Biomedical Research 2017; 28 (5): 2031-2035 *ISSN 0970-938X*
www.biomedres.info

An intrusion detection system approach using conditional random field for detecting attacks on web-based telemedicine system.

Bala Krishnan R[1*], Manikandan G[2], Rajesh Kumar N[1], Raajan NR[3], Sairam N[2]

[1]Department of Computer Science and Engineering, Srinivasa Ramanujan Centre, SASTRA University, India

[2]School of Computing, SASTRA University, India

[3]School of Electrical and Electronic Engineering, SASTRA University, India

Abstract

The attainment of security level in telemedicine communication systems might be a sensitive research area because of its content confidentiality. The proposed Intrusion Detection System (IDS) research work presents a model for detecting attempts of attacks to acquire unauthorised access to a web based telemedicine system. The process of attacks detection would be performed by applying Conditional Random Field (CRF) based approach because of its structured prediction feature. The proposed mechanism deals with structured prediction approach and hence it is suitable for identifying the attacks over the web based telemedicine application. From the experimental observations, the proposed scheme offers high efficiency with accuracy in terms of detecting the attacks (Probe, R2L, U2R and DoS). This model presents proficient results over the medical data communication in terms of efficient attack detection time, false alarm rates along with good classification accuracy.

Keywords: Intrusion detection system, Anomaly detection, Condition random field, Structured prediction, Pattern matching.

Accepted on September 21, 2016

Introduction

The growth of internet technologies and its usage leads the modern communication technology era. The growth of internet over the fields like Medicine, tele-consulting creates possibility for the occurrence of intrusive behaviours and it results to system failure. One of the popular web based telemedicine based consultation system for asthma disease is Arnasa [1], it is a decision support system supports for health professionals by holding the implicit knowledge from clinical practice guidelines. There is possibility for the occurrence of intrusive behaviours over this kind of Decision Support System (DSS) for data extracting and make the system failure. To safe guard this kind of DSS, an efficient intrusion detection and avoidance system needs to be incorporated in the tele-consulting web application for the detection of intrusive attacks and to improve the data classification accuracy of the system for obtaining a better Quality of Service (QoS). The mechanism IDS have been classed into two major categories: Misuse or signature-based detection and anomaly detection [2,3]. Misuse detection identifies the intrusive attacks by known intrusive signature patterns whereas the anomaly detection finds the abnormal request by checking it with patterns from Knowledge Base (KB). For the need of attack identification the proposed research plan introduces an efficient intrusion detection System for web-based telemedicine application using a structured prediction procedure called Conditional Random Field (CRF). In most of the existing intrusion detection systems, intrusions could be detected by applying machine learning and rule based data mining approaches. Some of the popular mining algorithms are Naive Bayesian [4], support vector machine [5], particle swarm optimization [6,7], genetic algorithms [8], neural networks [9,10] etc. The rule based approaches works well in terms of attacks detection but the need for the IDS over a web based telemedicine application expects attack detection along with classification accuracy. For attaining the classification accuracy the structured prediction scheme CRF would be deployed in the proposed work. The rest of the article is posed as follows. Section 2 supplies an overview of related works carried out in IDS with data mining approaches; Section 3 offers the proposed scheme. Experimental observations are stated in Section 4 and conclusion is stated in Section 5.

Related Works

The intrusion detection system has been implemented in various applications like: Media, Medicine, Telecommunication, and Tele-Consulting. The accuracy of the IDS application depends on the mining approach which is to be incorporated in the model. Popular approaches of IDS are machine learning, biological information, image [11], statistical and data mining [7]. An approach for efficient

parameters identification for attacks classification is stated by the authors Bala Krishnan et al. [4]. The authors Garcia Adeva et al., [11] deal with intrusion detection using text mining for web-based telemedicine systems. The growth of internet medium and its facilities enable the multimedia tools for online medical consultation. The proposed model is also designed to perform an efficient online medical consultation system with security. The data which is transformed to the application may contain intrusive contents that may leads to system failure. The attacks may extract the secret data about the patient from the database. At present, some tele-consultation applications are available on demand but they focus on the medical terminologies alone with less focus on the security aspects. The major drawback of the IDS designing is: Observing attacks and other security ravishments, which have not been precluded by elementary protection techniques [12]. An approach for identifying the intrusive attacks on biometric images is stated by the authors Bala Krishnan et al., [13]. A large amount work has been prepared to employ the characteristics of IDS and is stated by the authors Durst et al., [14].

