Intrusion Detection Techniques in Wireless Sensor Network using Data Mining Algorithms: Comparative Evaluation Based on Attacks Detection

YOUSEF EL MOUROUBIT

ANOUAR BOUIRDEN
Laboratoire Thermodynamique et Energétique
Faculty of Sciences, Ibn Zohr University
AGADIR, MOROCCO

AHMED TOUMANARI
Equipe Signaux, Systèmes et Informatique (ESSI)
National School of Applied Sciences, Ibn Zohr University
AGADIR, MOROCCO

NADYA EL MOUSSAID
Equipe Signaux, Systèmes et Informatique (ESSI)
National School of Applied Sciences, Ibn Zohr University
AGADIR, MOROCCO

***Abstract*—Wireless sensor network (WSN) consists of sensor nodes. Deployed in the open area, and characterized by constrained resources, WSN *suffers* from several *attacks, intrusion and security vulnerabilities.* Intrusion detection system (IDS) *is one of the essential* security mechanism against attacks in WSN. In this paper we present a comparative evaluation of the most performant detection techniques in IDS for WSNs, the analyzes and comparisons of the approaches are represented technically, followed by a brief. Attacks in WSN also are presented and classified into several criteria. To implement and measure the performance of detection techniques we prepare our dataset, based on KDD'99, into five steps, after normalizing our dataset, we determined normal class and 4 types of attacks, and used the most relevant attributes for the classification process. We propose applying CfsSubsetEval with BestFirst approach as an attribute selection algorithm for removing the redundant attributes. The experimental results show that the random forest methods provide high detection rate and reduce false alarm rate. Finally, a set of principles is concluded, which have to be satisfied in future research for implementing IDS in WSNs. To help researchers in the selection of IDS for WSNs, several recommendations are provided with future directions for this research.**

Keywords—Keyword: Wireless sensor network; Anomaly Detection; Intrusion detection system; classification; KDD'99; Weka

I. INTRODUCTION

Wireless sensor networks are composed of several sensors deployed in areas where the aim is to collect data and forward it for the analysis. It has become an increasingly interesting field of research in solving such challenging real-world problem, as environmental monitoring [1], military applications, geographical sensing, traffic control, and home automation. The properties of WSN show that that sensor node is completely restricted by resources, including memory, energy, computing, communication and bandwidth. [2]. Therefore, the deployment of these kinds of networks with their resource restrictions makes their security issue essential, and vulnerable to various security threats. Key management and authentication have been used to protect WSNs from different attacks, encryption and authentication are the first security measures as the first line of defense for protecting WSN [3]. But cryptography based on secret key management are not enough to protect the WSN, because even in the presence of this first line of defense, several attacks may extract sensitive information, and use them for malicious reason. However, Detection-based approaches are then proposed to protect WSNs from intrusion and attacks, as a second line defense, after the failure of the cryptographic techniques [4], Intrusion detection system (IDS) observes and analyzes the events generated in the network system to identify maximum security problems. IDSs are used to monitor the network to detect anything unusual. [5]. This concept was originally proposed by Anderson [6]. There are two principal approaches for detection, intrusion: Misuse detection based on rules, these rules will look for signatures on the network and then system operations try to catch known attack that should be considered as Misuse [7] [8]. Anomaly detection [9], which based on the normal behavior of a system, it compares normal activities against observed events to identify significant deviations. The main scope of this paper is to improve that random forest technique is an efficient anomaly detection technique for IDS in WSN, with a comparative evaluation study for the most recent and performants anomaly detection technique used in IDS for WSN. In Section 2 we present a classification of existing attacks in WSN by several criteria. Section 3 introduces a survey of ids in WSN, and analyzes four recent anomaly intrusion detection techniques using in IDS for WSN: (K-means, Naives Bayesian, SVM, Random Forest), showing their principles, advantages and drawbacks.

Simulation environment and results are presented in section 4, we simulated last techniques on KDD dataset using Weka tool, and results are based on matrices of confusion, detection rate, time of execution and memory consumption. At the end of the paper a conclusion is introduced, and a set of recommendations are suggested to boosting the performance of intrusion detection in WSN for future researches.

II. ATTACKS CLASSIFICATION IN WSN

An attack is a set of techniques, used to cause damage to a network by exploiting flaws in it. Attacks know several possible classification, the most used are grouped into the following categories:

A. According to the origin or source attacks:

Two categories are distinguished: internal and external attacks: An external attack is triggered by a node that does not belong to the network or does not have access permission. The aim of this attack is to cause congestion in the network, the spread of incorrect routing information, or completely close the network. The internal attack is done by a malicious internal node. Defense strategies generally aim to protect the network against external attacks. However, internal attacks are the most serious threats that can disturb the WSN [10][11].

B. Based on the nature of attacks:

We can distinguish between passive and active attacks, the passive attack is limited to listening and analyzes exchanged traffic. This kind of attacks is difficult to detect and easier to realize, because the attacker does not make any modification on exchanged information.

The aims of the attacker can be the knowledge of the significant nodes in the network (cluster head node), or knowledge of confidential information by analyzing routing information. In the active attacks, an attacker tries to modify or remove the messages transmitted on the network, inject his own traffic or replay of old messages to disturbing the operation of the network. [12].

