy rules. The GUI of the Configuration Panel is shown in Fig. 3.

The Report Panel provides a facility to access the anomalies report. The reports can be generated with respect to the identity of the network segment, room identity, and sensor

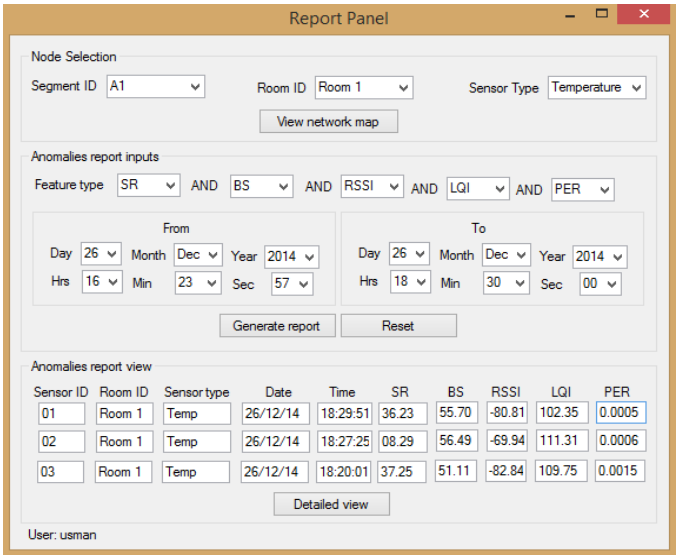

Fig. 4. GUI of the report panel.

type for selected date and duration of time. Three anomalies records, in the reverse chronological order, are displayed within the Report Panel window. A complete report can be viewed by clicking on the "detailed view" button. The GUI of the Report Panel is shown in Fig. 4.

In order to compute the regions and consequently setting up the cross-layer expert system for the experiments, the data traffic of 1000 iterations from the V_3 to V_2 type node was sampled. The values of the features SR, BS, LQI, and RSSI of each data packet were saved. On the other hand, the values of the PER were computed for every five data packets. The observations of the node and link features are plotted in Fig. 5 and Fig. 6, respectively, and the statistics of the observations are given in TABLE II.

The value of the parameter n was set as 50 and as a consequence the values given in TABLE III were obtained to define the regions for the experiments. The 10% randomly generated anomalous traffic was included in the dataset. The cross-layer rule-base was created using the configuration panel and by following the structure described in Section III-C. The size of the developed mobile agent was 762 bytes including both code and data. The mobile agent cannot be transmitted as a single data packet because of its large size [15]. Therefore, it was segmented into the eight packets. The first seven packets had $7 \times (102 + 25) = 889$ bytes size, where 102 and 25 were payload and header sizes, respectively. Similarly, the last packet had the size of $1 \times (48 + 25) = 73$ bytes. The trust decrement value was set as 0.33 for the observations in the tolerance region. The mobile agent was transmitted only if the trust value was 0. The trust value was reset to 1 whenever it reached to 0 for experimentation purpose. In practice, however, the proposed algorithm will generate an alarm to the user as soon as the trust value will reach to the lower bound.

In order to provide the baseline to the detection accuracy results, the experiments were also performed with a well-established crisp-logic classification algorithm, namely, decision tree. The detection accuracies for the crisp-logic case

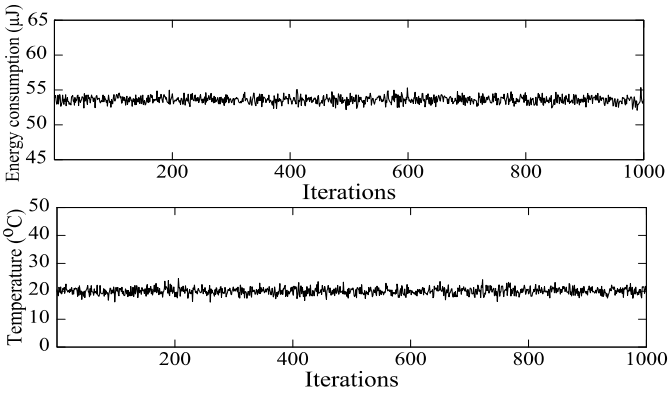

Fig. 5. Node features.

