Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Anomaly detection of network-initiated LTE signaling traffic in wireless sensor and actuator networks based on a Hidden semi-Markov Model

*June-ho Bang, Young-Jong Cho *, Kyungran Kang*

College of Information & Computer Engineering, Ajou University, Suwon, Republic of Korea

ARTICLE INFO

Article history:
Received 4 January 2016
Received in revised form 5 October 2016
Accepted 14 November 2016
Available online 23 November 2016

Keywords:
Wireless sensor and actuator network
LTE signaling attack
Hidden semi-Markov Model
Anomaly detection
Intrusion detection system

ABSTRACT

LTE signaling attack is a serious threat to a wireless sensor and actuator network whose facilities are dispersed and connected with LTE technology on a large scale, in order to conduct a particular mission. An LTE attacker generates a lot of signaling initiating packets, named wakeup packets, to saturate the LTE network's resources. Existing LTE signaling attack detection schemes are merely based on measuring the mean wakeup packet generation rate. Since resulting from extensive amounts of facilities involved in a normal management process, severe fluctuations of signaling traffic are ordinarily expected in the wireless sensor and actuator network, and those mean-based schemes cannot effectively distinguish between attacks and normal traffic. In this paper, we propose an advanced LTE signaling attack detection scheme based on a Hidden semi-Markov model, which captures the spatial–temporal characteristics of normal wakeup packet generation behavior. Our proposed detector takes the log-likelihood of a node's wakeup packet generation as the test criterion for normality. Through simulations with various parameter settings, we verified that the proposed scheme effectively distinguishes attacker nodes from normal nodes.

1. Introduction

Facilities for industrial automation, transportation, agriculture, smart grids, and military operations incorporate numerous types of devices such as control servers, sensors and actuators, which are usually dispersed over a large area (Calle-Sanchez et al., 2012; Guan et al., 2011; Güngör et al., 2011; Hasan et al., 2013; Lien et al., 2011, 2012; Peng et al., 2011; Souryal and Golmie, 2011; Tingting and Bin, 2010; Yan et al., 2013). The devices frequently exchange control messages to handle the facility events in a timely manner. In this paper, we call a large-scale network with such particular missions a wireless sensor and actuator network (WSAN) (Akerberg et al., 2011; Chen et al., 2010; Güngör and Hancke, 2009).

LTE can provide a certain level of quality of service (QoS) for time-critical applications and also large cell coverage such as a radius of 10 km (Talukder et al., 2013). Thus, LTE has been considered to be one of the most promising candidate access technologies for the WSAN. In Fig. 1, we show a general WSAN configuration where remote sensors and actuators are connected with LTE technology to control servers in the fixed wired network.

Originally, an LTE network is designed to assign LTE communication resources, named LTE bearers, to wireless nodes on a demand basis (Holma and Toskala, 2011; Kreher and Gaenger, 2010; Rao and Gajula, 2011). The procedure for the LTE bearer assignment process is called LTE signaling. In Fig. 2, a representative LTE network configuration for WSAN is given. When a message from the wired network arrives at a serving

* *Corresponding author.*
E-mail addresses: june3731@ajou.ac.kr (J. Bang), yjcho@ajou.ac.kr (Y.-J. Cho), korykang@ajou.ac.kr (K. Kang).
http://dx.doi.org/10.1016/j.cose.2016.11.008

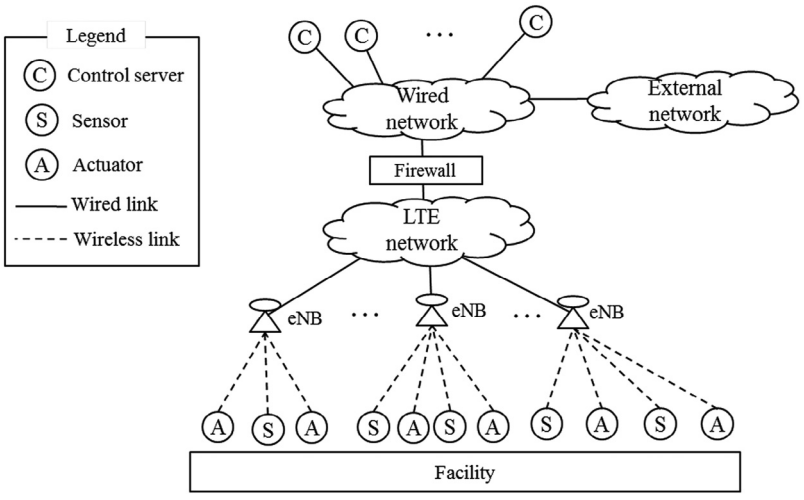

Fig. 1 – General WSAN configuration.

gateway (SGW), if the destination node of the message is not bearer assigned, an LTE signaling process is initiated. All of the involved network elements including several eNBs (evolved Node Base stations) participate in the LTE signaling process, exchanging numerous control messages that consume their computing resources. The LTE signaling triggering message is called a "*wakeup packet*" (Gupta et al., 2013). Bearer assignment expires when the involved nodes have not communicated for bearer-timeout (BT) units of time.

An LTE signaling attacker transmits malicious messages to sensors and actuators in every BT + ε (Bassil et al., 2012; Gupta et al., 2013; Jang et al., 2014; Jover, 2013; Lee et al., 2009). The messages initiate meaningless signaling processes, which are a big burden to deplete involved node's computing resources. Sometimes, due to the signaling attack, the LTE network experiences intermittent disconnection (Gupta et al., 2013). Therefore, the control messages experience larger delays or frequent losses, and time-critical facility management for conducting the relevant mission becomes paralyzed.

In this paper, we propose an LTE signaling attack detection scheme for WSANs. We assume that control servers are vulnerable to be compromised as puppet-like attackers to execute an effective signaling attack. Therefore, the attack detector is assumed to be located at the SGW in the LTE network, as shown in Fig. 2.

Broadly speaking, network intrusion detection schemes can be categorized into misuse detection and anomaly detection (Bhuyan et al., 2014). Their difference is mainly on what kind of prior information they use for detection. Misuse detection schemes use attack signatures obtained from reports on prior former attack reports. They raise alarm whenever an observed network traffic pattern is sufficiently similar to an attack signature. On the other hand, anomaly detection schemes use normal traffic pattern. They raise alarm whenever an observed network traffic pattern is sufficiently distant from the assumed normal traffic pattern.

The wireless node triggered signaling attack detection scheme proposed in Gupta et al. (2013) is misuse detection. This scheme maps each wireless node to a vector in a hyperplane. The coordinates of a vector represent the variation of destination addresses in data packets transmitted by the correspondent node, and the wakeup packet generation rate, etc. When a wireless node is reported as a signaling attacker, the corresponding vector becomes an attack signature. This scheme partitions the hyperplane using the notion of support vector machine and determines a wireless node as an attacker if it shares the same partition with a given signature. Misuse detection needs to pile up a large amount of attack signatures in order to achieve good detection accuracy. Due to the difficulty in implementing such a large signature knowledge base, researchers prefer anomaly detection schemes.

The cumulative sum (CUSUM)-based signaling attack detection proposed in Lee et al. (2009) is an anomaly detection

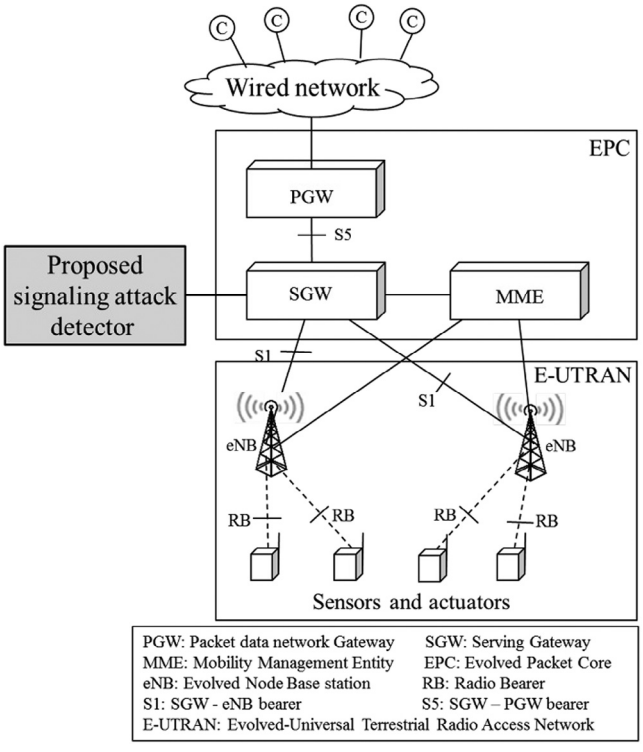

Fig. 2 – Reference LTE network configuration for a WSAN.

scheme. This scheme updates the CUSUM statistic whenever a wakeup packet arrives at the SGW. If the wakeup arrival rate is higher than the average wakeup arrival rate for a certain enough time, then the scheme raises alarm. The CUSUM-based detection overly simplifies legitimate traffic behavior merely by an average behavior. If a legitimate node's wakeup packet generation rate fluctuates, the CUSUM-based detection scheme is likely to raise frequent false alarms.