C. Classification by attacks techniques:

The spoofed, altered, or replayed routing information attack, and sinkhole attack: need to make a probe step before starting to attack, thereforeattack we can classified these attack as probe attacks. Selected forwarding, jamming, tampering: which uses illegitimate data forwarding to make attack, is known as a dos attacks? Hello floods caused by internal attacks, is classified as U2R attack. Sybil, wormholes, hello floods, and acknowledgment spoofing make the attack through the weakness of the system then they would be classified as R2L attack. In the table below we present the following main types of attacks, sorted by four principals attack classes.

TABLE I. ATTACKS CLASSES

Attack class	Attack techniques
Probe	Spoofed Routing Information attack, Altered Routing Information Attacks, Replayed Routing Information, Sinkhole
DOSS	Selected Forwarding, Jamming, Tampering
U2R	Hello Floods
R2L	Sybil, Wormholes, Achnowledgement Spoofing

D. According to the various protocol layers and proposed mechanism defense:

The following main types of attacks, are sorted by their assignments to the relevant layers of the protocol stack. For each attack, a list of proposed mechanism defense is presented [13][14]:

Protocol layer	Attacks	Defenses
physical	Jamming	Priority messages, monitoring, authorization, redundancy, encryption[14] Spread-spectrum, priority message, lower Duty cycle, region mapping, mode change Tamper-proofing, hiding
	Tampering	
Data link	Collision	Error-correction code
	Exhaustion	Rate limitation
	Unfairness	Small frames
network	Spoofed, Altered or relayed routing information	Detection on MintRoute[4]
	Selective forwarding	
	Sinkhole	
	Sybil attack	Identity certificates[11]
	Wormholes	Dawwsen proactive routing rotocol[13] suspicious node detection by signal strength,[10]
	Hello flood attacks	Suspicious node detection by signal strength[10]
	Achnowledgment spoofing	Encryption,authentication, monitoring
Transport	Flooding	Client puzzles
	De-Synchronization	Authentication

TABLE II. ATTACKS, PROTOCOL LAYER AND DEFENSE MECHANISM

III. RELATED WORK

It has become clear that we cannot achieve the satisfactory level of security in WSN only by using cryptographic techniques, as these techniques fall prey to insider attacks. The attacker can compromise and retrieve the cryptographic material of a number of nodes [15]. In order to counter this threat some additional techniques such as intrusion detection system (IDS) has to be deployed. Any kind of unauthorized or unapproved activities are called intrusions. An IDS is a collection of the resources, methods and tools to help identify, evaluate, and report intrusions [16]. WSN led researchers develop strategies about providing stable networking and communications, and also about how to secure these strategies with limited resources.

In [17], a hierarchical framework for intrusion detection as well as data processing is proposed. Throughout the experiments on the proposed framework, they stressed the significance of one hop clustering. The authors believed that their hierarchical framework was useful for securing industrial applications of WSNs with regard to two lines of defense. Krontiris et al. [18] proposed a distributed IDS for WSNs based on collaborative neighborhood watching. In a simulation environment, the authors evaluated the effectiveness of their IDS scheme against blackhole and selective forwarding attacks. In [19] provided an IDS for WSNs that was based on detection of packet level receive power anomalies. The detection scheme was focused on transceiver behaviors and packet arrival rates of the neighboring nodes of a particular node. In [20], a distributed cluster based anomaly detection algorithm was proposed. They minimized the communication overhead by clustering the sensor measurements and merging clusters before sending a description of the clusters to the other nodes. The authors implemented their proposed model in a real-world project. They demonstrated that their scheme achieves comparable accuracy when compared to centralized schemes with a significant reduction in communication overhead. The table below presents a brief list of constraints and the corresponding requirements of IDS in WSN:

TABLE III. IDS REQUIREMENT FOR WSN

Constraints and challenges of WSN	**Requirement of IDS**
• *N*o infrastructure in WSNs to support operations such as communications, routing, real time traffic analysis, encryption, etc. • Nodes are prone to physical capture, tampering or hijacking which compromises network operations. • Compromised nodes may provide misleading routing information to the rest of the WSN leaving the network un-operational (blackhole, wormhole, sinkhole attacks). • Wireless communication is susceptible to eavesdropping, which would reveal important data to adversaries and/or to jamming/interfering, which would cause DoS in the WSN. • There is no trusted authority; decisions have to be concluded in a collaborative manner.	• *N*ot introduce new weaknesses to the system, • Need little system resources and should not degrade overall system performance by introducing overheads, • Run continuously and remain transparent to the system and the users, • Use standards to be cooperative and open, • Be reliable and minimize false positives and false negatives in the detection phase.

There are two basic approaches in IDS according to the used detection techniques [21]:

Misuse detection technique compares the observed behavior with known attack patterns (signatures). Action patterns that may pose a security threat have to be defined and stored in the system. The advantage of this technique is that it can accurately and efficiently detect instances of known attacks, but it lacks an ability to detect an unknown type of attack.

Anomaly detection: The detection is based on monitoring changes in behavior, rather than searching for some known attack signatures. Before the anomaly detection based system is deployed, it usually must be taught to recognize normal system activity (usually by automated training). The system then watches for activities that differ from the learned behavior by a statistically significant amount. The main disadvantage of this type of system is high false positive rate. The system also assumes that there are no intruders during the learning phase.

