Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 2 – No.8, September 2015 – www.caeaccess.org

Performance Prediction Model for Network Security Risk Management

Akinyemi Bodunde Odunola
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile- Ife
Nigeria

Amoo Adekemi Olawumi
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile- Ife
Nigeria

Aderounmu Ganiyu Adesola
Department of Computer Science and Engineering, Obafemi Awolowo University, Ile- Ife
Nigeria

ABSTRACT

Network security risks compromises network asset, resources, rules, policy and guidelines. These compromises adversely affect the network performances by altering or tampering with the three set network security objectives i.e. Confidentiality, Integrity, and Availability (CIA) of network. The overall goal of network management is to maximize network performance. The proactive management of security risk of a network is thus a necessity requirement for effective and efficient performances of the network. This study aimed at developing a framework that automatically performs predictions on network security situations. This study presented a prediction model based on Bayesian Network applied to predetermine the effect of network security risk factors on the network Confidentiality, Integrity and Availability. The proposed model utilized the probability characteristics of Artificial Intelligence method to address the challenges being faced by network administrators in using objective metrics to measure their network security and justify the performance of their network, rather than relying on their instinct or experience. The proposed scheme measures the security risk quantitatively and predicts network performances using objectives metrics.

General Terms

Network Security, Modelling, Risk Management

Keywords

Prediction, Network Performance, Bayesian Network

1. INTRODUCTION

Network management is a critical issue in today's rapidly changing network environment. Data Communication Network consists of heterogeneous network, therefore the growing complexities of the networks requires the use of effective network management techniques. One of the essential components of network management is performance management. The network provides diverse applications to end users. Ensuring the workability of these applications as they are added to the network is challenging. Network performance information is needed at every stage in a network's evolution. These stages include requirements, architecture, design, implementation, routine maintenance, and upgrades. Network performance is a key determinant of an application's end-to-end performance and user experience. Network performance management has traditionally been thought of as a problem of providing guaranteed Quality of Service (QoS) capabilities in the network. It provides functions to evaluate and report upon the behavior and the effectiveness of the network or network elements. Its role is to gather and analyze statistical data for the purpose of monitoring and correcting the behavior and effectiveness of the network, network elements, or other equipment and to aid in planning, provisioning, maintenance and the measurement of quality.

The performance management of a network is closely related to the security management of a network. The continued growth in the number of network elements, end users, interfaces, protocols and vendor makes the network vulnerable to threats and attacks, which result to security risks. Network Security Risks are events that could result in the compromise of network asset, resources, rules, policy and guidelines. These compromises adversely affect the network performances by altering or tampering with the three set network security objectives i.e. Confidentiality, Integrity, and Availability (CIA) of network [1]. Network Security Risk is the potential impact that a threat can have on the Confidentiality, Integrity, and Availability of network systems resources and services by exploiting a vulnerability of the networks, thus, inefficiencies in the performance of the network are inevitable.

This study aimed at developing a framework that automatically performs predictions on network security situations. The study was motivated by the demands of the knowledge of network security risk management by emerging threats, vulnerabilities and risks in the networks. The task is challenging due to the diversity of the Network accompanied with a dynamic environment due to different kinds of network elements that demand comprehensive performance management.

The existing security solutions are very complex and costly. What is rapidly needed is a flexible, adaptable and affordable security solution, which provides greater autonomy. Therefore, it is necessary to review the way security system architectures are designed by investigating new technologies that could help make easier and cost-effective new solution [2]. The management of security risk of a network is therefore a necessity requirement for effective and efficient performances of the network.

Hence, there is a need for tools to enable early detection of network security risk problems which in turn should quickly alert the network administrator of the problem area. As a result, this study discussed the significance of a flexible decision support system for network security managers deciding between interventions, to ascertain that network services are delivered at the right time, available at the right place, present in the right shape, satisfying quality requirements and obtained at the lowest possible costs.

The rest of this paper is arranged as follows: Section 2 discusses the related works while Section 3 described the modelling process while Section 4 described the prediction algorithm and the conclusions are discussed in Section 5.