Other researchers proposed HMM (Hidden Markov Model) as a normal behavior description model for anomaly detection. The HMMs in such researches represent a system's normal resource utilization behavior (Sugaya et al., 2009), routing message transmission behavior of mobile ad-hoc network nodes (Ye et al., 2010), system function call behavior (Shi and Sun, 2012), and cloud system states (Hong et al., 2015). HMM's limitation is that the latent variable's state sojourn time is distributed only with geometric distribution. Therefore, HMM lacks the ability to represent various latent transition behaviors. For this reason HsMM (Hidden semi-Markov Model) has been proposed to overcome the shortcoming of HMM. HsMM is a HMM with arbitral state sojourn time and is well known for time-series behavior analysis.

Web server attack detection schemes proposed by Xie and Yu (2009a, 2009b) and Xie et al. (2013) represent web search behaviors by HsMM. They used HsMM for modeling a web client's page request behavior (Xie and Yu, 2009a), aggregate webpage request behavior (Xie and Yu, 2009b), and web server's cache access behavior (Xie et al., 2013). The HsMM-based attack detections effectively distinguish abrupt traffic surges by flash crowds from those by attacks.

However, no HsMM-based attack detection scheme for LTE signaling attacks has been reported up to date. To fill in the blank, we use HsMM for modeling normal wakeup packet generation behavior and propose a new signaling attack detector. This paper assumes that the wakeup packet generation process at a control server has a particular spatial–temporal characteristic. To exploit this characteristic, we use HsMM as an effective way for modeling the spatial–temporal characteristic of the wakeup packet generation process. We will discuss what the spatial–temporal characteristic means in detail later. The proposed detector compares real-time spatial–temporal characteristic of a server's wakeup packet generation with the employed HsMM. The detector raises an alarm if they are far different from each other, or updates the HsMM parameters otherwise. Our detector has better detection accuracy compared with prior schemes as will be shown with simulation results. We show that even though an attacker is highly intelligent, it is almost unable to avoid our detector by generating wakeup packets mimicking the complete characteristic of an ordinary control server to avoid our detector.

The contribution of this paper is three-fold. First, we propose a new attack decision criterion which exploits the spatial–temporal characteristics hiding in the generation mechanism of wakeup packets. Second, we propose an HsMM that captures the spatial–temporal characteristic of a control server's signaling behavior. Third, we propose an advanced LTE signaling attack detector and prove its superiority by various simulations. We show that our proposed detector results in much less frequent false alarms.

Our work is organized as follows. Section 2 introduces the system model in detail. An algorithm description for the proposed signaling attack detector is given and discussed in Section 3. Section 4 presents the simulation results and verifies our analytical models. Conclusions are drawn in Section 5.

2. System model

In this section, we describe the system model to derive an anomaly detection in a WSAN. First, we explain how an LTE signaling process is initiated by a wakeup packet, and the detailed generation mechanism of wakeup packets in control servers is given. Next, we introduce a general HsMM and propose a particular HsMM to model the wakeup packet generation process.

2.1. *Wakeup packet generation model*

In this subsection, we explain a control server's wakeup packet generation process. Control servers $\{c_k\}$ send commands to actuators $\{a_j\}$ based on sensing data transmitted by sensors $\{s_i\}$ as shown in Fig. 3. Measuring facility states, sensors periodically send sensing data to control servers to handle facility state changes. Depending on the information, each control server determines which task should be done in the facility and sends commands to relevant actuators. The actuators perform the commands and result in physical effects on the facility. The sensors feedback new observed outcomes to the control servers. In this paper, we will focus on an arbitrarily tagged control server c_k and describe its behavior.

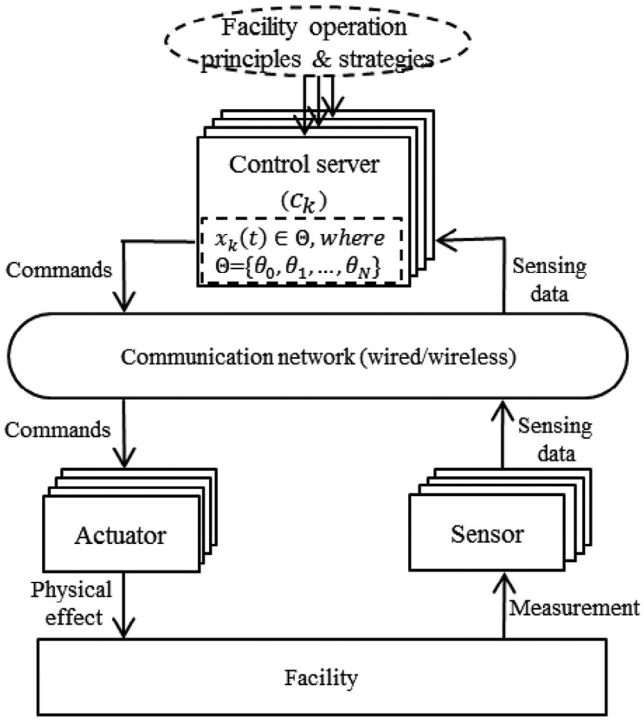

Fig. 3 – Message transactions between nodes in a WSAN.

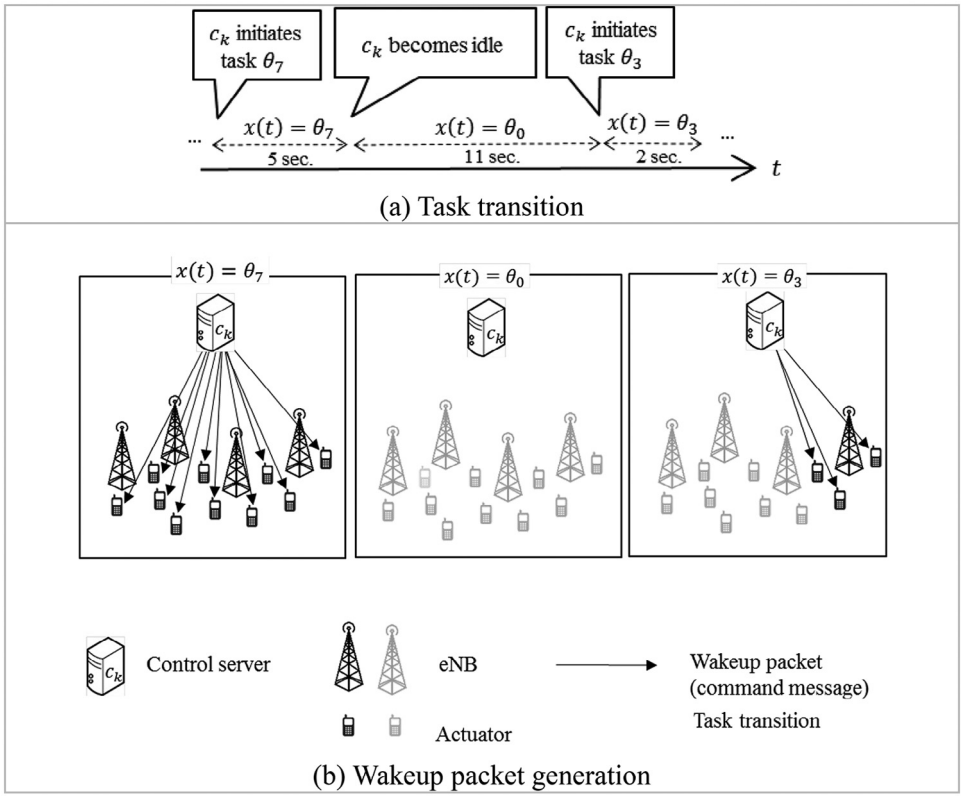

Fig. 4 – Wakeup packet generation model at control servers.

Command messages transmitted by c_k may include several wakeup packets. At the time of arrivals, depending on whether the target actuator of the wakeup packets is bearer assigned, bearer setup signaling for the actuator is probabilistically initiated in the LTE network. The spatial–temporal characteristic of wakeup packet generation caused by these command messages is therefore dependent on the current task transitions executed on the tagged server c_k, which are explained with an example in Fig. 4.

Let $\Theta = \{\theta_m\}$ be a set of server tasks, including θ_0 which is a null task. At time t, c_k works on a task $x(t) \in \Theta$. In Fig. 4(a), it chooses θ_7 and works for 5 seconds. After that, finding there is nothing to do in the facility from sensing data, it becomes idle in θ_0 for 11 seconds. And then, θ_3 is chosen as the next task for 2 seconds. The eNBs drawn in dark black color within a box titled with $x(t) = \theta_m$ in Fig. 4(b) are serving eNBs engaged in relaying at least a wakeup packet during the execution of task θ_m. Fig. 4(b) leads to our idea on describing the spatial–temporal characteristic of wakeup packet generation, which is dependent on the transition behaviors of $x(t)$.

When the tagged server c_k is doing a task like θ_7 in which a larger number of actuators dispersed over a wider area are involved as compared to tasks like θ_0 or θ_3 in Fig. 4(b), it would transmit more commands to actuators in a wider spatial area than the cases with θ_0 or θ_3. As a result, θ_7 induces more wakeup packets over a wider area with a higher generation rate than those for θ_0 and θ_3 in Fig. 4(b).