Anomaly may be caused by security threats, or faulty sensor nodes in the network or unusual phenomena in the monitoring zone [22]. Isolated node failures can bring down the whole network, which is malicious to reliability of WSN. Researches in this field are yet absent to present the latest progress of developing anomaly detection in WSN. However, our paper expects acting as a guideline of selecting efficient and appropriate anomaly detection techniques, not just based on analyzing, comparing, and evaluating those particular approaches, but also according to the results of simulation, which shows the classification rate, confusion matrix, consumption of memory, and time to build every approach.

IV. RSTUDY ANALYSIS AND EVALUATION OF ANOMALY DETECTION TECHNIQUES IN WSN

A. Clustering approach

With K-means clustering algorithm, Rajasegarar et al [20] design a distributed detection scheme. Each common sensor node locally collects the input dataset to work out a normal profile. Then the cluster head collects all local normal profiles to accomplish the procedure of data processing, where a global normal profile is produced. After received the global normal profile, each common sensor node initiates the analysis and decision procedure to perform detection. In order to fit in distance-based clustering, the input dataset is normalized at each common sensor node with a preprocessing procedure.

Given a dataset v_{kj} , k=1…m, it is transformed to $u_{kj=(v_{kj}-\mu_{vj})/\delta_{vj}}$

Where μ_{vj} and δ_{vj} stand for the mean and standard deviation of the jth attribute in v_{kj} . Subsequently u_{kj} is normalized in the interval [0,1], according to

$\overline{u}_{kj} = (u_{kj} - minu_j)/(maxu_j - minu_j).$

Given a common sensor node s_i collecting a dataset X_i , s_i sends the local normal profile.

$(\sum_{k=1}^{m} x_k^i \sum_{k=1}^{m}(x_k^i)^2 , m, x_{max}^i, x_{min}^i)$

to the cluster head, where m stands for $|X_i|$. After the global normal profile is computed,

$(\mu_j, \delta_G^2, x_{max}^G, x_{min}^G)$. The cluster head sends it back to the common sensor nodes. After receiving the global normal profile, each common sensor node initiates detection locally, using a fixed-width clustering algorithm. If the Euclidean distance between a data point and its closest cluster centroid is larger than a user-specified radius o, a new cluster is organized with this data point as centroid. For reducing the number of resulting clusters, a cluster merging process is then conducted, through measuring the inner- cluster distances[35]. The clusters c1 and c2 merge if their inner- cluster distance d(c1,c2) is less than o. Finally, the average inter- cluster distance of K nearest neighbor (KNN) clusters is applied to identify anomalous clusters. Let ICDi be the average inter-cluster distance (KNN) of cluster i, AVG (ICD) and SD(ICD) be the mean and standard deviation of all inter-cluster distances respectively. If : ICDi>SD(ICD) + AVG(ICD), cluster i is viewed as anomalous[35].

B. Support Vector Machine Classifier

Support Vector Machines (SVMs) are supervised learning algorithms [24], which have been applied increasingly to anomaly detection in the last decade. One of the primary benefits of SVMs is that they learn very effectively from high dimensional data [25]. In WSN SVM is used to investigate spatial and temporal correlations of data for detecting suspect behavior of a node. Many researchers have tried to find possible methods to apply SVM classification for large data sets. Sequential Minimal Optimization (SMO) is a fast method to train SVM [26], which breaks the large Quadratic Programming (QP) problem into a series of smallest possible QP problems. In [27] Kim et all applied SVMs to host based anomaly detection of masquerades. One-class quarter-sphere SVM, as a representative algorithm of SVM, is also suited to distribute anomaly detection [28]. First, the local quarter-sphere is computed at each common sensor node. Second, the cluster heads collects these locally computed radii to work out a global radius. Detection is then launched at each common sensor node with the global normal profile.

C. Naïve Bayes Classifier

The naive Bayes classifier is usually used in WSN because of its simplicity, elegance, and robustness. A large number of modifications have been introduced, by the statistical, data mining, machine learning, and pattern recognition communities, in an attempt to make it more flexible. Novel approach was proposed in [29] to identify the faulty sensor node using Naïve Bayes classifier. The proposed Naïve Bayes framework was deployed for performing WSN faulty node(s) detection. A new attribute, the end-to-end transmission time of each packet arrived at the sink is analyzed using Naïve Bayesian classifier for determining the network status. This technique doesn't involve any additional protocol and extra resource consumption of sensor node, it suggests a list of suspicious faulty nodes to the user [29]. In the same context, based on mobile agent and using naïve Bayesian classifier an IDS is presented in [23]. The figure below presents the principal of naive Bayesian classifier.

m Number of classes C1, C2,….,Cm
d_{ct} Dimentional vector for class t d_{ct}= {dct1,dct1,…….,dctn}
where $\sum_i d_{cti}$=1
K total ksenses of network operation S = {S_1, S_2,….., S_k}
S_1 Is a product of the data that appear in the scene:

$$P(S_1|d_{ct})= \frac{(\sum_i N_i)!}{\prod_i N_i!} \prod_i (d_{cti})^{N_i} \qquad (1)$$