Proposed System

The architecture of the proposed intrusion detection system application on web-application is stated in Figure 1. The intrusion detection module resides in the application server. The application server handles the bio-medical requests from various sources. The proposed IDS module contains the following four components: Knowledge Base (KB), Request analyser, attacks classifier, and training agent. The component Services handler is available inside the application server for handling the normal requests.

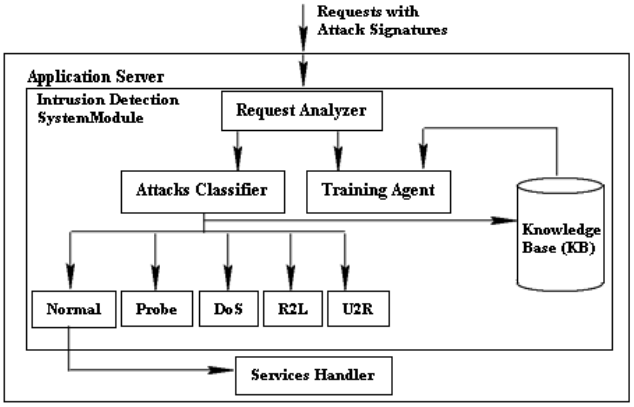

Figure 1. System architecture for proposed IDS.

The proposed model works by including the medical requests with KDD Cup'99 data set. The web-based telemedicine system application received requests or queries from various locations and the requests may hold intrusive contents. The intrusive contents are simulated using the KDD Cup'99 data set. The data set holds 42 attributes, which characterize the behaviour of network traffic. Among the 42 attributes, three attributes are represented using symbols and the rest are numeric data.

The proposed web-based telemedicine system application receives the client requests from various sources and the requests are handled by the 'Request Analyser' of the proposed intrusion detection system module, which is available inside the application server and it is the gateway for the requests from various sources. After receiving the request the analyser module analyses the request using the existing data set from KB and using the Conditional Random Field (CRF) procedure. The classified requests are again updated in the KB for future predictions and normal requests are submitted to the 'Services Handler' module for granting medical services or suggestions to the requested users.

CRF on IDS

A conditional random field is an undirected and unbiased graphical model and it could be effectively used to perform sequence labelling. The structured approach CRF is used to device the conditional distribution of variables (random). Let R_i be a random variables set with a data sequence and LS_i be the consequent label series, where i=1, … N. Let the Graph GH=(Vertex, Edges) and LS=LS_i U (Vertex). So that LS is indexed by the vertices of graph GH. Then, (R_i, LS_i) is a conditional random field, when trained on R_i, the variables LS_i abide by the feature of Markov with deference to the graph:

$$P (LS_i /R_i, LS_j, j \pm i) = P (LS_i/R_i, LS_j, i \sim j) \rightarrow (1)$$

Where, i ~ j states that the elements i and j are neighbours in the graph GH. It states that the CRF is a field which conditioned globally on LS_i. The CRFs are stated by the equation:

$$PCRF (LS_i/R_i) \alpha \exp (\beta_K f_K (e, y/e, R_i)) + \Sigma \Upsilon_K g_K (i, y/i, R_j) \rightarrow (2)$$

The element R_i is the sequence of input data for the classification, LS_i is a sequence of label and the characteristics f_k and g_k are picked out by the exploiter of this application. For example: The Boolean edge property f_K is true if the outcome R_i is 'tcp' which is devolved by the determination negotiator. The property g_k is true if the outcome R_i is "service of telnet" and element LS_i is confirmed as "attack". In this proposed scheme of classification, the work has been presented as an enhanced CRF based approach by extending the existing CRF mechanism [5] in which features for the attacks identification are selected randomly. Proposed IDS Application is trained at two levels: Phase I and Phase II. In Phase I, the proposed application is trained to detect the attacks. A numerical value has been assigned here for all features of the dataset and the proposed model fixes a threshold value to find the accurate the features for all the types of attacks. Identified characteristics are maintained in a list named: 'Attacks-List (AL)'. The IDS application in Phase I work with the AL data to detect an attack and it executes on the base of the cooperative involvement assessment of each character by utilizing the conventions. The training depends on the previous knowledge (dataset). The Phase II is the live implementation of the model; it detects the online attacks on the web-based telemedicine system Application and stores the attacks on the knowledgebase for

the future reference. The classified normal requests are then forwarded to the services handler of the application server to handle the client requests.

