Information Security Risk Assessment by Using Bayesian Learning Technique

Farhad Foroughi*

***Abstract*— The organisations need an information security risk management to evaluate asset's values and related risks. The risk management is usually a human activity which includes risk assessment, strategy development and risk mitigation by using managerial resources. The significant part of risk assessment which identifies threats and vulnerabilities, is very relevant to the past incidents, their likelihood and impacts. The problem is the risk identification and evaluation of new assets according to their properties, current security controls and consequences of before incidents. According to this problem, a system that could assist experts or works on behalf of them to assess the risks during the normal working processes is required. The system should be reactive and autonomous because it is needed to respond immediately and independently of events. An intelligent software agent is the best solution for this aim. It learns risk of past experiences regarding to risk factors and asset's properties, and predicts the probability of new risk for a new instance. This article will describe an intelligent system which is based on Bayesian Learning Technique for Information Security Risk Assessment.**

***Index Terms*—Bayesian Belief Network, Bayesian Learning, Information Security, Intelligent Agent, Risk Assessment.**

I. Introduction

All organisations and businesses are in the target of information security attacks. Those who work by using e-services are most in danger. The attacks could be from hackers, viruses or internal employees. There is no way to find which kind of attacks may occur but all know that this would lead to lose a large amount of money, time and other resources. The organisations spend millions of dollars on technical security equipments such as firewalls, IDSs, encryption tools and anti-viruses to protect them against the threats. Nonetheless, always there is a cleaver intruder or unknown vulnerability that may make a successful attack. [1]

Regarding to CSI/FBI survey 2007, 13% of companies which are participated in the survey have no idea that how much they spent for security in last year. The 48% of them suppose that they should invest just 1% of IT budget for security awareness but just 39% are using ROI (Return on Investment) to ensure how much is enough to spend on security. The 46% of companies have obviously found at least one security incident in the past 12 months but only 29% of them have security risk management techniques in progress. What is the most challenge for these companies? The answer is simple. They don't know about what they have, and what they need. They want to know which asset or technology has a security risk and for which one, they have enough security control to protect. [2]

To manage this challenge, the organisations need an information security risk management to evaluate asset's values and related risks. The risk management is usually a human activity which includes risk assessment, strategy development and risk mitigation by using managerial resources. The risk assessment is a process which identifies the assets, their values, threats and their consequences. A significant part of risk assessment which identifies threats and vulnerabilities, is very relevant to the past incidents, their likelihood and impacts.

* Farhad Foroughi is with University of Sunderland

II. The Problem

The risk is the logical time of likelihood to impact. The likelihood is the rate of occurrence and the impact is the weight of loss. In this definition, the prioritization of amount of loss and rate of occurrence is crucial. It means, the handling of greatest probability against of greatest loss may make a difficulty in risk calculation. For example, we have an asset with high probability of a threat but low loss versus another asset with low probability of attack and high amount of loss. Which one is more under the risk?

The answer to this question is not easy. It depends on the asset characteristics, the existing controls and before experiences. In these situations, an expert opinion needed to make a decision, but it could not help without any scientific theory or technology to support that.

The continual risk assessment is a problem in large organisations and complex business environments which produce or use information assets. In other mean, the problem is the risk identification and evaluation of new assets according to their properties, current security controls and consequences of before incidents.

III. The Solution

A. Intelligent Software Agent

We need a system that could assist experts or works on behalf of them to assess the risks during the normal working processes. The system should be reactive and autonomous because it is needed to respond immediately and

independently of events. It should also be communicative and cooperative with logs and reports which are made in relation with other databases and past experiences. The learning capability is very significant for this system because it should learn from past incidents and others which made by itself. The flexibility is also important because the factors and parameters may change during the time or special circumstances. [3]

An intelligent software agent is the best solution for this aim. It could perform various tasks on behalf of human experts and has all properties which the system needs. It learns risk of past experiences regarding to risk factors and asset's properties, and predicts the probability of new risk for a new instance. It could also dynamically adjust itself by new decisions which are made and their results. This will increase the accuracy of the prediction. For this reason and because we need to predict the probability of risk, the Bayesian learning theory is the best choice for this intelligent agent. The Bayesian learning theory is based on conditional probability and the risk evaluation is an uncertain prediction under conditional assumption. We have data set of past incidents and consequences. The instances in that data set classified by common asset attributes and common threat and vulnerability groups. It could link the assumptions and make a probabilistic prediction. We just need to make the data set as the knowledge and training data for the learning method and define the optimal hypothesis. [1]

B. Risk Assessment and BBN

The first step in risk management is establishing risk assessment and asset identification. The potential risk identification could run after this assessment. A risk is the probability of cause of a problem when a threat triggered by vulnerabilities. The source of the problem is vulnerability and the problem itself is threats. Threats are much related to the characteristics of the assets and vulnerabilities are relevant to the security controls. [1]

We need to develop a causal diagram which could represent the probable source of security breaches to evaluate the risks. In this case, the Bayesian Belief Network is the choice because it could graphically represent the probabilistic relationships regarding to the data set which we have. For better result and most real prediction, the model should set up a list of risk factors and impacts which are common in all incidents. The BBN could be made by creating the structure of the network and the probability estimation of each node. The fist one will present by diagram and the second one will calculate through mathematical procedure which is associated to the training data set.