Where N_i is the number of data I in sceneS_1.

$$L= \arg max_c \; [\; \log P(D_{ct}) + \sum_i N_i \; log d_{cti}] \qquad (2)$$

D. Random Forest Classifier :

Random forests are based on collection learning method for classification, that operate by constructing a multitude of decision trees, at training time and outputting the class, that is the mode of the classes output by individual trees. Random tree, on the other hand, involves construction of multiple decision trees randomly [30]. Each tree is constructed using the following algorithm:

Step1: Let the number of training cases be *N*, and the number of variables in the classifier be *M*. ***Step2***: We are told the number *m* of input variables to be used to determine the decision at a node of the tree; *m* should be much less than *M*. ***Step3***: Choose a training set for this tree by choosing *n* times with replacement from all *N* available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes. ***Step4***: For each node of the tree, randomly choose *m* variables on which to base the decision at that node. Calculate the best split based on these *m* variables in the training set. ***Step5***: Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier). A novel data mining approach based on random forests was proposed to characterize and classify a similar large scale physical environment in [31]. The proposed data mining formulation, allows better performance in terms of tradeoff between energy efficiency and accuracy. Compared to a single decision tree algorithm, RFs runs efficiently on large datasets with a better performance. In [30] Random Forests (RF) is used as a classifier for the proposed intrusion detection framework. RF gives better performance in designing IDS that is efficient and effective for network intrusion detection. the advantages and inconveniences of the studied techniques are presented in the following table:

TABLE IV. ADVANTAGES AND INCONVENIENCES OF STUDIED TECHNIQUES

approach	**advantages**	**inconveniences**
K-means	-Fast and easier to understand. -Gives best result when data set are distinct.	-Sensitive to initialization -Low detection accuracy
Naïve-bayes	-Low computation complexity -High detection accuracy	-Increased communication overhead required for sending full data from common nodes to cluster heads. -Central point of failure as anomalous detection is accomplished only at cluster heads
SVM	-No central points of failure, all nodes have the same capability of detection -Reduced energy consumption by transmitting support vectors between nodes instead of all captured data	There must be an efficient way to select relevant features instead of delete one at a time and rank the important one the biggest limitation of the support vector approach lies in choice of the kernel
Random Forest	-Runs efficiently on large databases -Provides effective methods for estimating missing data -High detection accuracy and low false positive rate.	have been observed to over fit for some datasets with noisy classification/regression tasks the variable importance scores from random forest are not reliable for all types of data

V. EXPERIMENT RESULTS

A series of experiments were conducted to simulate and evaluate each approach, to define the efficient detection technique for ids in WSN. We used several critical evaluation metrics: Confusion matrix, general classification rate, time to build model, memory consumption. We prepared our data set, based on the standard KDDCup'99 intrusion detection dataset [32], into following five step, using Weka tool:

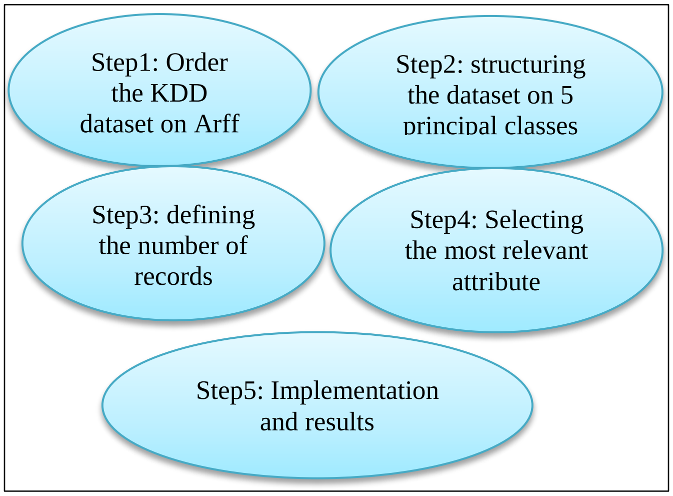

Step1: Order the KDD dataset on Arff
Step2: structuring the dataset on 5 principal classes
Step3: defining the number of records
Step4: Selecting the most relevant attribute
Step5: Implementation and results

Fig. 1. Preparing dataset steps

Step1: in this step we structured all records on Attribute-Relation File Format (ARFF), which is an input file format used by the machine learning tool WEKA [33].

Step2: In this step we classed all types of attacks, on four principal categories. As shown the table [2].

Step 3: the main aims of this step is defining the number of records treated for each class as presented in table below, we used 70% in training stage and 30% in the test stage for each class.

TABLE V. NUMBER OF RECORDS

Class	Instances Number
Normal	10233
Dos	41748
Probe	441
R2L	96
U2R	92

Step4: In general, a characteristic is good if it is relevant to the concept of class but not redundant to one of the other functions. Reduction of the attributes is a process of choosing a subset of the original attributes which feature space is reduced optimally at an endpoint.

