A Novel Intrusion Detection Algorithm for the Protection of Databases

PENG XU*

Security Engineering, The Third Research Institute Of Ministry Of Public Security, Shanghai Province, Xuhui Shanghai, 200030, China

***Abstract* — In order to improve the detection of intrusion and enhance the protection of computer databases a new intrusion detection algorithm is proposed in this paper. The evolved detection and reaction methods are used to construct our distributed and cooperative Statistical Anomaly-based Intrusion Detection and Reaction System (SAIDRS). The performance of SAIDRS is evaluated using simulation experiments. Different mobility patterns and attack scenarios are utilized in the simulated network. The evaluation results show that SAIDRS performs well in detecting and reacting to routing attacks with a noticeable improvement in the network throughput.**

Keywords - Intrusion detection technology; computer database; external force entry conditions

I. INTRODUCTION

In recent decades, the explosive growth of the Internet has resulted in an increasing number of people using the Internet in their daily life and business. However, risks from network attacks cause damage to social stability and economic development. As one of main techniques for protecting network security, intrusion detection technology can detect network attacks before they induce widespread damage and provide important basis for establishing defence strategy. With the continuous expansion of the network, a variety of new security vulnerabilities and network attacks are emerging, which put forward a higher requirement for the performance of intrusion detection system (IDS).Data mining is an intelligence analysis technique, which is able to discover the useful knowledge from amount of data.

Compared with using data mining technologies to the entire original data set, using a subset of the original data set selected by sample reduction method can reduce the cost of data mining and speed up the data mining process, and sometimes even able to achieve better result. In order to select a subset of high quality from the original data set, Li's [1] paper proposes a stratified sample method based on class's centroid. In this method, a new concept is introduced to measure the representative power of an instance with respect to its class in a given data set, and a strategy is proposed to divide the data set into subsets of equal size. A subset can be selected from the original training data set and will be used as training data to build intrusion detection model.

The prevention mechanisms are defined as a first line of defence to protect MANET from attacks by using several approaches, such as authentication, encryption, and digital signature. Generally, the approaches used in prevention solutions are based on cryptography methods. The cryptography can be used to ensure data integrity in transmission, hide valuable and secure data, and achieve authentication services. Cryptography-based security solutions include asymmetric and symmetric keys. The symmetric key cryptography, which is based on a pre-distributed shared key, requires the least computation and bandwidth power to establish and distribute the shared key. However, due to the absence of any centralized point in MANET, it is impossible to securely distribute and reconfigure the shared key. On the other hand, the asymmetric key cryptography uses two keys, (private and public keys) to encrypt, decrypt, authenticate, and verify messages. Thus, since the private keys never need to be transmitted to anyone, asymmetric key cryptography increases the security with less prior configuration. But it causes a considerable communication overhead and more computation power.

While the network brings convenience to people, its own fragility offers intrusion opportunities for hackers and malicious attackers. Along with the diversity and complexity of intrusion attack, high performance intrusion detection techniques are required, and so the study of on-line detection, adaptive detection and multiclass detection techniques becomes current hotspot. To improve the performance of multiclass intrusion detection system, this dissertation focuses on the study of multiclass intrusion detection methods against the characteristics of the easy classification, easy mixed, imbalanced and new unknown types of attacks, and proposes an adaptive multiclass intrusion detection ensemble model. To improve the overall detection accuracy and efficiency of intrusion detection system further, various ensemble structures of classifiers are studied in Chi's paper [2]. Combining the advantages of different detectors detecting different attack types, an intrusion detection ensemble model with the three levels of hybrid structures is proposed. The first level detector based on the principal direction divisive partitioning clustering detects the easy classification attacks. The second level detector based on the feature extraction of the weighted non-negative matrix decomposition and the projection pursuit direction divisive partitioning clustering detects the easy mixed and the imbalanced types of attacks. The third level detector based on the ART2neural network recognizes the new unknown types of attacks. This ensemble model develops every single detector's advantages, is able to detect the easy classification

attacks quickly, and improves the detection accuracy of the easy mixed and small class of attacks. It can detect new unknown types of attacks and learn their profiles adaptively. So the model in paper has a better overall performance [3-4].