2. RELATED WORKS

Application of Artificial Intelligence (AI) tools to risk management has been utilized by some authors to manage uncertainties and security risks in a network. A study presented in [3] deals with uncertainty in software project management. The authors make use of combination of Bayesian network and knowledge engineering method of Artificial Intelligence to analyze risk. The outcome of the study affirmed that one approach of risk assessment is the application of Bayesian framework. In [4], also Bayesian Network was applied in modelling students' behavior by evaluating Bayesian networks precision for detecting the students learning styles. The Bayesian network was used to model different aspects of a student behavior while he/she works with the system. Then, it infers his/her learning styles according to the modeled behaviors. The result showed that Bayesian Network can be used to detect students learning styles. Bayesian Network was applied as an approach to traffic flow forecasting, specifically in the management of Urban Traffic Control Systems (UTCSs) and freeway systems [5]. The authors presented a new approach based on Bayesian networks to predict the traffic flow of the object link, even in case of incomplete data. It was shown that a Bayesian Network can be used to predict future events even where there is no historical data. Also, Bayesian network was applied in [6] to managing ecological assets. The author worked on parameterization and evaluation of a Bayesian Network for use in an ecological risk assessment. The work is based on combining expert and data-based estimations. It was shown that Bayesian Network can combine both subjective and objective data. In cyberspace environment, Bayesian Networks was used by [7] for cyber security analysis in order to capture the uncertain aspects in cyber security. The cyber security uncertainty was modelled using Bayesian Networks. The presented work showed that Bayesian Network is a modeling approach that correctly captures uncertainties.

Linear Regression is another Artificial Intelligence tool employed in [8]. A prediction model of network security situation based on Regression Analysis was presented. Linear regression was proposed as a method for network security situation evaluation. A prototype system was designed for data collection and regression fitting. The study shows that Regression Analysis complexity rate is low and less time-consuming. The regression prediction model reflects the physical network's security situation in a certain range of threshold value. The weakness of this method is that it can only work on small dataset and lack of scalability.

Graphical-based method is another Artificial Intelligence tool employed in [9]. The problem of security risk assessment and mitigation was addressed by proposing a dynamic security risk management using Bayesian Attack Graphs (BAG). The Bayesian Attack Graphs (BAGs) is used to model vulnerability exploitations in a test network. It was shown that the attack graphs-based risk management framework using Bayesian networks enables a system administrator to quantify the chances of network compromise at various levels and also help in risk mitigation procedure by identifying the most critical and probable attack path in the network. Conversely, the attack graphs can get complex as the network attacks sequences increases i.e. lack of scalability. It is also a scenario-based approach.

More recently, in [1], the theoretical background of the performance prediction model for data communication network security risk was revealed. The study presented the procedures that support dynamic decision-support model that will predetermine the impact of network security risk on the selected network domain given the causal- effect model. The study provided a system that will monitor and report the security status or posture of a network to enhance network performance and facilitate efficient quality of services.

In this paper, attempt will be made to implement the performance predictive model for managing security risks in a Data Communication Network. The proposed model will utilize the probability characteristics of Artificial Intelligence method known as Bayesians Network to address the challenges being faced by network administrators in using objective metrics to measure their network security and justify the performance of their network, rather than relying on their instinct or experience.

3. THE PREDICTION MODEL

The requirements and methodological issues to build a prediction model for network performances in the face of security risks was presented in [1]

According to [1], an adaptive network performance prediction was formulated as a Bayesian Network (BN)-based problem. A Bayesian network was employed as a tool to represent uncertain, ambiguous or incomplete knowledge of the client-server model of a data communication network domain. Bayesian networks make use of probability theory to represent the uncertain knowledge. Bayesian networks are probabilistic networks derived from Bayes theorem based on the Bayesian theories, which allows the inference of a future event based on prior evidence.

Applying Bayes' theorem:

$$P(F|E) = \frac{P(E|F)\,P(F)}{P(E)} \qquad (1)$$

Where:

F - represents Risk Causal Factor

E - represents Risk Effect

P(F) - represents Prior Probability of Risk Causal Factor (F) i.e. the Unconditional Probability

P(E|F) -represents the Conditional Probability of Risk Effect (E) given the Risk Causal Factor(F)

P(E) - represents Marginal (unconditional probability) of the Risk Effect (E) also called the normalizing constant or prior predictive distribution i.e. the probability that risk effect occurs when there is no specific information about the event or factor that influence it.

P(F |E)- represents Posterior probability of the Risk Causal Factor (F) given the evidences of Risk Effect (E).