In our experiment, Weka tool is used for reduction function. **CfsSubsetEval** with **BestFirst** approach is applied to the set of training data to obtain the relevant features for the classification process. Each subset was analyzed using correlation analysis to identify important features. The best known Measuring correlation is the linear correlation coefficient. For a pair of variables (x, y), the linear correlation coefficient r (x, y) is given by the expression below:

$$r(x,y) = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^4 - (\sum x)^4)(n\sum y^4 - (\sum y)^4)}} \ldots$$

The main principle of CfsSubsetEval method is evaluating the value of a subset of attributes by considering the individual predictive ability of each element as well as the degree of redundancy between them. It generates subsets of features that are highly correlated with the class while having a low cross correlation [34]. The results are presented in the table below:

TABLE VI. MOST RELEVANT ATTRIBUTES

Search Method	CFS Subset Evaluator + Best first
Selected attributes	5,6,9,11,12,14,31,32
Attributes names	src_bytes; dst_bytes; urgent; num_failed_logins; logged_in; root_shell; srv_diff_host_rate; dst_host_count

Step5: In this step we implemented each technique on our dataset, using Weka tool. Below the result obtained based on confusion matrix, detection rate, time of execution and memory consumption.

A. Confusion Matrix:

In order to assess these techniques we take the confusion matrix, illustrated below:

TABLE VII. CONFUSION MATRIX APPROACHES

K-means confusion matrix

Classified Attacks	a	b	c	e	f
Normal	4090	6106	0	0	37
Dos	4808	31254	0	0	5686
U2r	37	55	0	0	1
R2L	38	58	0	0	1
Probe	148	151	0	0	142

Naïve Bayes confusion matrix

Classified Attacks	a	b	c	e	f
Normal	8253	150	36	709	1085
Dos	309	39189	4	10	2236
U2r	0	0	92	1	0
R2L	4	0	8	82	3
Probe	9	5	0	15	412

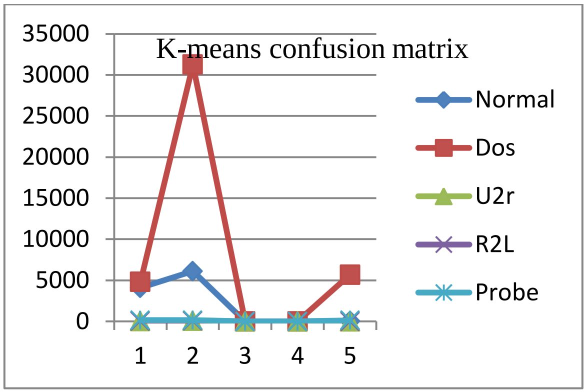
SMO confusion matrix

Classified Attacks	a	b	c	e	f
Normal	10207	15	2	9	0
Dos	13	41735	0	0	0
U2r	1	0	92	0	0
R2L	14	0	0	83	0
Probe	23	0	0	0	418

Random forest confusion matrix

Classified Attacks	a	b	c	e	f
Normal	10230	0	0	3	0
Dos	2	41745	0	0	1
U2r	1	0	92	0	0
R2L	8	0	0	89	0
Probe	7	2	0	0	432

Generally each column of the matrix represents the number of occurrences of an estimated class, while each row represents the number of occurrences of a real class (or reference). The results are presented in the following figures:



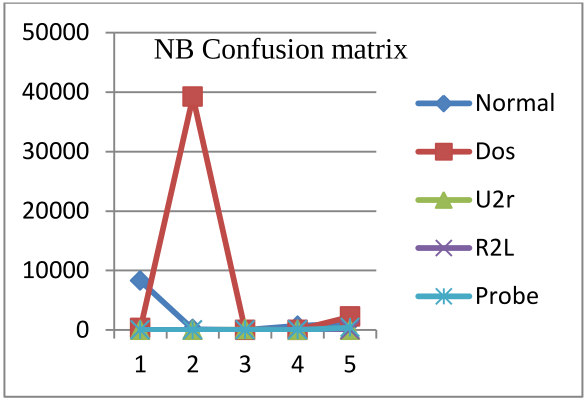
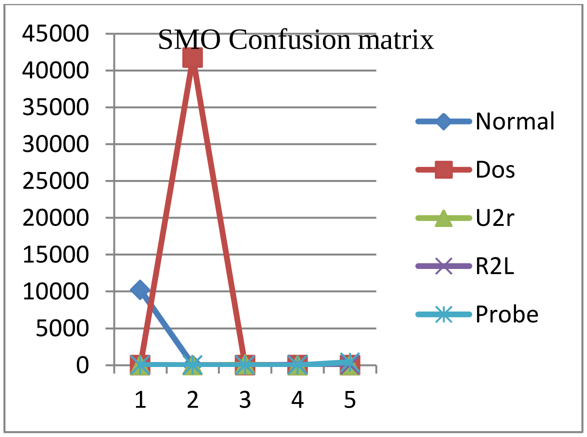