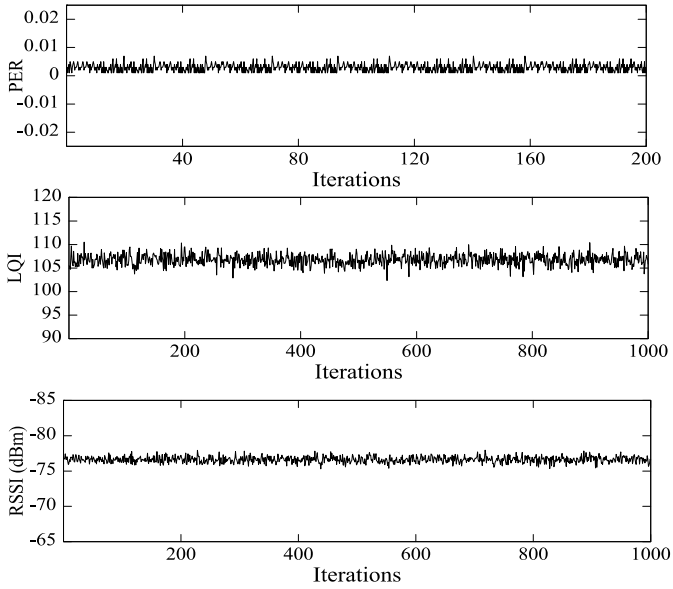

Fig. 6. Link features.

TABLE II
SAMPLED CROSS-LAYER DATA STATISTICS

Feature Category	Feature	Mean	Standard Deviation
Node	*SR*	20.08	1.43
	BS	53.70	0.50
Link	*RSSI*	-76.65	0.33
	LQI	106.75	1.28
	PER	0.0033	0.001

TABLE III
CROSS-LAYER EXPERT SYSTEM PARAMETERS

Feature	a^*	b^*	c^*	d^*	e^*	f^*
SR	18.65	19.26	19.88	20.28	20.90	21.51
BS	53.20	53.41	53.63	53.77	53.99	54.20
RSSI	-76.98	-76.83	-76.70	-76.60	-76.46	-76.32
LQI	105.47	106.02	106.57	106.93	107.48	108.03
PER	0.0020	0.0024	0.0029	0.0031	0.0036	0.0040

were 98.8%, 98.5%, 98.7%, 98.4%, and 98.7% for SR, BS, LQI, RSSI, and PER features, respectively. On the other hand,the detection accuracy was steady at 100% for the proposed scheme, as shown in Fig. 7. The proposed scheme, however, requires the domain expertise to appropriately setup the cross-layer rule-base to detect anomalies with high accuracy.

For estimation of the energy consumption by the mobile agent transmission, the experiments were performed by employing the cross-layer approach (i.e., the proposed

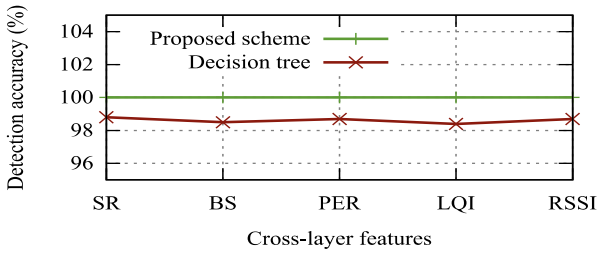

Fig. 7. Detection accuracy.

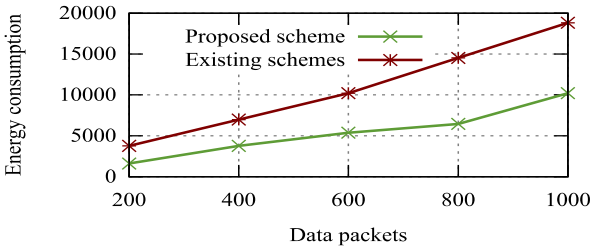

Fig.8. Energy consumption.