Procedure: IDS training/construction phase (Phase I)

Input: Dataset with all Parameters

Output: Parameters for attack classification and attack type

Step 1: Read all the input parameters of a request on Telemedicine system.

Step 2: Identify the parameters that are essential for the attack classification using CRF store it in the 'List'.

Step 3: Apply the CRF procedure Equations 1 and 2 on the list to identify the attack pattern.

Step 4: Store the obtained attack with pattern in Knowledge Base (KB).

Procedure: proposed CRF based classification mechanism

Input: KDD CUP'99 Dataset with telemedicine request, Knowledge Base (KB).

Output: Classified dataset with attack labels.

Step 1: Read the input.

Step 2: Split the telemedicine request and structured parameters from the request.

Step 3: Store the telemedicine request in the array 'Bio-Request' and rest of the parameters in an array 'IDS-CRF'.

Step 4: Extract the features from the 'IDS-CRF' for attacks and normal request classification.

Step 5: Apply the knowledge (trained CRF based classifier) from the KB on the processed testing dataset 'IDS-CRF'.

Step 6: Observe the outcome from the trained classifier.

Step 7: Fix the label for the dataset that is either attack or normal.

Step 8: Check the label with attack to identify the exact pattern of attack.

Step 9: If the attack is 'Normal', forward the request 'Bio-Request'' to services handler module with request structure else the request will be terminated.

The Proposed CRF based IDS on web-based telemedicine application works for identifying the four types of attacks with nature such as Probe, DoS etc., and a normal request will be forwarded to the services handler module. The services handler modules handles the request and send back the response for the bio-request and the outcomes of the IDS module that is either attack with label or normal request are regularly updated in the knowledge base.

Experimental Observations

The proposed web-based telemedicine system application has been implemented using J2EE platform using Apache Tomcat Server 7.0.65 and Java 1.7 for business logic implementation with Servlets and Java Server Pages. The database My-Sql 5.1.59 has been utilized as backend. The performance of the proposed algorithm is assessed on a Pentium 2.6 GHz system (Core 2 Duo) with RAM capacity of 4 GB on Windows 8 platform. The IDS module on the proposed web-based telemedicine system application have been designed to observe four cases of intrusive attacks, where Att_i is the number of attributes that have been considered for the classification from the dataset, Num states the numeral figure of training dataset, Tag denotes a numeral figure for labels and loop states the total iterations count. The complexity measure in terms of time for the proposed IDS application with the four attack types and it overall complexity are stated in the Table 1 [4].

Table 1. *Attacks type and its time complexity.*

S. No	Attack type	Time complexity
1	Probe	O ((Att_i) Tag^2 Num × Loop)
2	DoS	O ($(Att_2\text{-}Att_1)$ Tag^2 Num × Loop)
3	U2R	O ($(Att_3\text{-}Att_2)$ Tag^2 Num × Loop)
4	R2L	O ($(Att_4\text{-}Att_3)$ Tag^2 Num × Loop)
	Overall complexity	O (Att_4 Tag^2 Num × Loop)

As for the proposed system testing, 10 percent of the total records from KDD Cup'99 dataset is taken as training data and 10 percent of total records with adjusted labels has been taken as testing data for the experiments in the client server environment. The requests from various sources have been generated and the model has been tested with 2.15, 024 requests. The outcomes of the experiment are the attack classes, which are Normal, Probe, DoS, U2R and R2L.

The accuracy of the proposed IDS module on the web-based telemedicine application has been computed through the metrics such as precision, recall and F-measure, which affect the accuracy of the system. The parameters for the metrics computations are stated in the Table 2.

Table 2. *Matrix (Confusion) for the attacks rating.*

Attack type	Connection	
	Attack	**Normal**
Attack	True Positive-TP	False Negative-FN
Normal	False Positive-FP	True Negative-TN

The equations for the metrics computation are stated below:

Recall (RC)=TP/(TP+FN) → (3)

Precision (PN)=TP/(TP+FP) → (4)

F-Measure (FM)=$(1+\beta^2)$ × RC × PN/β × (RC+PN) → (5)

Where RC is the ratio between the relevant data and successfully detected data, PN is the ration of data conditioned for the detection and F-measure states the PN versus RC value in terms of fraction.