Fig. 2. Confusion Matrix Approach



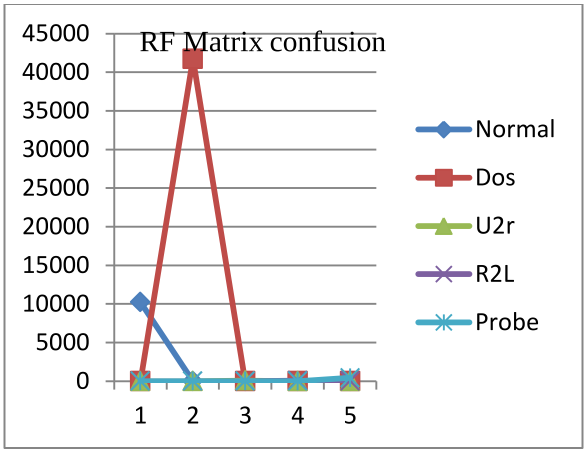


Fig. 3. Confusion Matrix Approach

According to results above and Based on Dos attack: K-means a classify 31254 Dos attack from 41749 real dos attack (74,86%), however 6106 instances are classified into normal class, 55 as U2R attack, 58 as R2L attack and 151 as Probe attack. Naïve bayes is able to classify 39189 Dos attacks from 41749 real Dos Doss attack (93.87%), while 150 instances is classified as normal attack, and 5 as Probe. SMO classified 41735 Dos attack from 41749 (99,96%), and 15 instances into normal class. Finally random forest classified 41745 Dos attack from 41749 real Dos attack (99,99%), and 2 instances as a Probe attack.

B. Classification Rate

The purpose of classification is to minimize the probability of error Detection algorithms are usually evaluated using the detection rate. A simple way to perform an intrusion detection, is to use a classifier to determine whether certain traffic data observed is normal or attacks. We present the classification rate on two sides: Global records classification and general rate classification.

Global records classification: The table below presents for each technique the global number of correctly and incorrectly classified records:

TABLE VIII. INSTANCES CLASSIFICATION

Approach	Correctly Classified Instances	Incorrectly Classified Instances
K-means	35486	17126
Naive bayes	48028	4584
SMO	52535	77
Random forest	52588	24

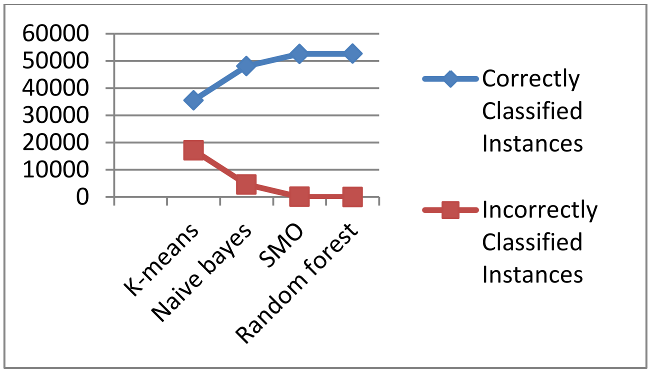


Fig. 4. Instance Classification

As shown in figure above, we note that random forest has the higher number of correctly classified instances and the lower number of incorrectly classified instances, however we observe the complete opposite for K-means technique.

General rate classification: The following figures represents the rate classification of each class, normal class is represented by the Blue color, Red for Doss class, U2R Blue sky, green for R2L class, and pink color for the Probe class. A better classification is obtained if the represented classes are well separated. According to the results we deduce that the Random Forest classifier is more effective and efficient than other approaches with a classification rate of 99.9544%.

Below the Complexity variables: (**N**: instances number, **M**: Attribute number, **C:** Classes number, **V**:attribute value).

According to the results, the SVM method is the most complex [0((NM)^2] , which explains its high memory consumption with 38,444KB, and his long time compilation.

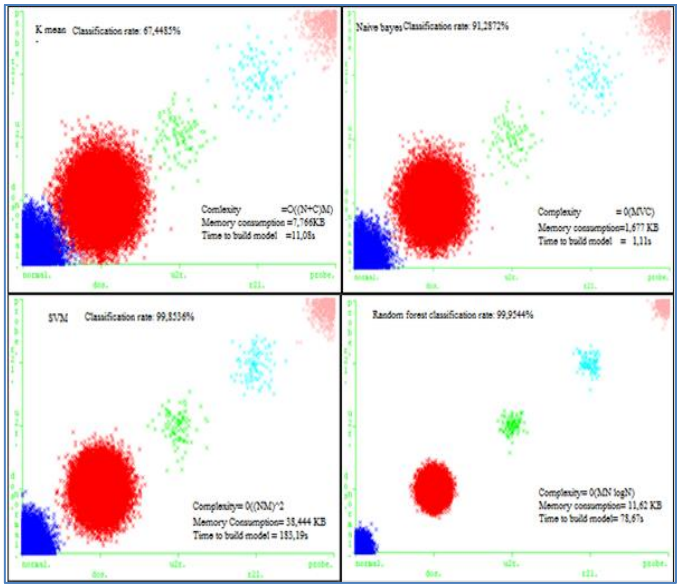


Fig. 5. Classification rate

The memory consumption of these techniques are compared with properties of sensor node that we can use in deployment of wireless sensor network, we choose MICA2 and Telosb. Knowing that MICA2 is equipped with a processor running at 7.37 MHz, 4KB of RAM, 128KB of flash memory and a radio transmitter on 433 MHz. For Telosb, is equipped with an 8 MHz clock processor, 10K RAM, 48K of program memory, and 1024K flash storage.

In the figure 5, we compare the memory consumption of studied techniques and node sensor ability. Time to build the approaches is presented in figure 6.