TABLE IV
MEMORY, PROCESSING TIME, AND ENERGY RESULTS

Procedure	RAM (bytes)	ROM (bytes)	Processing Time (ms)	Energy Consumption (µJ)
Phase 1	81	4013	12.73	113.12
Phase 2	1439	73303	282.86	2554.76

scheme) and by without employing the cross-layer approach (i.e., the existing schemes [6]-[9]). In the case of the existing schemes (i.e., without consulting the communication link-state for the mobile agent transmission decision), the energy consumption for the mobile agent transmission was between 3764.32 µJ to 18821.60 µJ for 200 to 1000 data packets, respectively. On the contrary, in the case of the proposed scheme, the energy consumption was between 1613.28 µJ to 10217.44 µJ for 200 to 1000 data packets, respectively, as shown in Fig. 8. The results of these experiments indicate that the proposed scheme can save 42.85% to 54.29% energy as compared to the existing schemes, which do not consider the communication link-state before the transmission of mobile agents.

The algorithm implementation results for the memory consumption, processing time, and energy consumption are given in TABLE IV. These results establish two facts: (i) the proposed algorithm is suitable for the low resource sensor nodes and (ii) the implementation results are consistent with the theoretical results presented in Section IV-B.

The key findings from the experiments are following: (i) initially the domain knowledge is required to properly setup the proposed anomaly detection system, (ii) the user can then control/ track the system performance through user friendly configuration/ report panel, (iii) the proposed scheme can detect cross-layer anomalies with high accuracy, (iv) in the case of the poor communication link-state, the mobile agent cannot be reliably transmitted and the anomaly detection

system has to rely on other actions such as notifying the user about anomalies, and (v) the proposed scheme offers energy efficient and more reliable service to transmit mobile agent.

VI. Conclusion

A robust anomaly detection system is indispensable in the smart home to timely notify users about the anomalies caused by the transmission errors, node faults, or attacks. This study has contributed towards the design and implementation of a novel cross-layer anomaly detection scheme for smart home sensor networks. The proposed scheme employs simple, yet practically effective statistical procedures along with the fuzzy logic to detect cross-layer anomalies. It also offers the facility to transmit mobile agents after consideration of the communication link-state. The proposed scheme is implemented on a testbed and results indicate its ability to detect cross-layer anomalies with high accuracy and also its capability to decrease the energy consumption caused by the mobile agent transmissions in the poor communication link-state situations. An application software is developed to manage the anomaly detection system in a smart home environment, which empowers the user to adjust the statistically derived values of the parameters to tune the performance of the system. The modular design of the application software makes it suitable for integration into the main application software, which controls the smart home.

Future work on the proposed scheme will aim to achieve two primary goals: (a) building different profiles for different natures of sensor nodes with respect to their hardware and designated roles in the smart home, and (b) incorporating role-based access control mechanism into the application software to offer different levels of privileges to the system administrator and users.

References

[1] Y. Xue, X. Chang, S. Zhong, and Y. Zhuang, "An efficient energy hole alleviating algorithm for wireless sensor networks," *IEEE Trans. Consumer Electron.,* vol. 60, no. 3, pp. 347-355, Aug. 2014.

[2] F. J. Bellido-Outeirino, J. M. Flores-Arias, M. Linan-Reyes, E. J. Palacios-Garcia, and J. J. Luna-Rodriguez, "Wireless sensor network and stochastic models for household power management," *IEEE Trans. Consumer Electron.,* vol. 59, no. 3, pp. 483-491, Aug. 2013.

[3] R. Harper, *Inside the Smart Home*, 1st ed., Springer, 2003, pp. 1-2.

[4] J. Wang, Z. Zhongqi, B. Li, S. Lee, and R. S. Sherratt, "An enhanced fall detection system for elderly person monitoring using consumer home networks," *IEEE Trans. Consumer Electron.,* vol. 60, no. 1, pp. 23-29, Feb. 2014.

[5] D.-M. Han and J.-H. Lim, "Smart home energy management system using IEEE 802.15.4 and zigbee," *IEEE Trans. Consumer Electron.,* vol. 56, no.3, pp. 1403-1410, Aug. 2010.