As stated in [1], the prediction model will use Bayes' law to find the probability of the effect of a security risk on network performances in terms of its Confidentiality, Integrity and Availability given that one of the possible causes the risk has occurred. Thus, Equation (1) was re-written as:

$$P(F_i|E) = \frac{P(E|F_i)P(F_i)}{\sum_{j=1}^{n} P(E|F_j)P(F_j)}$$

$$i = 1,2,\cdots,n \qquad j = 1,2,\cdots,n \qquad (2)$$

This require formulation of Bayesian network model which consists of two parts; namely

i. A qualitative graphical structure of the relationships in the model; and

ii. A quantitative structure represented by the probability distributions that are indicated by the graph.

In building a Bayesian network, the procedures shown in Figure 1 for constructing Bayesian network structures [10] was employed in this study. The three important steps that were taken in this study are discussed as follows:

(i) Choose a set of variables that describes the application domain

(ii) Choose an ordering for the variables.

(iii) Start with the empty network and add variables to the network one by one according to the ordering.

(iv) To add the *i-th* variable X_i,

i. Determine a subset $pa(X_i)$ of variables already in the network $(X_1,...,X_{i-1})$ such that

$$P\ (X_i|X_1,\ ...,\ Xi_{-1})\ = P(X_i|pa(X_i))$$

(Domain Knowledge is needed here.)

ii. Draw an arc from each variable in $pa(X_i)$ to X_1.

Figure 1: Procedures for constructing Bayesian network structures

3.1 Constructing the graphical representation

Constructing Bayesian Network graphical representation decides what the variables of interest are, which will become the nodes in the BN, The network structure was obtained based on the conceptual causality model or the domain knowledge presented by [1]. To generate this Bayesian Networks graphical representation, the method employed in this study was firstly to generate a random Directed Acyclic Graph (DAG), and then to generate the conditional probability distributions for the graph.

3.1.1 Generating Random Directed Acyclic Graph (DAG)

The pseudocode in Figure 2 returns a graphical model which is directed acyclic graph where nodes represent variables and directed arcs (arrows) represent the conditional probability relationships assumed in the model. The variables and the arcs between the variables in the causality model presented in [1] are the inputs to the graph.

3.1.2 Quantification of the Network

Quantification of the network is equivalent to the parameterisation of the model. It involves assigning states and probabilities to each variable on the network. The states for each node represent the potential values or conditions that the node can assume. This study utilizes a probability distribution as the states type for each node based on the on the level of uncertainty inherent in network security risks. The discrete nodes contain probability distribution over the states of the variables. These probabilities were encoded during the quantification of the network. Since, the network security risk is considered as a function of the probability/likelihood of a threat materializing through vulnerability, and the impact of that event. Thus, the probability distribution of the likelihood of occurrence that a network system is intentionally or unintentionally exploited resulting in loss of Confidentiality, Integrity, or Availability is measured using the scoring system provided in Figure 3

```
Procedure GenerateMulti-ConnectedDAGs
  {This procedure create a connected Directed Acyclic Graph with n
    nodes}
  Set n ←NoOfNodes
  Set N ←1
  CreateGraph(n)
  Repeat
          Generate uniformly a pair of distinct nodes x and y;
           For each x, y Î E do
           If flag [x,y] = false Then  {If the arc does not exist in the
          actual graph, add the arc, provided that the underlying
          graph remains acyclic};
           CreateGraph(n)
              N=N+ 1
  Until N = n
  Return the current graph after N iterations
  End GenerateMulti-ConnectedDAGs

Procedure CreateGraph(n)
  {This procedure create a network with n nodes, where all nodes
          have just one parent, except the first node that does not
          have any parent i.e. Graph G=(V, E), with E={{x,y}: x∈V
          and y∈V and x≠y} }
  Set G← {V, E}
  Set E ←{ Ø }
  Set V ←1
  Repeat
          Generate uniformly a pair of distinct nodes x and y;
          for each x, y Î V do
            set E ¬ E  U {x,y}
          flag [x,y] ¬ true
  Until    |E| = |V|*(|V|-1)/2
  End CreateGraph(n)
```