We can observe these spatial characteristics of wakeup packet generation during the execution of θ_m numerically, by a metric representing the uncertainty E_m of serving eNBs for θ_m, named "eNB entropy", as follows:

$$E_m = -\sum_{l \in EB_m} P_{eB}(m, l) \cdot \log P_{eB}(m, l), \tag{1}$$

where EB_m is a set of serving eNBs involved in executing θ_m, and $P_{eB}(m, l)$ is the probability of wakeup packets relayed by eNB l during θ_m execution. Thus, E_m represents an approximate spatial location sparsity for relaying wakeup packets for executing θ_m.

Let T_m be the time duration of θ_m and R_m be the wakeup packet generation rate during θ_m, respectively. In case of Fig. 4(b), E_7 and R_7 would have larger values than E_3 and R_3, respectively, because many more eNBs contribute evenly to wakeup packet delivery for θ_7. Both E_m and R_m are random variables dependent on θ_m, not deterministic values. Although c_k generates the same set of command messages whenever it works for θ_m, only some subsets of them become wakeup packets depending on whether destination actuators are bearer assigned or not. In Fig. 4(b), E_7 and R_7 would have a probability distribution skewed to larger values while E_0, R_0, E_3 and R_3 would have distributions skewed to smaller values. T_m is also considered as a random variable since task duration depends on numerous factors in the facility and so cannot be determined as fixed in general.

We assume that our signaling attack detector located at the SGW, as given in Fig. 2, can observe the eNB entropy and wakeup packet generation rate in a real-time manner, but $x(t)$ is latent from observation. Therefore, real-time eNB entropy and wakeup

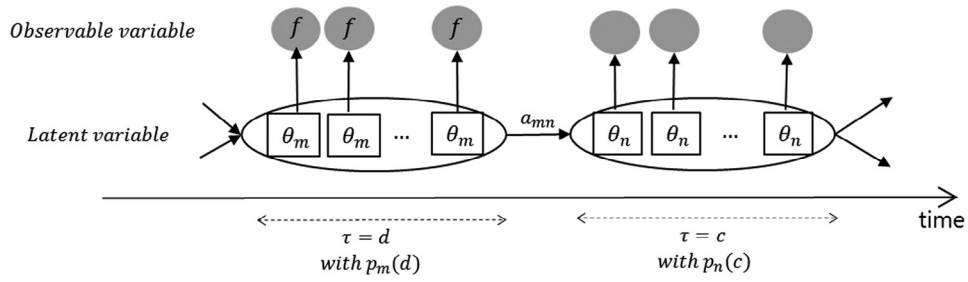

Fig. 5 – Basic Hidden semi-Markov Model (Xie et al., 2013).

packet generation rate are modeled as mixtures of random variables with hidden states.

As a control server alternates between numerous tasks with different characteristics, it appears to generate wakeup packets whose generation rate is fluctuating and therefore whose related eNB entropy is also varying. We observe that previous signaling attack detection schemes based on average rate estimation are likely to falsely determine a normal control server as attacker if it persists with a task with a larger wakeup packet generation rate for a certain long-enough time. Thus, in this paper, rather than relying on the mere average rate computation, we resort to a different approach based on a behavior context description model. We believe HsMM to be best to qualify our needs in that regard. In the following subsection, we describe a general HsMM and propose a particular HsMM to formularize the process for wakeup packet generation.

2.2. *HsMM for wakeup packet generation*

HsMM models a discrete time stochastic process which composes latent and observable variables. The latent variable is modeled to take a value from a state of a semi-Markov chain at a time. The observable variable is a random variable which is dependent on the current state of the latent variable. The basic structure of an HsMM is shown in Fig. 5. Values of the latent and observable variables are represented with rectangles and circles, respectively.

An instance of HsMM is characterized by parameter $\lambda = (\Theta, \{\pi_m\}, \{a_{mn}\}, \{b_m(f)\}, \{p_m(d)\})$ where the components are described in Table 1. The semi-Markov chain of the latent variable is composed of the triplet $(\Theta, \{a_{mn}\}, \{p_m(d)\})$. The latent variable $x(t)$, $t = 1, 2, \ldots, \rho$, takes a value (state) from Θ at a time. If the latent variable takes a value θ_m at t, it holds θ_m until $t + d - 1$ and transits to take another value at $t + d$. The next value is determined to be θ_n with probability a_{mn} where $\sum_{n \in \Theta} a_{mn} = 1$ for all m. d is the value holding time (state sojourn time) of θ_m, and its distribution is specified by $p_m(d)$ where $\sum_{d=1}^{D} p_m(d) = 1$ and D is the maximum state sojourn time. π_m is the probability that the initial latent value $x(1)$ is θ_m.

The observable variable has a value f with probability $b_m(f)$ if the latent value is θ_m. The sequence of latent variables is not observable, but affects the distribution of observable variables. Therefore, the observation appears to be a mixture of random variables, and we can only infer the latent sequence indirectly by the observation sequence. An observation sequence $f_1^{\rho} = (f_1, f_2, \cdots, f_{\rho})$ in ρ consecutive times from time 1 to ρ is conditionally independent (Bishop, 2006), given a sequence of the latent variables $x(t), t = 1, 2, \cdots, \rho$, so that the probability of observing f_1^{ρ} is $\prod_{t=a}^{b} b_{x(t)}(f_t)$ where f_t is an observation at time t.

In this paper, we divide time into equal sized timeslots, named "*Observation Window*" (OW), and model a control server's wakeup packet generation behavior as a discrete time stochastic process using HsMM. Table 1 summarizes HsMM parameters for modeling a WSAN.

In our HsMM, Θ is a set of server tasks. We assume the LTE signaling attack detector installed on the SGW cannot directly observe the server task sequence $x(t), t = 1, 2, \ldots, \rho$, but $x(t)$ can only be inferred from observation sequences of the eNB entropy and wakeup packet generation rate. $x(t)$ is modeled as a semi-Markov chain $(\Theta, \{a_{mn}\}, \{p_m(d)\})$ as shown in Fig. 6. $x(t)$ takes a value from the states of the semi-Markov chain. a_{mn} is the transition probability that θ_n is chosen next given the current task is θ_m, and $p_m(d)$ is the probability distribution of θ_m's duration time in terms of the number of OWs. $\{\pi_m\}$ is the probability that the initial task $x(1)$ is θ_m. Let us denote by $OW(t)$, the t-th observation window. After a control server c_k completes θ_m at $OW(t)$, it chooses θ_n as the next task at $OW(t+1)$ with probability a_{mn} and terminates θ_n at $OW(t + d)$. d, the time duration of θ_n in terms of OWs, is determined to be an integer value with probability $p_n(d)$. Both $\{p_m(d)\}$ and $\{a_{mn}\}$ are determined reflecting the facility operation principles, strategies, and the control server's role. Revisiting the example in Fig. 4(a) with our semi-Markov chain, the c_k chooses

Table 1 – Definition of HsMM parameters.

	General HsMM	HsMM for WSAN
$\Theta = \{\theta_0, \theta_1, \theta_2, \ldots, \theta_N\}$	A set of latent states	A set of server tasks
$\pi_m = P[x(0) = \theta_m \mid \lambda]$	Initial latent state probability	Initial server task selection probability
$a_{mn} = P[x(t) = \theta_n \mid x(t-1) = \theta_m, \lambda]$	Latent state transition probability	Server task transition probability
$b_m(f) = P[f \mid x(t) = \theta_m, \lambda]$	Conditional distribution of observation variable	Conditional distribution of observable variables which correspond to eNB entropy and wakeup generation rate
$p_m(d), d \in \{1, 2, 3, \ldots, D\}$	Latent state sojourn time distribution	Task duration distribution

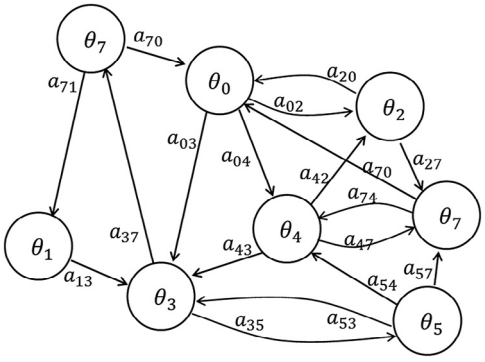

Fig. 6 – An example of a semi-Markov chain for server task transition.

θ_7 as initial state with probability π_7 and stays for 5 seconds with $p_7\left(\frac{5}{|\mathrm{OW}|}\right)$ where |OW|, the length of an OW, is assumed to be 1 second. After then, it chooses θ_0 as the next task with probability $a_{7,0}$ and stays on there for 11 seconds with $p_0\left(\frac{11}{|\mathrm{OW}|}\right)$ until it transits to θ_3.