II. KEY TECHNOLOGIES INVOLVED IN THE SYSTEM

The distributed and cooperative architecture used in our proposed IDS mainly relies on IDS-agent attached to each mobile node. The IDS-agent model (as shown in Figure 1) consists of four components, namely, data collection component, data process component, intrusion detection component, and response component. However, those components are responsible for the detection and reaction to both neighboring and remote intruders in MANET [5].

The main tasks of data processing module are: (1) obtaining direct observations about neighbors' activities (i.e., outgoing and incoming packets counts), (2) collecting delay time and throughput data about all monitored paths (i.e., paths existed in routing), and (3) gathering and filtering indirect security-related information about both neighboring and remote nodes. The input data for our proposed IDS is obtained from three main sources:

(1)the activities of neighboring nodes that include routing and packets forwarding processes, (2) characteristics of path that replay ping broadcasted by the IDS-agent, and (3) the security-related information regarding the behavior of both neighboring and remote node (i.e., global information provided by other IDS-agents). Since the existence of misbehaviors' causes a noticeable deviation in some of the path characteristics, the deviation in path features are utilized in this research and computed statistical so as to detect the suspected nodes (i.e., pre- classify nodes).

Nevertheless, the behavior of suspected nodes will further be verified based on weighted set of local and global data. However, by filtering indirect information based on their sources' behaviors, only information from trusted nodes are accepted and associated with local information, and thus, the cooperative anomaly detection is achieved.

After a specific period of monitoring the system activities; where the data is obtained and be ready for analysis, this module performs the following tasks: (1) prepares the obtained data from different sources (i.e., neighbour node behaviour, path behaviour, and non-neighbour node behaviour) in a way that can be used as input for the intrusion detection module, (2) shares the computed features of neighbours' activities by broadcasting them via routing protocol packets (i.e., utilizing the unicasted ZREQ). However, data preparing includes: (1) analysing path features and pre-classify all monitored nodes based on their paths classifications, (2) calculating the statistics (i.e., average and variation) and ratios of all selected behaviour features those reflecting the activities of neighbouring nodes, (3) filtering all the gathered global information by relying on their sources behaviour in the view of monitor (i.e., the host of IDS-agent), and (4) aggregating the accepted indirect information with direct information by computing a weighted average vector of each feature, and use the feature vectors to update the record of each monitored node in the monitoring table.

The main tasks of intrusion detection module are: (1) detecting anomalous data points in data set of each behaviour feature (i.e., the weighted-average vectors), and (2) making the decision regarding the behaviours of monitored node based on its accusation times during the previous monitoring periods. In statistic, there are many techniques to detect outliers (anomalies) in data. However in this research, we propose a threshold-based detection scheme to detect attacks by determining the statistical changes of the ratios of selected features. Due to the resource constraints in MANETs, we introduce simple statistical-based algorithms that can detect anomalous behaviors without wasting too much resource. Based on the proposed scheme, our anomaly detection algorithms are applied to all behavior metrics in order to detect the anomalous events of each monitored node. Meanwhile, the decision regarding the behavior of monitored node is made by comparing its accusation times with maximum allowed accusation. Where, the accusation record of a node is periodically updated based on the findings of anomaly detection algorithms. However, by using the associated data from different sources, each IDS-agent, based on its location, takes part in the final decision regarding node behavior [6-8].