Figure 2: Directed acyclic graph specification

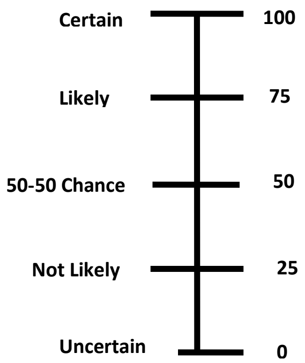

Figure 3: Scoring rules for probability distributions

As depicted in Figure 3, to determine the probability distributions of the likelihood of occurrence of the network security risks, the scoring functions are described as follows:

i. CERTAIN (100%) is defined as having a definite chance of occurrence

ii. LIKELY (75%) is defined as having a significant chance of occurrence

iii. 50-50 CHANCE (50%) is defined as equally probable chance of occurrence

iv. NOT LIKELY (25%) is defined as a modest or insignificant chance of occurrence

v. UNCERTAIN (0%) is defined as having a indefinite chance of occurrence

3.2 Specifying the numerical relationships among the variables

This section discusses how the prior, conditional and posterior probabilities distributions of the network security risk were determined:

3.2.1 Determining the Prior Probabilities Distributions

The probability of a security risk is a numeric value that states the likelihood of the specific security risk effect to occur, in a given period of time. Prior probabilities are the original probabilities of the security risk, which will be updated with new information to create posterior probabilities. Prior knowledge is summarized in a density *P(F)* (Equation 1), which represents prior probability of Risk Causal Factor (F). This function models a state of knowledge about the unknown true value of Risk Causal Factor (F). Bayesian analysis involves probing the data with various different prior beliefs about these parameters. In this study, the approach to determining the probability that a specific security risk effect will occur is subjective; expert knowledge is used as the sole source of information about the prior probability of security risks, since no statistical or historical data is available.

3.2.1.1 Eliciting priors from the experts

The obtained Bayesian Network knowledge structure information was collected by eliciting experts' opinion using a knowledge gathering framework based on questionnaire. This entails acquisition and integration of multiple experts' knowledge weighting the individual competence properties and the data knowledge. The estimation of a prior expectation of the probability of a node associated with each state was elicited from experts i.e. probability distributions were elicited from experts. The uncertainties associated with each relationship are quantified in the probability distribution.

Experts were asked to give and review their opinion on the rate of occurrence and magnitude of certain defined incidence within the causality network presented in [1]. Experts were asked to give their judgments on a response scale or give a numerical estimate based on the scoring rules (Figure 3) on the question sheets only for sections where they were confident of the response. The questionnaire is treated offline in this paper.

3.2.1.2 Combining elicited probabilities from the experts

The role of experts in this study is important because their judgments provide valuable information as regards the subject matter, particularly in view of the limited availability of ‘‘hard data’’ regarding many important uncertainties in security risk analysis. The motivation for the use of multiple experts is simply the desire to obtain as much information as possible.

Combining these experts’ probability distributions summarizes the accumulated information and provides sufficient information about the system. Thus, the aggregation of information across information sources is thus called the formation of a consensus of experts.

The mathematical method for combining experts’ probability distributions employed in this study is based on axiomatic approaches as presented in [11]. The method employed is called Linear Opinion Pooling (LOP) or Linear Averaging, which is calculated as a weighted linear combination of the experts’ probabilities. Linear Opinion pool is a method of combining a number of different opinions about some unknown quantity, θ to generate a single pooled opinion about θ. It is presented as follows:

$$p(\theta) = \sum_{i=1}^{n} w_i p_i\ (\theta) \qquad (3)$$

Assuming that for every $i \in \{1, \cdots, n\}$

$$0 \leq w_i \leq 1\ ; \quad \text{and}$$

$$\sum_{i=1}^{n} w_i = 1 \qquad (4)$$

Where:

n represents the number of experts

$p_i(\theta)$ represents expert i’s probability distribution for unknown θ,

p(θ)represents the combined probability distribution,

w_i represents the non-negative weights of the experts

The weight is used to represent the relative quality of the different experts, thus the determination of the weights is a subjective matter. The attribute of the experts used in this study is the number of years of experience as a network administrator in a networked environment as described in Table 1. The weights of the experts must be non-negative and sum to one.

Table 1: Allocation of weights to experts' opinion

No of years of experience	Assigned weight
10 and above	1.0
9	0.9
8	0.8
7	0.7
6	0.6
5	0.5
4	0.4
3	0.3
2	0.2
1	0.1

3.2.2 Determining the Conditional Probabilities

This section seeks to determine the conditional probability of security risks, *P(E|F)* in Equation (1), which is the probability that one of the possible Risk Effect (E) will occur, given the evidences of a Risk Causal Factor (F) or on the condition that Risk Causal Factor (F) has already occurred. The occurrence of any evidences of Risk Causal Factor (F) will always change the probability of Risk Effect (E). Thus, the two events are related and are called “dependent events”.