Two observation variables at $OW(t)$ comprise $F(t)=(E(t),R(t))$, a *feature vector*, where $E(t)\in\{0,1,2,\ldots,L\}$ and $R(t)\in\{0,1,2,\ldots,r_{max}\}$ are finite non-negative integer-valued random variables representing eNB entropy and wakeup packet generation rate, respectively. We formalize $E(t)$ so that it has non-negative integer values bounded above by a finite integer, L, as follows:

$$E(t)=\left\lceil (L/\log|EB|)\times\left(-\sum_{l\in EB}p_{eb}(l)\cdot\log(p_{eb}(l))\right)\right\rceil \quad (2)$$

where EB is the set of eNBs involved, and $p_{eb}(l)$ is the proportion of wakeup packets served by eNB l in $OW(t)$. Note that $E(t)$ has two factors of $(L/\log|EB|)$ and $(-\sum_{l\in EB}p_{eb}(l)\cdot\log(p_{eb}(l)))$ where the latter term takes the genuine form of entropy definition whose range is $[0,\log|EB|]$, assuming the eNB number l as a random variable with a probability mass function $p_{eb}(l)$. In order for $E(t)$ to have integer values in $\{0,1,2,\ldots,L\}$, we multiply $(L/\log|EB|)$ and apply ceiling function to the genuine eNB entropy. We use a convention that $0\cdot\log 0=0$. $R(t)$ is simply the number of wakeup packets generated at $OW(t)$. Therefore, $R(t)$ is a finite non-negative integer-valued random variable bounded above by r_{max}, where r_{max} is determined considering the task characteristics of c_k and the real WSAN operating constraints, such as the number of actuators or bandwidth of the wired network to which c_k is directly connected. $F(t)$ takes a value $f=(e,r)$ with probability distribution $\{b_m(f)\}$ given $x(t)=\theta_m$. When $x(t)=\theta_0$, c_k does not generate any commands. Therefore, $b_0(f)=1$ if $f=(0,0)$ and $b_0(f)=0$, otherwise.

We assume that $E(t)$ and $R(t)$ are conditionally independent given $x(t)$ (Bishop, 2006). Without any knowledge about $x(t)$, we easily suppose that $E(t)$ and $R(t)$ are usually in a positive correlation. This is because if a large value of $E(t)$ is observed, we believe many eNBs were involved to handle a large number of wakeup packets, consequently expecting a large value of $R(t)$. Meanwhile, let us assume that we observe the exact value of $x(t)$, the ongoing task θ at $OW(t)$. If θ has a characteristic that the number of actuators associated with θ is small and remains almost unchanged, contrary to the case without any knowledge about $x(t)$, we still assure the same characteristic holds even though the number of eNBs involved in θ becomes large. This shows a case that $R(t)$ is in a range of small values even though we observe a large value of $E(t)$. Thus, depending on the property of the tasks in progress on the server, the aforementioned unconditional positive correlation between $E(t)$ and $R(t)$ does not always hold. We express this conditional independence property as $b_m(f)=b_m(e)\cdot b_m(r)$ where $b_m(e)$ and $b_m(r)$ are probability distributions of the eNB entropy and wakeup packet generation rate given $x(t)=\theta_m$, respectively.

Furthermore, observations in each OW are conditionally independent so that the probability of successive observations, f_a^b from $OW(a)$ to $OW(b)$, is $\prod_{t=a}^{b}b_{x(t)}(f)$ given $x(t)$. $E(t)$ and $R(t)$ are random variables, each of which represents spatial and temporal characteristics of wakeup packets generated in $OW(t)$. If the control server works on a task that involves a lot of actuators located in a larger area and all of the serving eNBs evenly handle the wakeup packets, both $E(t)$ and $R(t)$ are likely to be larger values close to L and r_{max}, respectively. On the other hand, $E(t)$ and $R(t)$ would have smaller values close to 0 if the control server is related to a smaller number of actuators and eNBs at time t.

HsMM appropriately describes the fluctuations of both the eNB entropy and wakeup packet generation rate. To verify this, let us suppose with an example in Fig. 6 that a server frequently takes θ_1, θ_3, θ_5, and θ_7 in sequence. In this example $\{\theta_1,\theta_5\}$ induces low wakeup packet generation rate and eNB entropy, while $\{\theta_3,\theta_7\}$ induces high wakeup packet generation rate and eNB entropy. As the control server frequently takes the sequence, the control server would show fluctuations in wakeup packet generation rate and eNB entropy repeatedly. Those frequent fluctuations can be modeled in HsMM by high values of $\{\alpha_{13},\alpha_{35},\alpha_{57}\}$ along with larger probability masses on larger values in $\{b_1(f),b_5(f)\}$ and that on low values in $\{b_3(f),b_7(f)\}$.

If an attack detector is well designed with an HsMM which has parameters quite close to true values, we can avoid undesirable situations where a normal control server is falsely judged as a signaling attacker even though it generates wakeup packets with higher rate for a certain long-enough time. This is a notable improvement from the previous mean-value-based signaling attack detection schemes. Furthermore, it is more difficult for a genuine attacker to avoid the HsMM-based detection. To do so, prior to attack the attacker has to obtain enough information on the dynamics of spatial characteristic such as the eNB entropy. Also, simply by eavesdropping, it is almost impossible for the attacker to obtain all the necessary information such as server task transitions, relevant sets of actuators for each tasks, and actuator–eNB pairing information. Therefore, a compromised attacker that does not learn the related information has no alternative but to assume an average situation where the counterfeit wakeup packets are destined evenly to all the serving eNBs. As a result, this case would induce a value converging to maximum eNB entropy. In the next section, we propose an LTE signaling attack detection scheme which employs HsMM for analyzing a control server's wakeup generation behavior, to determine normality.

3. Signaling attack detection algorithm

In this section we elaborate on our proposed signaling attack detection algorithm shown in Fig. 7. For every server c_k, the detector calculates normality of real-time wakeup packet generation pattern in ρ consecutive OWs based on the HsMM stated in Section 2. If the normality value is lower than a predetermined decision threshold, the detector determines c_k as a compromised attacker and raises an alarm. Otherwise, λ is adjusted according to the ρ observations so that the detector copes with behavioral change of c_k with time.

Prior to installation, the detector derives an initial HsMM parameter, λ, from a prior history which has been acquired in the SGW while handling wakeup packets and observing the c_k's task transition. A history of wakeup packet handling consists of its arrival times at the SGW and the conveyed identifiers about the serving eNBs.

Once the detector is installed, the SGW feeds a log entry to the detector every time a wakeup packet is handled. The log entry has the same format with the history entry. The detector accumulates log entries during each OW. At the end of each OW, the detector calculates a feature vector and stores it in the feature repository. If the detector finds f_1^ρ, a sequence of ρ feature vectors from OW(1) to OW(ρ) stored in the repository, the detector calculates the log-likelihood of feature vectors, $\log\left(P\left[f_1^\rho|\lambda\right]\right)$.

If $\log\left(P\left[f_1^\rho|\lambda\right]\right)$ is lower than a predefined decision threshold h, the detector determines c_k as a compromised attacker and raises an alarm. Otherwise, the detector determines c_k as a normal control server and updates HsMM parameters using the maximum a posteriori probability (MAP) estimates of λ reflected in f_1^ρ. As a whole, the proposed detector takes the log-likelihood of observations about wakeup packet generation as a promising test criterion for normality.

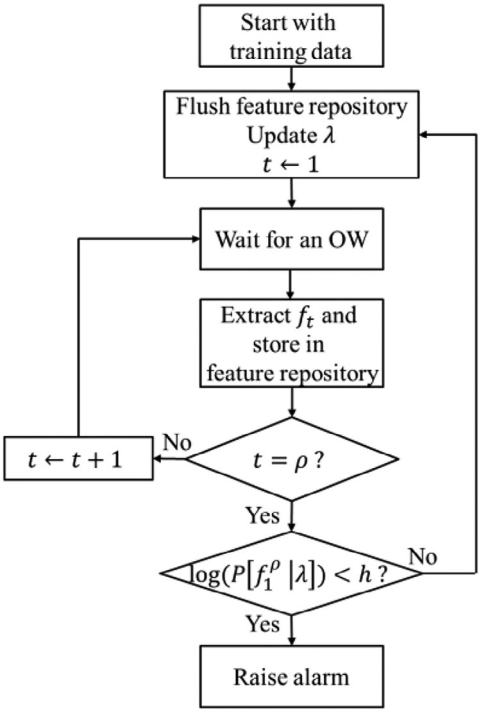

Fig. 7 – Flowchart of the proposed signaling attack detector.