[6] M. Usman, V. Muthukkumarasamy, X.-W. Wu, and S. Khanum, "Wireless smart home sensor networks: mobile agent based anomaly detection," *in Proc. IEEE 9th International Conference on Ubiquitous Intelligence & Computing* and *9th International Conference on Autonomic and Trusted Computing*, Fukuoka, Japan, pp. 322-329, Sept. 2012.

[7] M. Usman, V. Muthukkumarasamy, and Xin-Wen Wu, "A resource-efficient system for detection and verification of anomalies using mobile agents in wireless sensor networks," *Journal of Networks,* vol. 9, no. 12, pp. 3427-3444, Dec. 2014.

[8] M. Pugliese, A. Giani, and F. Santucci, "Weak process models for attack detection in a clustered sensor network using mobile agents," *in Proc. International Conference on Sensor Systems and Software*, Pisa, Italy, pp. 33-50, 2010.

[9] M. Ketel, "Applying the mobile agent paradigm to distributed intrusion detection in wireless sensor networks," *in Proc. IEEE Southeastern Symposium on System Theory*, New Orleans, USA, pp. 74-78, March 2008.

[10] K. Kapitanova, S. H. Son, and K.-D. Kang, "Using fuzzy logic for robust event detection in wireless sensor networks," *Ad Hoc Networks*, vol. 10, no. 4, pp. 709-722, June 2012.

[11] Y. Gu and T. He, "Dynamic Switching-Based Data Forwarding for Low-Duty-Cycle Wireless Sensor Networks," *IEEE Trans. on Mobile Comp.*, vol. 10, no. 12, pp. 1741-1754, Dec. 2011.

[12] J. J. Buckley, W. Siler, and D. Tucker, "A fuzzy expert system," *Fuzzy Sets and Systems,* Vol. 20, no. 1, pp. 1-16, Aug. 1986.

[13] L. Tang, K.-C. Wang, Y. Huang, and F. Gu, "Channel characterization and link quality assessment of IEEE 802.15.4-compliant radio for factory environments," *IEEE Trans. Ind. Info.*, vol.3, no.2, pp. 99-110, May 2007.

[14] Y. Wang, G. Attebury, and B. Ramamurthy, "A survey of security issues in wireless sensor networks," *IEEE Commun. Surv.*, vol. 8, no. 2, pp. 2-23, Feb. 2007.

[15] *Part 15.4: Wireless Medium Access (MAC) and Physical Layer (PHY) specifications for low-rate wireless Personal Area Network (LR-WPANs),* IEEE Std. 802.15.4, 2006.

BIOGRAPHIES

Muhammad Usman received the MS (Computer Science) from the PMAS-AAUR, Pakistan. He is currently associated with the School of Information and Communication Technology, Griffith University, Australia, where he is a mature PhD candidate. Prior to joining Griffith University, he has held several academic and industrial positions in different parts of the world. His current research interests include design of mobile agent-based anomaly detection systems for smart homes and formal verification of communication protocols. He has published over twenty research papers in international journals and conferences.

Vallipuram Muthukkumarasamy (M'86) obtained B.Sc. Eng. with 1st Class Hons. from the University of Peradeniya, Sri Lanka and obtained Ph.D. from Cambridge University, England. He is currently attached to the School of Information and Communication Technology, Griffith University, Australia as an Associate Professor. His current research areas include investigation of security issues in wireless networks, sensor networks, trust management in MANETs, key establishment protocols and medical sensors. He is currently leading the Network Security Research Group at the Institute for Integrated and Intelligent Systems at Griffith University. He has also been providing leadership to innovative learning and teaching practices. He has received a number of best teacher awards.

Xin-Wen Wu (M'00-SM'14) received the Ph.D. degree from the Chinese Academy of Sciences. He was with the Chinese Academy of Sciences, the University of California, San Diego (as a post-doctoral researcher), and the University of Melbourne (as a research fellow). He was affiliated with the School of Information Technology and Mathematical Science, University of Ballarat, Australia. In April 2010 he joined Griffith University, Australia, as a faculty member of the School of Information and Communication Technology. His research interests include network and data security, coding techniques, and information theory and its applications. He has published one book, three book chapters and over sixty research papers in IEEE transactions, other journals, and conferences proceedings.

View publication stats