In this study, sampling-based sensitivity analysis was used to determine how sensitive these conditional probabilities to small changes in the parameters and/or evidence values. The approach to sensitivity analysis was to define reasonable ranges for each of the hypothesis or the parameter value i.e. the Risk Causal Factor (F) vary each parameter from its lowest to highest reasonable value while holding the other variables, Risk Effect (E), fixed, and examine the resultant changes in the target value. In short, Sampling-based sensitivity analysis was used to determine the likelihood (L) of the probability of the data (Risk Effect (E)) given the hypothesis or parameter value (Risk Causal Factor (F)) i.e.

$$L = P(data|hypothesis) = P(E|F)$$

Where:

$$P(F) > 0 \qquad (5)$$

Sampling from a probability distribution P(X) is a process of generating complete instantiations $X_1,\ldots,X_n$ of variables *X*. Thus, the sampling-based sensitivity analysis technique employed to estimate the likelihood is using Monte Carlo algorithms based on Markov chain simulation i.e. Markov Chain Monte Carlo (MCMC) simulation

From the chain rule of probability:

$$P(X_1, X_2, \cdots, X_n) = P(X_1)P(X_2|X_1)\cdots P(X_n|X_1, \cdots, X_{n-1}) \qquad (6)$$

$$= \prod_{i=1}^{n} P(X_i|X_1, \cdots, X_{i-1}) \qquad (7)$$

Thus,

$$P(X) = P(X_n, X_{n-1}, \cdots, X_1) \qquad (8)$$

$$P(X) = P(X_n|X_{n-1}, \cdots, X_1)P(X_{n-1}|X_{n-2}, \cdots, X_1)\cdots P(X_1) \qquad (9)$$

Thus, the markov chain model for this study is written as:

$$P(X) = P(X_n|X_{n-1})P(X_{n-1}|X_{n-2})\cdots P(X_2|X_1)P(X_1) \qquad (10)$$

$$P(X) = P(X_1)\prod_{i=2}^{n} P(X_i|X_{i-1}) \qquad (11)$$

Thus, the Markov-Chain Monte Carlo (MCMC) numerical simulation was performed by sampling the likelihood (L) function of the probability of the data (Risk Effect (*E*)) given the hypothesis or parameter value (Risk Causal Factor (*F*)) at discrete points in the multi-dimensional Bayesian network. The algorithms sample from Markov Chains which converge to the required probability distribution and hence give a simulation from this distribution i.e. from the MCMC points, probability distribution of parameters are obtained to get the best fit value of the parameters.

The MCMC algorithm employed for the sampling is the Random Walk Metropolis algorithm presented in Figure 4. This algorithm constructs a Markov Chain by choosing a random initial starting point in parameter space, and computes the target density by iteratively generating a step in parameter space from a proposal distribution and a new trial point for the chain by proposing small probabilistic symmetric “jumps” (the chain moves from one value to another in one step) centered on the current state of the chain. It then computes the target density at the new point, and accepts it or not according to some specified probability. If the point is not accepted, the previous point is repeated in the chain.

```
AIM: Sampling a probability distribution function P(x) i.e. the target density
Let f(x) be a probability distribution that is proportional to the desired distribution P(x)
        i.e.   $f(x)= P(x)$
Initialization:
    i.   Set $i=0$
    ii.  Choose an arbitrary point $x_0$ to be the first sample
    iii. Choose an arbitrary probability density $Q(x|y)$ which suggests a candidate for the next sample value x, given the previous sample value y.
         Where:
                 $Q(x|y)=Q(y|x)$

Compute the target density:
 For each iteration t: (Generate the sequence by repeating the following cycle, with $x_n$ being the previously selected point at each iteration)

    iv.  Select a new trial point $x^*$ for the next sampling, according to the symmetric proposal $Q(x^*|x_n)$
    v.   Calculate the acceptance ratio (to decide whether to accept or reject the candidate)

              $\alpha = \frac{f(x^*)}{f(x_n)}$

         Since,  $f(x)= P(x)$

              $\alpha = \frac{f(x^*)}{f(x_n)} = \frac{P(x^*)}{P(x_n)}$

    vi.  IF $\alpha \geq 1$, ( the candidate is more likely than $x_t$)
             Then  $x_{n+1}=x^*$ (Accept the candidate)
             ELSE  $x_{n+1}=x_n$ (Reject the candidate)
    vii.     $i=i+1$

Note: Note that the acceptance ratio $\alpha$ indicates how probable the new proposed sample is with respect to the current sample, according to the distribution P(x).
```