Calculating $\log\left(P\left[f_1^\rho|\lambda\right]\right)$ and the MAPs of λ has huge time complexity. To reduce the time complexity, our proposed detector employs the well-known forward–backward algorithm (Xie and Yu, 2009a, 2009b; Xie et al., 2013; Yu, 2010; Yu and Kobayashi, 2003). Let $\alpha_t(m,d)$ be a forward variable and $\beta_t(m,d)$ be a backward variable which are defined as follows:

$$\alpha_t(m,d) = P\left[f_1^t, x(t) = \theta_m, \tau = d|\lambda\right], \tag{3}$$

$$\beta_t(m,d) = P\left[f_{t+1}^\rho | x(t) = \theta_m, \tau = d, \lambda\right], \tag{4}$$

where τ is a remaining time to the next state transition at time t. For all t,m and d, both $\alpha_t(m,d)$ and $\beta_t(m,d)$ can be calculated iteratively by the forward–backward algorithm. Once they are calculated, the detector obtains the following three types of joint distributions from forward and backward variables as follows:

$$\xi_t(m,n) = P\left[f_1^\rho, x(t-1) = \theta_m, x(t) = \theta_n|\lambda\right], \tag{5}$$

$$\eta_t(m,d) = P\left[f_1^\rho, x(t-1) \neq \theta_m, x(t) = \theta_n, \tau = d|\lambda\right], \tag{6}$$

$$\gamma_t(m) = P\left[f_1^\rho, x(t) = \theta_m|\lambda\right]. \tag{7}$$

$P\left[f_1^\rho|\lambda\right]$ and the MAPs of four HsMM parameters conditioned on f_1^ρ can be calculated as follows:

$$P\left[f_1^\rho|\lambda\right] = \sum_{m,d} P\left[f_1^\rho, x(\rho) = \theta_m, \tau = d\right] = \sum_{m,d} \alpha_\rho(m,d), \tag{8}$$

$$\hat{\pi}_m = \gamma_1(m) \Big/ \sum_n \gamma_1(n), \tag{9}$$

$$\hat{a}_{mn} = \sum_{t=1}^{\rho} \xi_t(m,n) \Big/ \sum_{t=1}^{\rho} \sum_k \xi_t(m,k), \tag{10}$$

$$\hat{b}_m(f) = \sum_{t=1}^{\rho} I_{f=f_t} * \gamma_t(m) \Big/ \sum_{t=1}^{\rho} \gamma_t(m), \tag{11}$$

$$\hat{p}_m(d) = \sum_{t=2}^{\rho} \eta_t(m,d) \Big/ \sum_{t=2}^{\rho} \sum_d \eta_t(m,d). \tag{12}$$

The forward–backward algorithm is capable of calculating $\log\left(P\left[f_1^\rho|\lambda\right]\right)$ and new HsMM parameters only in $O\left(|\Theta|K + |\Theta|D\rho + |\Theta|^2\right)$ time complexity, where K is the number of distinct values that f can take on.

Reestimating λ is necessary only when c_k's task characteristic and its transition behavior change over time. If both of them are time-invariant, reestimation is unnecessary, and the detector needs to calculate only normality value related to Eq. (3) and Eq. (8).

Table 2 – Four different server configurations.

	eNB entropy ($b_n(e)$)	Wakeup packet generation rate ($b_n(r)$)
Configuration 1	Beta-binomial distribution	Truncated Poisson distribution
Configuration 2		Binomial distribution
Configuration 3		Hypergeometric distribution
Configuration 4		Integer-valued normal distribution

4. Simulation results

4.1. Simulation setup

In this section, we give simulation results to verify the superiority of our HsMM-based signaling attack detector. The simulation environment is implemented with R (CRAN-R, n.d., Bulla et al., 2010). We resort to empirical simulation with various control server configurations and attack types.

In our simulation environment, a normal control server has 10 tasks including a null task. State sojourn times are distributed with truncated logarithmic distributions with different parameters for each state. Both L and r_{max} are set to be 100 and uniform initial state distribution is assumed, $\pi_n = 0.1$ for all states. We assumed that the observation probability mass function is factorable as $b_n(e, r) = b_n(e) \cdot b_n(r)$ and simulated four different server setup configurations shown in Table 2.

We assume that a compromised control server tends to generate mixed traffic types of wakeup packets both for normal management process and from an attack burst. To take into account this assumption, we simulated four attack traffic types with different burst intervals and burst sizes whose values are either deterministic or random as shown in Fig. 8. When an attack burst is included, it induces a larger eNB entropy. In our simulation, a compromised control server's eNB entropy was sampled from a beta-binomial distribution which is skewed to larger values when it generates attack bursts.

We also implemented the CUSUM detector (Lee et al., 2009) with which the detection accuracy of our proposed detector is compared. To evaluate detection accuracy, we simulated the proposed detector and the CUSUM detector to report attack judgments based on observations of 50 consecutive OWs, i.e., with an open window size of 50. For each server configuration and attack type, we simulated more than 50,000 OWs.

4.2. Discussions

Fig. 9 compares the accuracy of the proposed detector and the CUSUM detector under deterministic attack (attack type-1). False positive ratio (FPR) and true negative ratio (TNR) are measures of report reliability, and they have a cost–benefit relationship with each other. Adjusting the detection threshold h, we can increase the detection sensitivity of our proposed scheme and the CUSUM detector. Increased detection sensitivity results in more attack alarms with higher FPR and TNR. According to Fig. 9, our proposed scheme benefits higher TNR, costing smaller FPR compared with the CUSUM detector. We observe that the proposed scheme achieves high detection accuracy satisfying high TNR (>0.9) while costing reasonable FPR (<0.1) in all simulation scenarios. On the other hand, the CUSUM detector costs more than 40% of FPR to have 80% of TNR. We can find the reason behind such performance gaps with Figs. 10 and 11.

Fig. 10 shows CUSUM statistic values of a normal control server and those mixed with attack type-1. The CUSUM statistic with attack type-1 has generally higher values. However, the CUSUM statistic of a normal server also fluctuates severely so that its maximum value is close to that with attack type-1. In the simulation of Fig. 10, the normal control server executes a task which has a very high wakeup packet generation rate from time 34,000 until 38,000. As a result, the CUSUM statistic peaks to 1060, which is close to the largest value (1090)

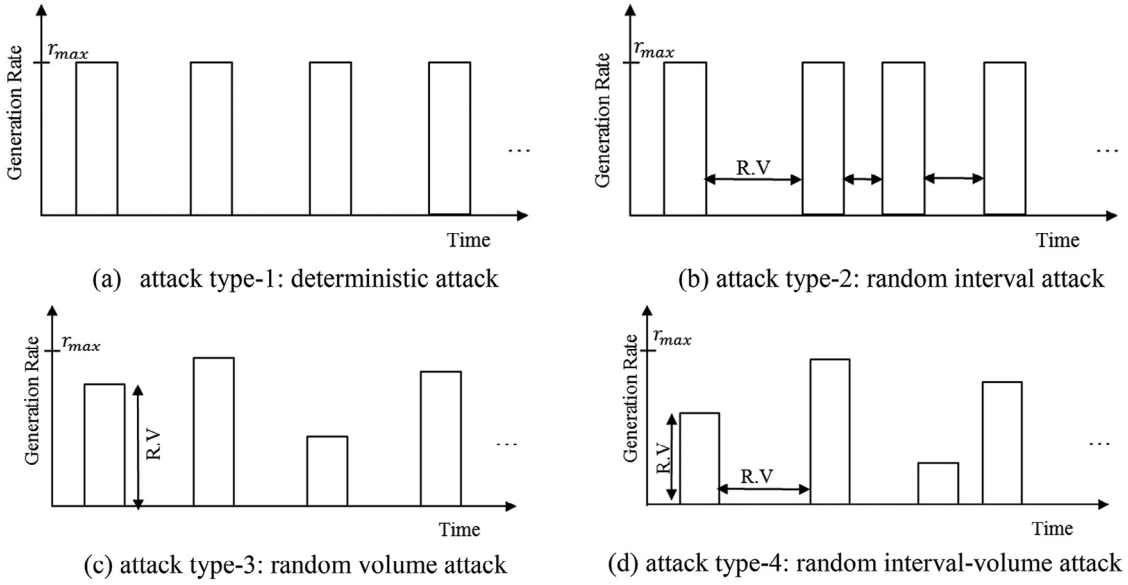

Fig. 8 – Four different attack types in wakeup packet generation.