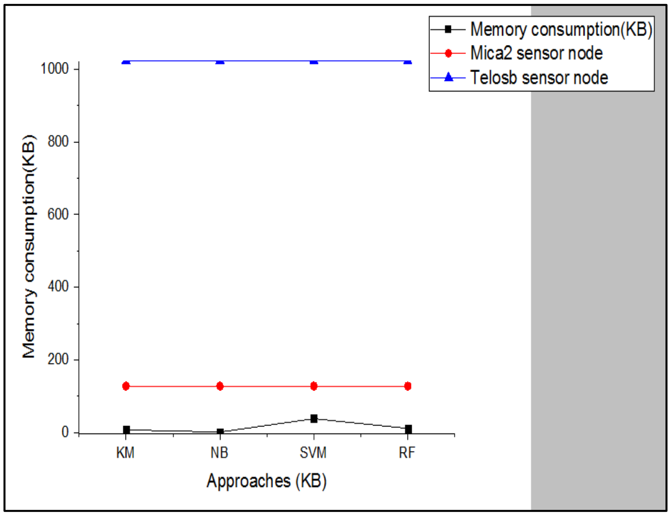


Fig. 6. Memory Consumption

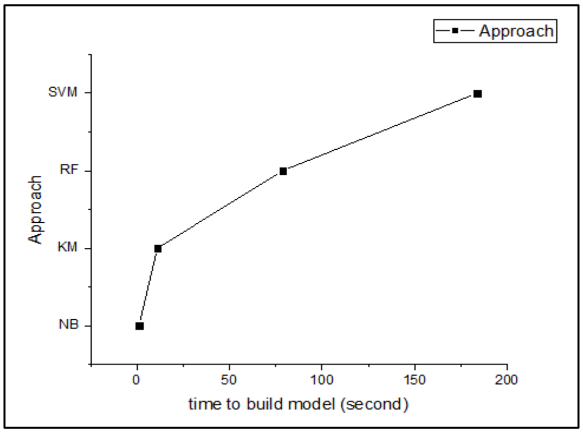


Fig. 7. Approaches building time

According to results, it is clear that memory is enough to compile each approach on Mica2 node or Telosb node, but for increasing the lifetime of the node, and taken on consideration the main aims of these techniques, detecting the different attacks (classification rate),we can say that Random forest technique is the efficient technique for detecting intrusion in wireless sensor network, with a higher rate classification (99.9544 %), reasonable required memory (11,62 KB), and building time(78,67 s). Indeed, the superiority of Random Forest intrusion detection technique, SVM, Naïve Bayes and K-means respectively, can be clearly deduced, in this order, according to confusion matrix, classification rate, memory, complexity, building time and memory consumption we can classify these techniques, from the higher to lower performant technique. Classification based on suitable feature selection is one of the main factors which reach the performance of IDS, especially in WSN.

VI. Conclusion

The key challenge of evolving intrusion detection system in WSN is to identify attacks with high accuracy, and satisfied the required constraints and challenges, to prolong the lifetime of the entire network. This aims could be attained from several ways. Firstly paying much more attention to detection techniques used for attacks detection is characterized by efficiency and ability. Secondly, reconstructing detection mechanism with a distributed manner, to reducing the communication overhead. This paper has compared and evaluated the newest anomaly detection intrusion techniques used in wireless sensor network, to improve the efficient technique for IDS in WSN. According to the results, it is highly recommended to use the data mining techniques to detect effectively the intrusions and attacks in WSN. The decision of choosing efficient IDS is a compromise between technique employed and performance metrics. However, many issues are still open and need further research efforts such as hierarchical clustering patterns, using machine learning in resource management problem of wireless sensor networks, developing a classifier that is trained well with network patterns, selecting and preprocessing an appropriate dataset. In addition, taking smart strategies into account such as compressing the input dataset, narrowing the scale of attributes set and simplifying the procedure of analysis and decision could make lots of progress for IDS to satisfy the requirement constraint of WSN without losing the security and reliability.

References

[1] Akyildiz,I.F.,Su,W.,Sankarasubramaniam,Y.,Cayirci,E.,2002.Wirelesssensor networks: asurvey. Comput Networks 38,393–422.

[2] Lopez J, Zhou J. Overview of wireless sensor network security. In: Wireless sensor network security. IOS Press, incorporated; May 2008. p. 1–21.

[3] Perrig A, et al. SPINS: security protocols for sensor networks .Presented at the 17th ACM international conference on mobile computing and networks, 2001.

[4] Xiao Y, Rayi VK, Sun B, Du X, Hu F, Galloway M. A survey of key management schemes in wireless sensor networks. Comput Commun 2007;P: 30-23,14–41.

[5] Jaiganesh, V., Mangayarkarasi, S., & Sumathi, P.(2013). Intrusion Detection Systems: A Survey and Analysis of Classification Techniques. International Journal of Advanced Research in Computer and communication Engineering ,Vol. 2, Issue 4, April 2013

[6] Anderson JP. Computer security threat monitoring and surveillance. Fort Washing- ton, Pennsylvania: James P Anderson Co; April 1980.

[7] Kuperman, Benjamin A. "CERIAS Tech Report 2004-26 A CATEGORIZATION OF COMPUTER SECURITY MONITORING SYSTEMS AND THE IMPACT ON THE DESIGN OF AUDIT SOURCES." (2004).