The main task of response module is making the decision of proper response regarding the service requests based on the current security information. The pervious researches of IDS have taught us the half of fact that, a robust response system using an effective mechanism can improve the network security. However, without detecting the attack or identifying the attacker node, it is hard for IDS to make the proper response decision. Therefore, the response decision is the second part of our proposed IDS that presents an automatic detection and reaction of intrusions. In this research, we introduce a simple response mechanism aims to improve MANET security The mechanism is taking into account the resource constraints of MANET, however without interrupting the normal network traffic. The proposed response system is a combination of punishment, stimulation, and encouragement. By which, malicious (misbehaved) node would be enforced to behave well, fair (well-behaved) node would be stimulated to better or at least keep its (good) behaviour, and accused node would be encouraged for more participating and cooperating (i.e., routing and forwarding functions). Thus, IDS-agent can be able to take the appropriate actions wherever and whenever node behaviour is identified, based on the level of its mis behaviour (well, accuse, bad).

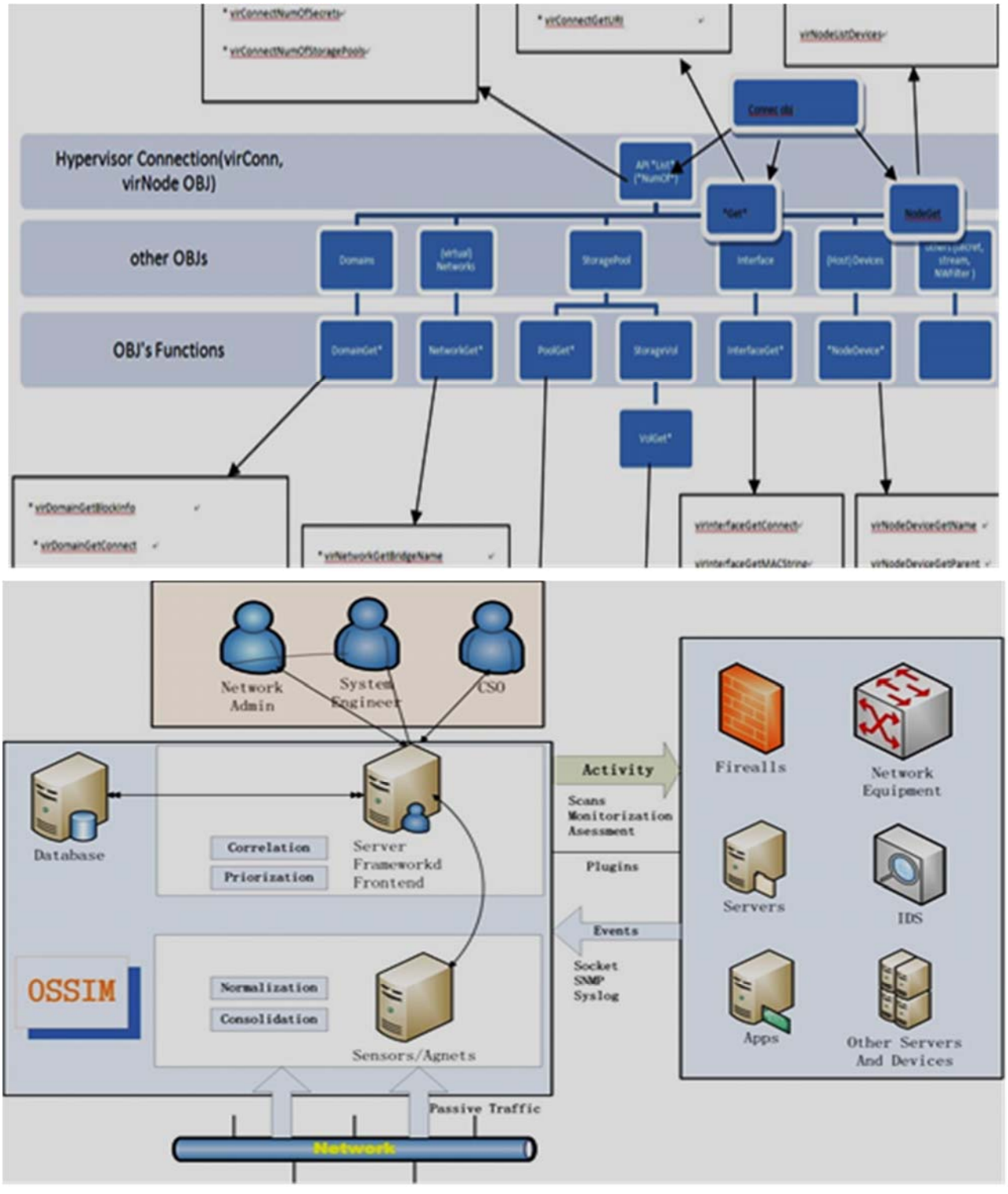

Figure 1. Fire hazard in concrete shield tunnel.