The detection rate of attacks and Rate of False Alarm (FAR) could be as computed using:

$DR=TP/(TP+FN) \rightarrow (6)$

$FAR=FP/(TN+FP) \rightarrow (7)$

The outcomes from the proposed IDS module with CRF mechanism in the web-based telemedicine application have been stated in the Table 3. It states the accuracy comparison of the proposed mechanism with existing in terms of precision, recall and F-measure values.

The proposed mechanism offers an efficient result in terms of attacks detection with high level of accuracy.

Table 3. *Accuracy metrics evolution with attacks.*

Attacks/Metrics	Probe	DoS	U2R	R2L
Existing RC [3]	97.8	97.05	62.3	27.08
Proposed RC	97.91	97.04	64.7	29.13
Existing PN [3]	88.1	99.98	55.07	94.7
Proposed PN	90.42	99.9	58.16	95.23
Existing FM [3]	92.7	98.5	58.1	42.0
Proposed FM	93.11	98.8	58.14	42.33

The accuracy of the proposed IDS module with CRF mechanism is compared with the existing Intrusive Attacks classification models using the algorithms such as decision tree, C 4.5 etc., and the observations show that the proposed model works well and offers better results in terms of its attacks detection rate.

The results are presented in the Table 4.

Table 4. *Accuracy evolution of classifiers.*

Classifiers	Attacks			
	Probe	R2L	DoS	U2R
C 4.5 (Decision Tree) [2]	80.8	97.0	1.80	4.60
Enhanced C4.5 [2]	81.5	97.12	6.24	12.57
MLP [2]	88.70	97.2	13.2	5.60
Proposed CRF based IDS	98.87	97.45	86.88	24.51

Conclusion

In this proposed research work a new intrusion detection system module have been presented with CRF mechanism in web-based telemedicine application, which improves the false positive intrusive attacks detection accurately.

Experimental results states that the proposed model works effectively and the outcomes are compared with the existing systems. Hence the proposed model is suitable for discovering the four major types of attacks over this kind of web-based telemedicine applications.

The proposed model could be planned to upgrade with temporal models on telemedicine solutions and it would the future direction of this research work.

References

1. Francisco JS, Juan MP, Iker UL, Juan JG, Diego LD. Towards a clinical practice guideline implementation for asthma treatment. Lecture Notes in Artificial Intelligence 2004; 3040: 587-596.
2. Shahbaa A, Karma M. Network intrusion detection based on hybrid intelligence system. AL Rafidain J Comp Sci Math 2012; 9: 81-98.
3. Sannasi G, Pandi V, Palanichamy Y, Arputharaj K. An intelligent CRF based feature selection for effective intrusion detection. Int Arab J Inform Technol 2016; 13: 271.
4. Bala Krishnan R. Efficient attributes identification practice on intrusion detection system dataset through prediction mechanisms. Far East J Electr Commun 2015; 1: 133-139.
5. Mukkamala S. Intrusion detection using neural networks and support vector machine. Proc IEEE Int Honolulu 2002.
6. Guangyou Y. A modified particle swarm optimizer algorithm. Proc Int Conf Electr Measur Instr 2007; 2: 675-679.
7. Shazzad K, Park J. Optimization of intrusion detection through fast hybrid feature selection. Int Arab J Inform Technol 2014; 4: 922-927.
8. Golovko V, Kochurko P. Intrusion recognition using neural networks. Proc Intel Data Acqu Adv Comp Sys Technol Appl IDAACS 2005; 2005: 108-111.
9. Hofmann A, Schmitz C, Sick B. Rule extraction from neural networks for intrusion detection in computer networks. Proc Int Conf Sys Man Cybern 2003; 2: 1259-1265.
10. Abdullah AM. Designing of intrusion detection system based on image block matching. Int J Comp Commun Eng 2013; 2: 605-607.
11. Garcia Adeva JJ. Intrusion detection using text mining in a web-based telemedicine system. Artif Intel 2005; 3809: 1009-1014.
12. Alsharafat W. Applying artificial neural network and extended classifier system for network intrusion detection. Int Arab J Inform Technol 2013; 10: 230-238.
13. Bala Krishnan R. An enhanced biometric based intrusion detection system for secure communication. Far East J Electr Commun 2016; 6: 121-131.
14. Durst R, Champion T, Witten B. Testing and evaluating computer intrusion detection systems. Commun ACM 1999; 42: 53-61.

An intrusion detection system approach using conditional random field for detecting attacks on web-based telemedicine system

*Correspondence to

Bala Krishnan R

Department of Computer Science and Engineering

Srinivasa Ramanujan Centre

SASTRA University

India