Figure 4: Random walk metropolis algorithm for sampling probability distributions

3.2.3 Determining the Posterior Distributions

The posterior probability P(F|E) represents posterior probability of the Risk Causal Factor (F) given the evidences of Risk Effect (E) i.e. it measures the probability of a security risk causal factor given information about the effect.

Given the likelihood that the Risk Effect (E) has been observed as a function of the unknown Risk Causal Factor (F) i.e. P(E|F) and original probability or the probability without further information i.e. the prior probability P(F), posterior probability updates belief and gives the probability of the Risk Causal Factor (F) that may explain the observed Risk Effect (E) through the Bayes' Theorem in Equation (1).

3.3 Making new evidence-based inferences or decisions

This section aims at providing an automated probabilistic reasoning of the likelihood of the security risk evidences in a network environment based on the BN model in Equation (1) i.e. using the BN to reason about the network domain. After observing the value of some variables, this involves conditioning upon the new information via a "flow of information" through the network. The process of conditioning is called probability inference or propagation also belief updating. This is performed by computing the posterior probability distribution for a set of query nodes, given values for some evidence (or observation) nodes. The reasoning considered in this study is predictive or causal reasoning (finding effects i.e. $F \rightarrow E$) which reason from new information about causes to new beliefs about effects,

The algorithm employed in this study is a class of approximate inference algorithm known as stochastic simulation algorithms. This class of inference is based on sampling, thus Markov Chain Monte Carlo (MCMC) algorithm specified in Figure 4 was utilized to make probabilistic inferences on the posterior probability distributions. The posterior distribution captures all information inferred from the data about the parameters. This posterior is then used to measure the security strength of a network in order to make optimal decisions or predictions about the network.

4. PREDICTION MODEL ALGORITHM

Prediction of the network performances in the face of security risks is based on the probabilistic approach of decision making under uncertainty. Thus, the procedure that a network administrator will utilized is as described in [1]. The formal framework is as presented in Figure 5.

5. SUMMARY

One of the most effective tools to manage and represent uncertainty has been the use of probabilistic graphical models. BN model have been proved suitable to model a large number of systems of this distinct nature. Thus, in this paper, BN model was utilized to manage the security risk in order to minimize the effects of uncertainties and information incompleteness in the network domain.

This study has demonstrated the use of the proposed model to predict the impact of security risks on the performances of the network so as to assess the future capacity needs and associated recommendations for performance monitoring and analysis of the network system. Thus, this study aimed at building a decision support system that provides network administrators with a tool to manage network security was realized.

6. CONCLUSION

This study is based on the conceptual model and the specifications formulated in [1]. It established a formal model with a level of detail sufficient to enable realistic predictions of operational network behavior and portray security measurements accurately.

The BN prediction model was built using information obtained from experts' knowledge elicitation. Bayesian probability updating ensures that the model is not static, but quickly adapts to new input and incorporates it with prior expert opinion in a mathematically tractable manner. The BN model has been shown to be easily adaptable to incorporate new input. This research is still ongoing. There is a need to establish how the model will work in a real life

7. REFERENCES

[1] Akinyemi Bodunde Odunola, Amoo Adekemi Olawumi and Olajubu Emmanuel Ajayi 2014. "An Adaptive Decision-Support Model for Data Communication Network Security Risk Management". International Journal of Computer Applications 106(8):1-7.

[2] Boudaoud, K., Boutaba, R. and Guessoum Z. 2000. Network Security management with Intelligent Agents. Network Operations and Management Symposium (NOMS 2000). 579-592.

[3] Paokanta, P. and Harnpornchai, N. 2009. Construction of Bayesian Networks for Risk Assessment of Software Project by Knowledge Engineering. 3rd International Conference on Software, Knowledge, Information Management and Applications. 154-158.