The linear differential equation can be expressed into the following simplified forms:

$$L(\nabla,\omega)f(x,\omega)=0,$$
$$L(\nabla,\omega)=T(\nabla)+\omega^2\rho\mathsf{J} \tag{1}$$

In which,

$$T(\nabla)=\begin{Vmatrix} T_{ik}(\nabla) & t_i(\nabla) \\ t_k^T(\nabla) & -\tau(\nabla)\end{Vmatrix},\ \mathsf{J}=\begin{Vmatrix}\delta_{ik} & 0\\ 0 & 0\end{Vmatrix},$$
$$f(x,\omega)=\begin{Vmatrix}u_k(x,\omega)\\ \varphi(x,\omega)\end{Vmatrix} \tag{2}$$
$$T_{ik}(\nabla)=\partial_j C_{ijkl}\partial_l\ ,\quad t_i(\nabla)=\partial_j e_{ijk}\partial_k\ ,$$
$$\tau(\nabla)=\partial_i\eta_{ik}\partial_k$$

Consider an infinite situation, we have the equation (3) in the following:

$$L^0=\begin{Vmatrix}C^0_{ijkl} & e^0_{kij}\\ e^{0T}_{ikl} & -\eta^0_{ik}\end{Vmatrix} \tag{3}$$

Consider the propagation, instead the equation (4) with the following form:

$$C(\mathrm{x})=C^0+C^1(\mathrm{x}),\quad e(\mathrm{x})=e^0+e^1(\mathrm{x}),$$
$$\eta(\mathrm{x})=\eta^0+\eta^1(\mathrm{x}),\quad \rho(\mathrm{x})=\rho_0+\rho_1(\mathrm{x}) \tag{4}$$

Then we have equation (5) to (8):

$$C^1=C-C^0,\quad e^1=e-e^0,$$
$$\eta^1=\eta-\eta^0,\quad \rho_1=\rho-\rho_0 \tag{5}$$

The containing inclusions can be simplified into the following integral equation set:

$$f(x,\omega)=f^0(x,\omega)+\int_V \mathcal{S}(x-x')(\mathrm{L}^1F(\mathrm{y}')$$
$$+\rho_1\omega^2\mathbf{g}(R)\mathrm{T}_1 f(\mathrm{y}')]S(\mathrm{y}')\mathrm{d}\mathrm{y}' \tag{6}$$

In view of the following relationship

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ik_3x_3'}dx_3'=\delta(k_3) \tag{7}$$

Equation (8) can be converted into the following form:

$$f(y,\omega)=f^0(y,\omega)+\int_s S(y-y',\omega)\mathrm{L}^1F(\mathrm{y}',\omega)\mathrm{d}\mathrm{y}'$$
$$+\rho_1\omega^2\int_s \mathbf{g}(y-y',\omega)\mathsf{J}f(\mathrm{y}',\omega)\mathrm{d}\mathrm{y}' \tag{8}$$

In which, S is cylinder cross section, $y=(x_1,x_2)$, and

$$\mathbf{g}(y-y',\omega)=\frac{1}{(2\pi)^2}\int_0^{\infty}\bar{k}\,d\bar{k}$$
$$\bar{k}=(k_1,k_2)$$
$$\int_0^{2\pi}\mathbf{g}(\bar{k},\omega)\exp(-i\mathrm{k}\square(\mathrm{y}-\mathrm{y}'))d\phi \tag{9}$$