[8] Govindarajan, M. "Hybrid Intrusion Detection Using Ensemble of Classification Methods." International Journal of Computer Network & Information Security 6.2 (2014).

[9] Hu J. Host-based anomaly IDS. In: Springer handbook of information and com- munication security. Springer Verlag; 2010.

[10] K. Sharma and M. K. Ghose; Wireless Sensor Networks: An Overview on its Security Threats; IJCA, Special Issue on “Mobile Ad-hoc Networks” MANETs; CSE Department, SMIT, Sikkim, India; 2010.

[11] Y. Zhou, Y. Fang and Y. Zhang; Security Wireless Sensor Networks: A Survey; IEEE Communication Surveys; 2008.

[12] David Boyle, Thomas Newe. “Securing Wireless Sensor Networks: Security Architectures”, Journal of networks, Volume 3, No. 1, 2008.

[13] Rouba El Kaissi, Ayman Kayssi, Ali Chehab and Zaher Dawy, (2005)” DAWWSEN: A Defense Mechanism against Wormhole ttack In Wireless Sensor Network”,Proceedings of the Second International Conference on Innovations in Information Technology (IIT‟05).

[14] A. D. Wood and J. A. Stankovic,(2002) “Denial of service in sensor networks”,Computer, 35(10):54–62, 2002.

[15] Ioannis Krontiris, Tassos Dimitriou, Thanassis Giannetsos, and Marios Mpasoukos. Intrusion detection of sinkhole attacks in wireless sensor networks. In Miroslaw Kutylowski, Jacek Cichon, and Przemyslaw Kubiak, editors, ALGOSENSORS, volume 4837 of Lecture Notes in Computer Science, pages 150–161. Springer, 2007.

[16] Richard Heady, George Lugar, Mark Servilla, and Arthur Maccabe. The architecture of a network level intrusion detection system. Technical report, University of New Mexico, Albuquerque, NM, August 1990.

[17] S. Shin, T. Kwon, G.Y. Jo, Y. Park, H. Rhy, “An experimental study of hierarchical intrusion detection for wireless industrial sensor networks”, IEEE Trans. Ind. Informat., volume 6, number 4, pages 744-757, 2010.

[18] I. Krontiris, T. Dimitriou and F.C. Freiling, “Towards Intrusion Detection in Wireless Sensor Networks”, Proc. 13th European Wireless Conference, 2007

[19] I. Onat and A. Miri, “An Intrusion Detection System for Wireless Sensor Networks”, IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2005.

[20] S. Rajasegarar, C. Leckie, M. Palaniswami, J.C. Bezdek, “Distributed Anomaly Detection in Wireless Sensor Networks”, 10th IEEE Singapore International Conference on Communication systems, 2006.

[21] A. Patcha and J.M. Park, “An overview of anomaly detection techniques: Existing solutions and latest technological trends”, Elsevier J. Computer Networks, volume 51, number 12, pages 3448-3470, 2007.

[22] Rajasegarar S, et al. Anomaly detection in wireless sensor networks. IEEE Wireless Communications 2008;15:34–40.

[23] Y.EL Mourabit, A. Toumanari, H.Zougagh, “A Mobile Agent Approach for IDS in Mobile Ad Hoc Network”, International Journal of Computer Science Issues, Vol. 11, Issue 1, No 1, January 2014

[24] M. Burgess. Computer immunology. In LISA ’98: Proceedings of the 12th USENIX conference on System administration, pages 283–298, Berkeley, CA, USA, 1998. USENIX Association.

[25] B.E Boser, I.M. Guyon and V.N. Vapnik. A training algorithm for optimal margin classifiers. In COLT ’92:Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152, New York, NY, USA, 1992. ACM. ISBN 0-89791-497-X.

[26] J.Platt, “Fast training of support vector machine using sequential minimal optimization,” Advances in Kernel Methods: support vector machine, MIT Press, Cambridge, MA, 1998.

[27] H-S. Kim and S-D. Cha. 2004. Efficient masquerade detection using svm based on common command frequency in sliding windows.IEICE Transactions On Information And Systems Volume, E87-D, 2446–2452.

[28] S. Rajasegarar, C. Leckie, M. Palaniswami and J.C. Bezdek, “Quarter Sphere Based Distributed Anomaly Detection in Wireless Sensor Networks”, IEEE ICC ’07, Glasgow, U.K., June 2007.

[29] Bill C.P. Lau a, Eden W.M. Maa, Tommy W.S. Chow, “Probabilistic fault detector for Wireless Sensor Network”, Expert Systems with Applications 41 (2014) 3703–3711.

[30] Hastie, T., et al., The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer, 2005. 27(2): p. 83-85.

[31] Abebe Tesfahun, D. Lalitha Bhaskari, “Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction”, 2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies.

[32] http://kdd.ics.uci.edu/databases/kddcup99/ kddcup.

[33] Zdrayko Markov, Ingrid Russel, “An introduction to the WEKA data mining system”, ITICSE '06 Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education, Pages 367-368.

[34] I. H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[35] M. Xie et al. “Anomaly detection in wireless sensor networks: A survey”, Journal of Network and Computer Applications 34 (2011) 1302–1325.