[4] García, P., Amandi, A., Schiaffino, S. and Campo, M. 2007. Evaluating Bayesian Networks' Precision for Detecting Students' Learning Styles. "Computers & Education". 49:794–808.

[5] Sun, S., Zhang, C. and Yu, G. 2006. A Bayesian Network Approach to Traffic Flow Forecasting. "IEEE Transactions on Intelligent Transportation Systems". 7(1):124-132.

[6] Pollino C.A, Woodberry O., Nicholson A., Korb K. Hart B.T. 2007. Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. Environmental Modelling & Software. 22(8): 1140–1152

[7] Xie, P., Li, J. H., Ou, X., Liu, P. and Levy, R. 2010. Using Bayesian Networks for Cyber Security Analysis. In Proceedings of the 40th IEEE/IFIP International Conference on Dependable Systems and Networks. 211-220.

[8] Wei-wei, X. and Hai-feng, W. 2010. Prediction Model of Network Security Situation Based on Regression Analysis. In proceeding of: Proceedings of the IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS). 616-619.

[9] Poolsappasit, N., Dewri, R. and Ray, I. 2012. Dynamic Security Risk Management Using Bayesian Attack Graphs. IEEE Transactions on Dependable and Secure Computing, 9(1):61-74.

[10] Pearl, J. 1999. Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference. Morgan Kaufman.

[11] Clemen, R. T. and Winkler R. L. 1999. Combining Probability Distributions from Experts in Risk Analysis. Risk Analysis, 19(2):187-203.

8. APPENDIX

Aim: Predicting network performances in terms of its Confidentiality(C), Integrity (I) and Availability (A), based on the evidences of security risks in a network domain

STEP 1- Initialization:

i. Choose a Network domain structure

ii. Construct the cause- effect model of the network domain

iii. Construct a Bayesian Network graphical representation of the domain {*The system is presented as a graph, and its structure function is established*}

iv. Given a dataset *(D_s)* of the network domain that contains information about security risks samples from n experts:

$$D_S = \{P(F_i): P(E_i(t))\}_{i=i,\cdots,n}$$

Where: F_i represents risks causal factors,
E_i represents risk effects

STEP 2- Input: *{quantification of the network i.e. assigning states and conditional probabilities for each variable on the network}*

i. The primary data are the unconditional probability distributions from the dataset:

$$F_i = \{P(F_i)\}_w \qquad w = 1,\cdots,k$$

{*This describes the likelihood of occurrence of security risks at t>=0, with marker w, weighted variable which represent the relative quality of the experts*}

Assumption:

a. The unconditional probability for the parent, Network Security Risk (NSR) must be one (1) i.e. the event of interest has occurred before time t

b. The unconditional probabilities for the three dependent factors used to define the Network Security Risk (NSR) i.e. Confidentiality Problems (C), Integrity problems (I), and Availability Problems (A) is assumed to have equal chance on Network Security Risk (NSR).

$$P(C) = P(I) = P(A) = \frac{1}{3}$$

STEP 3 - Compute the Conditional probabilities distributions:

i. Sampling-based sensitivity analysis - probability sampling *{to determine how sensitive these conditional probabilities are to small changes in the parameters and/or evidence values?}*

ii. Markov Chain Monte Carlo (MCMC) simulation:

a. Determine a probability function

b. Determine the random number generator, the source for selecting random numbers that are distributed uniformly on the proper unit interval

c. Determine a sampling rule for selecting samples for the model given a unit interval of random numbers

d. Record a count successes and failures

STEP 4 - Compute the Posterior probabilities distributions:

$$P(F_i|E) = \frac{P(E|F_i)P(F_i)}{\sum_{j=1}^{n} P(E|F_j)P(F_j)}$$

$$i = 1,2,\cdots,n \qquad j = 1,2,\cdots,n$$

STEP 5 - Output

Joint probability distributions: Risk impacts

i. Risk prediction modelling strategies for the response at time t are denoted *P(R)*. The dataset Ds are used to estimate model internal parameters.

ii. This yields a trained prediction model**:**

$$P(R)\{P(F_i)|D_S\} = \{P(F_i|E)\}_t$$

{This *assigns to security risk, i, a probability for the status at time t}*

$$P(R)\{P(F_i)|D_S\} = \{P(F_i|E)\}_t * P(F_i)$$

*{This probability is interpreted as a prediction for the unknown status of the network***}.**

Figure 5: The network performance prediction formal framework