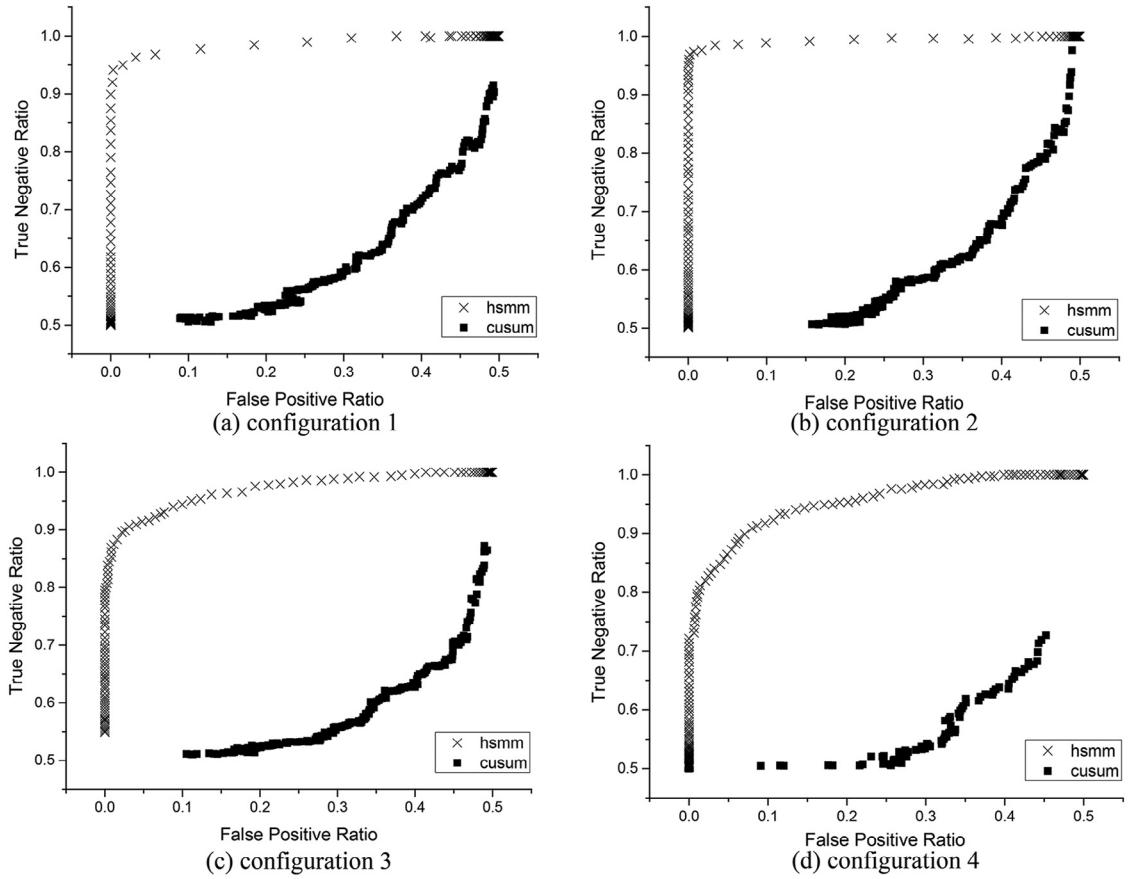

Fig. 9 – Accuracy comparisons of the two detection schemes under attack type-1.

of attack type-1's CUSUM statistic. Originally, the CUSUM statistic was designed to hover around 0 under normal condition. This means that the CUSUM notion about normal condition is assumed to have small behavior variance, which is quite inappropriate for describing the ordinary behaviors of control servers in WSANs. Usually, normal control servers which conduct real-time mission on a large scale tend to yield CUSUM statistics fluctuating like this. Therefore, the CUSUM statistic is unsuitable for signaling attack detection in WSANs.

Fig. 11 shows log-likelihood values of a normal control server and those with attack type-1 under various server configurations. The smallest value of the log-likelihood in the figures is –1000 because an R simulator variable has exp(–1,000) granularity. For all of the control server configurations, we can

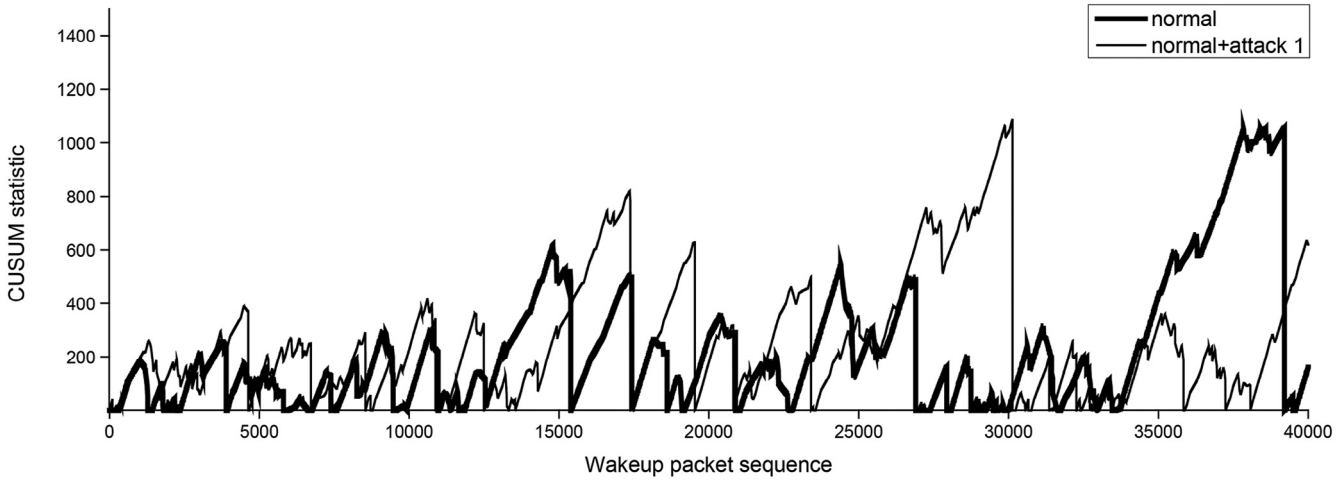

Fig. 10 – CUSUM statistic versus wakeup packet sequence.

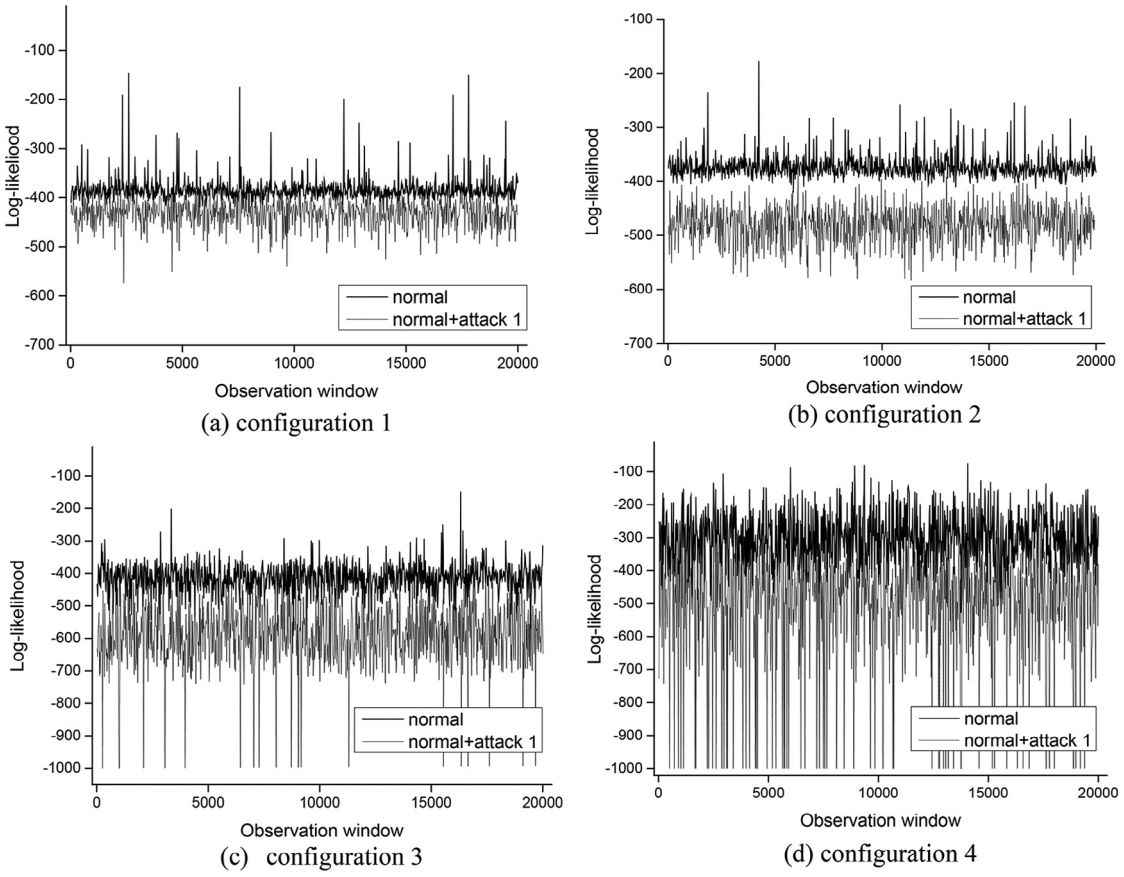

Fig. 11 – Log-likelihood for different server configurations.

effectively separate attack type-1 traffic from that of a normal control server by observing these values. The log-likelihood value of a normal control server has generally larger values and smaller variance, while that with attack type-1 has smaller values and larger variance. Therefore, after accumulating enough learning experiences the problem of determining a controllable decision threshold to obtain the required performance has a feasible solution for our proposed detector, which is almost unable to be achieved with the CUSUM statistic of Fig. 10. All of these observations from Figs. 10 and 11 confirm the effectiveness of the log-likelihood metric as a promising test criterion for normality.

Fig. 12 helps how to determine appropriate decision thresholds from analyzing log-likelihood distributions of a normal control server and attack type-1 under different server configurations. For example, in Fig. 11(b), the two histograms do not overlap except over a negligible region. If we set the decision threshold as –450, the proposed detector detects more than 95% of attack type-1 traffic without false alarm. In Fig. 11(d), the two histograms overlap each other in almost half of the region. Nonetheless, our proposed detector detects more than 20% of attack type-1 without false alarm if we set the decision threshold as –560. Although there is hardly any training data obtained under attack prior to installing our proposed scheme, simulation results in Fig. 12 assure that we can achieve low FPR and FNR by configuring the decision threshold as a value slightly lower than the minimum of log-likelihood of control server behavior.