The formula generates labels for each file block.

$$for(j=0;j\le n-1;j++); \tag{10}$$
$$\{W_j=r*(j+1);T_i$$
$$=[h(W_j)*m_j]^c \bmod N\}; \tag{11}$$
$$Output(T_0,T_2,...T_{n-1}); \tag{12}$$

And local fractional integral of $f(x)$ defined by Eq.9.

$$ {}_aI_b^{(\alpha)}f(t)=\frac{1}{\Gamma(1+\alpha)}\int_a^b f(t)(dt)^{\alpha}$$
$$=\frac{1}{\Gamma(1+\alpha)}\lim_{\Delta t\to 0}\sum_{j=0}^{j=N-1} f(t_{\mathrm{j}})(\Delta t_{\mathrm{j}})^{\alpha} \tag{13}$$

III. EXPERIMENT AND DATA ANALYSIS

Due to intrusion detection technologies are too new and there is no standard to compare against, therefore it is not easy to evaluate the performance of proposed IDS. However, many factors that affect the performance of MANET, such as the node mobility and the complex of routing mechanisms, are difficult to model mathematically. Therefore, an analytical evaluation of the proposed security solutions cannot enable us to investigate more MANT features. The simulation provides an alternative model of the analytical models, by which we can independently vary the experiment parameters to evaluate the performance of current or proposed solutions under different conditions. By using simulation, more realistic models with more environmental variables could be investigated and evaluated. Due to the complexity of the security solutions in MANETs, simulation plays an essential role in evaluating both the efficiency of current security solutions and the feasibility of proposed solutions.

In this research, in order to evaluate the effectiveness of our proposed solution in detecting and reacting to attacks aimed at routing protocols, we have built our network model to simulate the two routing attacks, which have been discussed earlier, by using Global Mobile Simulator (GloMoSim). GloMoSim is a clean and scalable simulation library designed for wired and wireless networks. This tool can provide parallel discrete-event simulation capability with integrated layered architecture. Based on the outcomes of an analytical evaluation of several possible simulators, G1oMoSim is chosen to implement our proposed anomaly-

based detection and reaction algorithms. That is, the current GloMoSim supports protocols for a purely wireless network. In addition to, GloMoSim as a set of library modules provides a sequential and parallel simulation of wireless networks. This feature is unlike other existing networks simulators, that G1oMoSim, based the parallel simulation, can scale large network with more than 106 nodes. Therefore, GloMoSim enables us to build scalable models for simulating MANETs. Moreover, as GloMoSim is built on top the PARSEC simulation environment and its libraries have been developed using the C-based simulation language (PARSEC). However, this language can be used to implement existing and under developing solutions using more realistic setting and scenarios.

In the simulation, our general methodology is to model a network with mobile nodes running our proposed IDS-agent under different scenarios so as to observe a fair performance evaluation. To achieve our goal, we consider an ad hoc network in a square region 1000m X 1000m, with 30 mobile nodes. Each of the nodes runs a separate copy of the proposed IDS. At the beginning of simulation, each node randomly locates on the square region. However, relying on the Random Waypoint mobility; the node moves from its current location to the new one in a speed uniformly chosen between 3 m/s and 10 m/s. Then, at the new location; the node pauses for a certain time (i.e., referred as pause time) before it moves to its next randomly chosen location. However, the pause time parameter is utilized to simply represent six mobility scenarios based on six different pause times, namely, 0, 30, 90, 150, 300, and 600 seconds. The performance of our proposed IDS is evaluated from two aspects; (1) the ability of detecting intrusive activities accurately, (2) the effectiveness of reacting to attacks in a way that mitigates the possible effects and achieves a better throughput. The ability of our proposed detection methods has been assessed in terms of the classic evaluation metrics; detection rate (DR) and false positive rate (FPR). As we have stated in our literature review, DR and FPR are the most common metrics those used in literature to identify and evaluate the proposed IDSs. However, these two classic metrics have no meaningful outcomes to measure the possible effects of attack. As our simulation goal includes the evaluation of our proposed response mechanism, our proposed IDS has been assessed in terms of the transmission rate (TR) to investigate the effectiveness of the proposed responses. However, to evaluate the performance in terms of the three mentioned metrics we simply follow this general methodology: for each scenario, we counted the number of detected attacks, the number of normal activities declared as attacks, the number of tracked attacks, the number of tracked normal activities, and the number of transmitted routing packets. Then, the outcomes of all simulated scenarios are accumulated and computed statistically to evaluate the performance metrics. Table 1 shows the experimental data set. Table 2 shows the operation time.