Fig. 13 shows performance comparisons of the two detection schemes with the other three attack types under server configuration 2. We see that for all these attack types, our scheme also has the better detection performances as well. With less than 5% of FPR, our proposed detector achieves more than 90% of TNR. On the other hand, the CUSUM-based attack detector requires more than 45% of FPR to achieve the same TNR performance as ours.

Fig. 14 compares detection performance with different values of L and r_{max} under configuration 4. Larger values of L and r_{max} result in clearer behavioral difference of wakeup packet generation among tasks. As we can see in Fig. 14, increasing L and r_{max} results in better detection performance with higher TNR. As long as a control server operates legitimately, our proposed detector can infer the task transition more clearly if the behavioral differences are more discriminative. Therefore, the detector is more likely to judge whether the control server as legitimate.

5. Conclusions

In this paper, we focused on developing an LTE signaling attack detection scheme for a WSAN which is a particular mission-oriented network. In WSAN, an attacker easily succeeds in

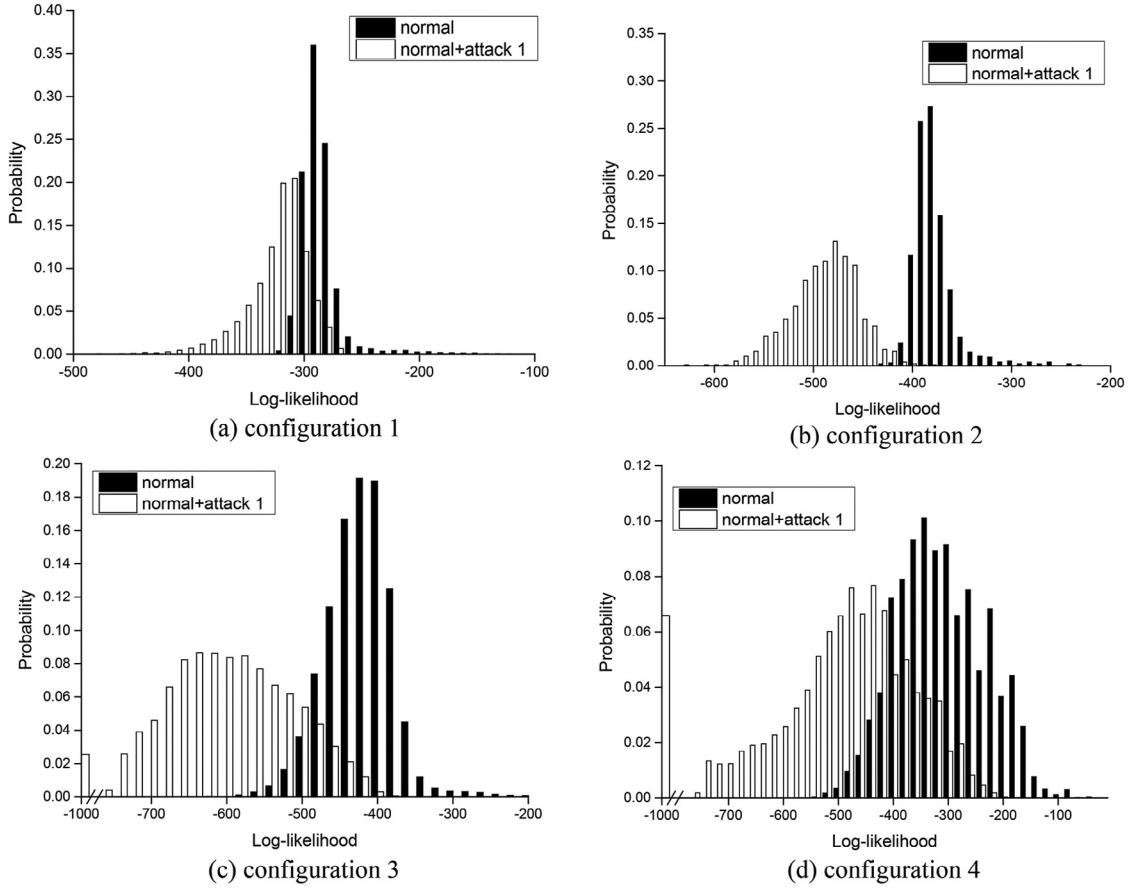

Fig. 12 – Determining appropriate decision thresholds for different server configurations.

degrading network QoS with relatively less attack traffic than is used with other types of aggression, such as the well-known Denial-of-Service attack. Furthermore, due to the enormous number of facilities involved in normal management processes, severe fluctuations of signaling traffic are to be expected with ordinary, normal operations. Therefore, conventional mean-rate-based attack detection schemes are inappropriate in our WSAN environment. We exploited the fact that there are task transitions in the scene behind the fluctuations. Furthermore, we take spatial characteristics of wakeup packet generation into consideration as a key observation to achieve better detection efficiency. We modeled these with HsMM and developed an advanced LTE signaling attack detector. We proved the superiority of our proposed attack detection scheme to a mean-rate-based detector, through various simulations.

Although this paper does not aim at giving general guidance for defining the task set Θ, we have experienced that how well Θ is defined decisively affects the performance of HsMM-based algorithms for WSANs, including our proposed scheme. An element of Θ can be either a single disjoint task or a partition of related tasks. In our case, we recommend that Θ be defined so that each element of Θ has distinct parameters for eNB entropy and wakeup packet generation rate distributions.

In our HsMM study, we assume discretized values of $E(t)$ and d to model the wakeup packet generation in WSAN. There is an application of HsMM to model continuous values of observations, each having Gaussian mixture distribution (Xie et al., 2013). However, adopting continuous random variables complicates the overall system modeling, and is not applicable to general cases.

Constructing initial λ requires sufficient amount of training network data, such as history about wakeup packet arrival to the LTE network or control server's task transition pattern. Xie and Yu (2009b) and Xie et al. (2013) argued that several hours of observing application layer message transaction are enough for obtaining accurate HsMM parameters. With larger amount of training data, HsMM-based attack detectors guarantee better detection accuracy. We cannot explicitly determine how much training data are enough for acceptable accuracy. Therefore, it is recommendable to collect as much training data as possible before installation of the detector.

Acknowledgement

This research was supported by a grant from the R&D Program of the Korea Railroad Research Institute, Republic of Korea.

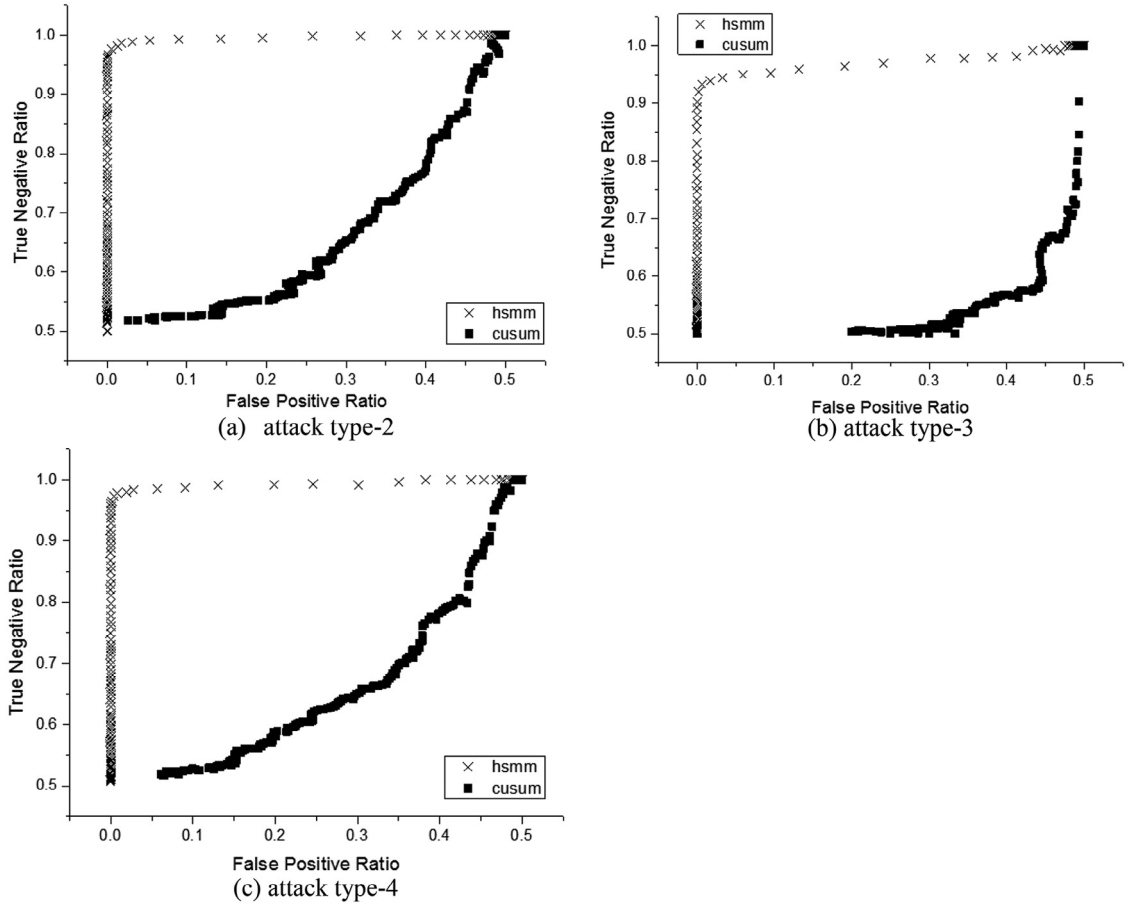

Fig. 13 – Performance comparisons under different attack types.