TABLE 1 EXPERIMENT DATA

Data set	Record number
Original	100480507
Probe	1408395
Target-probe	1408395
Training-probe	99072112
Training-result-probe	2213532384
Training-result-probe0	157521183
Predict-result-probe	96438515
Predict-result-probe0	1030747
Predict-result-probe	935794590
Predict-result-probe0	84286169
rmse	17

TABLE 2 EACH STEP OF THE PROCEDURE

Operation	Time(s)
Slopeone-Pjoin01	1598
Slopeone-Pjoin02	1787
Slopeone-Pjoin03	1676
Slopeone-Pjoin04	17636
Slopeone-Pjoin05	6626
Slopeone-Pjoin06	2417

IV. CONCLUSION

In order to improve the accuracy and sensibility of intrusion detection of computer database under external force entry conditions, a new intrusion detection algorithm is proposed in this paper. The evolved detection and reaction methods are used to construct our distributed and cooperative Statistical Anomaly-based Intrusion Detection and Reaction System (SAIDRS). In this research, in order to evaluate the effectiveness of our proposed solution in detecting and reacting to attacks aimed at routing protocols, we have built our network model to simulate the two routing attacks, which have been discussed earlier, by using Global Mobile Simulator. However, the evaluation results show that, SAIDRS performs well in detecting and reacting to routing attacks with a noticeable improving in the network throughput.

REFERENCES

[1] Lijun Li, Rentao Gu, Yuefeng Ji, Lin Bai, Zhitong Huang. All-optical OFDM network coding scheme for all-optical virtual private communication in PON. Optical Fiber Technology, 2013, pp. 13-27.

[2] Kaikai Chi, Yi-hua Zhu, Xiaohong Jiang, Xianzhong Tian. Practical throughput analysis for two-hop wireless network coding. Computer Networks, 2013, pp. 233-256.

[3] Kanagasabapathy A A, Franklin A A, "Murthy C S R. An Adaptive Channel Reconfiguration Algorithm for Multi-Channel Multi-Radio Wireless Mesh Networks". IEEE Transactions on Wireless Communications, 2010, 9(10):3064-3071.

[4] Xiong X, Yin F, Yue G, et al. "Research of Spatial-Temporal Multi-Channel Allocation Based on the Greedy Algorithm for Wirless Mesh Network". Telecommunications Science, 2012.

[5] Luiz Filipe M. Vieira, Mario Gerla, Archan Misra. Fundamental limits on end-to-end throughput of network coding in multi-rate and multicast wireless networks. Computer Networks, 2013, pp. 5717-5727.

[6] Tan X B, Xiong H, Wen H. “Research on the channel assignment problem for multicast in multi-channel multi-radio wireless mesh networks”. Proceedings of Chinese Control Conference, 2012:6556-6561.

[7] Kanagasabapathy A A, Franklin A A, “Murthy C S R. An Adaptive Channel Reconfiguration Algorithm for Multi-Channel Multi-Radio Wireless Mesh Networks”. IEEE Transactions on Wireless Communications, 2010, 9(10):3064-3071.

[8] Xiong X, Yin F, Yue G, et al. “Research of Spatial-Temporal Multi-Channel Allocation Based on the Greedy Algorithm for Wirless Mesh Network”. Telecommunications Science, 2012..

.