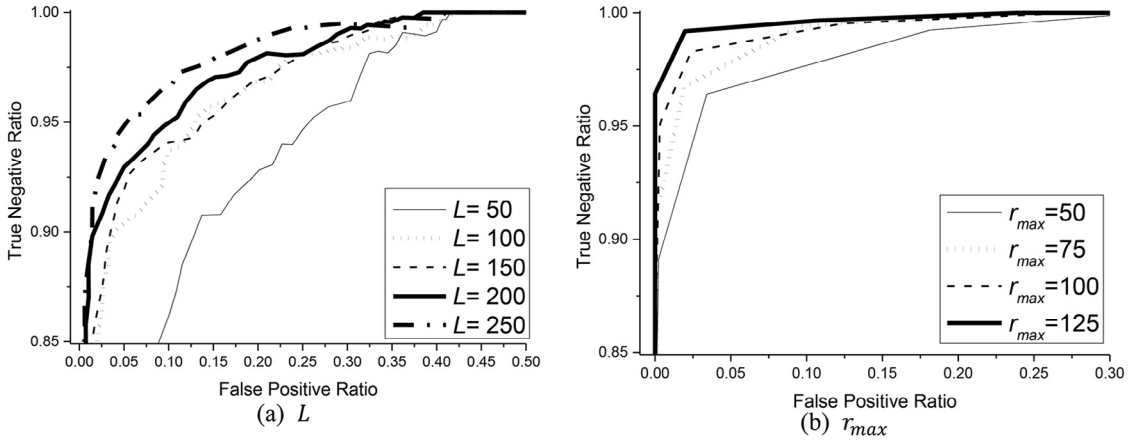

Fig. 14 – Performance comparison under different upper bounds on observation.

REFERENCES

Akerberg J, Gidlund M, Bjorkman M. Future research challenges in wireless sensor and actuator networks targeting industrial automation. In: 9th IEEE international conference on industrial informatics (INDIN). 2011. p. 410–15.

Bassil R, Chehab A, Elhajj I, Kayssi A. Signaling oriented denial of service on LTE networks. In: Proceedings of the 10th ACM international symposium on mobility management and wireless access. 2012. p. 153–8.

Bhuyan MH, Bhattacharyya DK, Kalita JK. Network anomaly detection: methods, systems and tools. IEEE Commun Surv Tutor 2014;16(1):303–36.

Bishop CM. Pattern recognition and machine learning. Springer; 2006.

Bulla J, Bulla I, Nenadić O. HsMM – an R package for analyzing hidden semi-Markov models. Comput Stat Data Anal 2010;54(3):611–19.

Calle-Sanchez J, Molina-Garcia M, Alonso JI. Top challenges of LTE to become the next generation railway communication system. WIT Trans Built Environ 2012;127:85–95.
Chen J, Cao X, Cheng P, Xiao Y, Sun Y. Distributed collaborative control for industrial automation with wireless sensor and actuator networks. IEEE Trans Ind Electron 2010;57(12):4219–30.
CRAN-R. n.d. CRAN. Available from: https://cran.r-project.org.
Guan K, Zhong Z, Ai B. Assessment of LTE-R using high speed railway channel model. In: Third international conference on communications and mobile computing (CMC). 2011. p. 461–4.
Gupta A, Verma T, Bali S, Kaul S. Detecting MS initiated signaling DDoS attacks in 3G/4G wireless networks. In: Fifth international conference on communication systems and networks (COMSNETS). 2013. p. 1–60.
Güngör VC, Hancke GP. Industrial wireless sensor networks: challenges, design principles, and technical approaches. IEEE Trans Ind Electron 2009;56(10):4258–65.
Güngör VC, Sahin D, Kocak T, Ergüt S, Buccella C, Cecati C, et al. Smart grid technologies: communication technologies and standards. IEEE Trans Industr Inform 2011;7(4):529–39.
Hasan M, Hossain E, Niyato D. Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches. IEEE Commun Mag 2013;51(6):86–93.
Holma H, Toskala A. LTE for UMTS evolution to LTE-advanced. Wiley; 2011.
Hong B, Peng F, Deng B, Hu Y, Wang D. DAC-Hmm: detecting anomaly in cloud systems with Hidden Markov Models. Concurr Comput Pract Exp 2015;27(18):5749–64.
Jang W, Kim SK, Oh JH, Im CT. Session-based detection of signaling DoS on LTE mobile networks. J Adv Comput Netw 2014;2(3):2–5.
Jover RP. Security attacks against the availability of LTE mobility networks: overview and research directions. In: 16th international symposium on wireless personal multimedia communications (WPMC). 2013. p. 1–9.
Kreher R, Gaenger K. LTE signaling, troubleshooting, and optimization. Wiley; 2010.
Lee PPC, Bu T, Woo T. On the detection of signaling DoS attacks on 3G wireless networks. Comput Netw 2009;53(1):2601–16.
Lien SY, Chen KC, Lin Y. Toward ubiquitous massive accesses in 3GPP machine-to-machine communications. IEEE Commun Mag 2011;49(4):66–74.
Lien SY, Liau TH, Kao CY, Chen KC. Cooperative access class barring for machine-to-machine communications. IEEE Trans Wirel Commun 2012;11(1):27–32.
Peng C, Li W, Bin Z, Shihua W. Feasibility study of applying LTE to Smart Grid. In: 2011 IEEE first international workshop on smart grid modeling and simulation (SGMS). 2011. p. 108–13.
Rao VS, Gajula R. Protocol signaling procedures in LTE. Radisys; 2011. p. 11. White Paper.
Shi S, Sun M. Study on HMM based anomaly intrusion detection using system calls. In: 2nd international conference on electronic & mechanical engineering and information technology. 2012. p. 139–44.
Souryal MR, Golmie N. Analysis of advanced metering over a Wide Area Cellular Network. In: 2011 IEEE international conference on smart grid communications (SmartGridComm). 2011. p. 102–7.
Sugaya M, Ohno Y, Nakajima T. Lightweight anomaly detection system with HMM resource modeling. Int J Secur Its Appl 2009;3(3):35–54.
Talukder ZH, Islam SS, Mahjabeen D, Ahmed A, Rafique S, Rashid MA. Cell coverage evaluation for LTE and WiMAX in wireless communication system. World Appl Sci J 2013;22(10):1486–91.
Tingting G, Bin S. A high-speed railway mobile communication system based on LTE. In: 2010 international conference on electronics and information engineering (ICEIE). 2010. p. V1-414–V1-417.
Xie Y, Yu SZ. A large-scale hidden semi-Markov model for anomaly detection on user browsing behaviors. IEEE/ACM Trans Netw 2009a;17(1):54–65.
Xie Y, Yu SZ. Monitoring the application-layer DDoS attacks for popular websites. IEEE/ACM Trans Netw 2009b;17(1):15–25.
Xie Y, Tang S, Huang X, Tang C, Liu X. Detecting latent attack behavior from aggregated Web traffic. Comput Commun 2013;36(8):895–907.
Yan Y, Qian Y, Sharif H, Tipper D. A survey on smart grid communication infrastructures: motivations, requirements and challenges. IEEE Commun Surv Tutor 2013;15(1):5–20.
Ye X, Li J, Li Y. An anomaly detection system based on Hide Markov Model for MANET. In: 6th international conference on wireless communications networking and mobile computing (WiCOM). 2010. p. 1–4.
Yu SZ. Hidden semi-Markov models. Artif Intell 2010;174(2):215–43.
Yu SZ, Kobayashi H. An efficient forward – backward algorithm for an explicit-duration Hidden Markov Model. IEEE Signal Process Lett 2003;10(1):11–14.

June-ho Bang is a Ph.D. student in the School of Computer Science and Engineering at Ajou University in Suwon, Korea. He received his M.S. degree in computer science from the university in 2012. He is currently interested in network optimization, performance analysis, and statistical anomaly detection system.

Young-Jong Cho received a B.S. degree in Electronic Engineering from Seoul National University, in 1983 and M.S. and Ph.D. degrees in Electronic Engineering from KAIST in 1985 and 1989, respectively. Since 1996, he has been a professor of the Department of Information and Computer Engineering at Ajou University. From 1985 to 1996, he had been a Principal Engineer at LG Electronics Co., Korea. In 1993, he visited AT&T Bell Lab. as a Visiting Research Fellow. In 2003, he was a visiting professor at GMU. His current research interests include performance analysis of mobile communication systems and statistical analysis of network data.

Kyungran Kang received a B.S. degree in Computer Engineering from Seoul National University, Korea in 1992, and an M.S. degree and Ph.D. degree in Computer Science from KAIST, Korea in 1994 and 1999, respectively. She has been a professor at Ajou University from 2004. Her research interests include computer networks and community computing.