In the year of 2002, the British standard Institute developed a guideline for information security risk assessment and identified the most common threats, vulnerabilities and risk factors. The model describes the asset attributes and security control categories which are critical for risk probability calculation. The asset attributes will indicate the impact and threats and the security control categories will represent the source of problems. The occurrence rate will also involve in risk calculation. [4]

C. Risk Calculation and Knowledge Requirements

According to BSI PD-3002:2002 and Data-Centric Quantitative Computer Security Risk Assessment research [5] the risk of an information system's asset could be determined by the following formula:

$$Risk = Impact \times Occurrence\ Rate \times (Threat \times Vulnerability)$$

From the same research, the threat is "potential violation of security" and vulnerability is a weakness in security controls which increase the probability of threat occurrence. Impact is the weight cost of losing an asset. This cost depends on the asset characteristics and its value for organisation. The asset's value for organisation could be presented by its classification. The occurrence rate is the count of a threat which is occurred in one year (Annualized Rate of Occurrence: ARO). The Combination of Impact and ARO is Annualized Loss Expectancy (ALE).

$$ALE = Single\ Loss\ Expectancy\ (SLE) \times Annualized\ Rate\ of\ Occurrence\ (ARO)$$

$$\Rightarrow Risk = ALE \times (Threat \times Vulnerability)$$

According to this result, we need information about Single Loss Expectancy, Annualized Rate of Occurrence, Threats and Vulnerabilities. The ARO is the rate of occurrence in the past and is available through logs. For SLE, we need to find the classification (we call it C in the formula) of the asset in organisation's documents and the properties of the asset. Regarding to [6] Research , the asset value depends on asset content. In information systems, each asset could have one or more factors of the following: [6]

- Financial Focus (AC1)
- Customer Focus (AC2)
- Process Focus (AC3)
- Renewal and Development Focus (AC4)
- Human Focus (AC5)

Furthermore, according to BSI PD-3002 (2002), the common threats in information systems could be categorized in the four groups: [7]

- Physical and Environmental (T1)
- Computer and Network (T2)
- Business Continuity (T3)
- Compliance (T4)

In addition, from the same guideline, the common vulnerabilities are related to the following security objectives:

- Personal Security (V1)
- Physical and Environmental Security (V2)
- Computer and Network Management (V3)
- System Development and Maintenance (V4)

By using Bayesian Belief Network (BBN) we could determine the relationship between these factors and their probabilities to risk evaluation. The BBN diagram is presented in figure 1 in appendix.

According to the BBN diagram:

$$P(Risk) = P(Impact) \times P(Occurrence\ Rate) \times P(Probability)$$

$$\Rightarrow P(R) = (P(Asset\ Value) \times P(Classification)) \times P(Occurrence\ Rate) \times (P(Threat) \times P(Vulnerability))$$
$$\Rightarrow P(R) = (P(AC1) \times P(AC2) \times P(AC3) \times P(AC4) \times P(AC5) \times P(C)) \times P(ARO) \times (P(T1) \times P(T2) \times P(T3) \times P(T4) \times P(V1) \times P(V2) \times P(V3) \times P(V4))$$

The above sentence means that probability of risk is equal of time of all factors' probabilities together.

On the other hand, regarding to Bayes Theorem, $P(h \mid D) = (P(D \mid h) \times P(h)) / P(D)$.

$P(h \mid D)$ means the probability of h happening given that D has happened. The $P(h \mid D)$ is the predictable risk because it means the probability of risk when instance D has happened. D is a set of training data. $P(D)$ denotes the prior probability that training data D will be observed. Because the h is independent on D, we could ignore $P(D)$.

Because instance h described by a set of attributes, we could use Naïve Bayes Classifier to simplify the formula. The Naïve Bayes Classifier will use when the target function $f(x)$ can take any value from some finite set attributes.

$$P(a1,a2,...,an \mid vj) = \prod P(ai \mid vj)$$

$$\Rightarrow P(C,AC1,AC2,AC3,AC4,AC5,ARO,T1,T2,T3,T4,V1,V2,V3,V4 \mid D) = P(AC1) \times P(AC2) \times P(AC3) \times P(AC4) \times P(AC5) \times P(C) \times P(ARO) \times P(T1) \times P(T2) \times P(T3) \times P(T4) \times P(V1) \times P(V2) \times P(V3) \times P(V4)$$

The Training data come from the log files and achieved incident reports of organisations. If there is no data available, this is possible to gather this information from international institutes which are making surveys and reports around the incidents such as CSI/FBI and SANS.

D. A Sample

Regarding to above formula and the training data which is made from company A's achieved logs by windowing technique, the risk of a new instance will calculate. The new instance is an asset by the following attributes:

Asset name: Annual Financial Report
C : Private, AC1: Yes, AC2:No, AC3:No, AC4:Yes, AC5:Yes, T1: Yes, T2:No, T3:Yes, T4:No, V1:Yes, V2:Yes, V3:Yes, V4:Yes, ARO:1

$$\Rightarrow P(C,AC1,AC2,AC3,AC4,AC5,ARO,T1,T2,T3,T4,V1,V2,V3,V4 \mid D) = P(AC1) \times P(AC2) \times P(AC3) \times P(AC4) \times P(AC5) \times P(C) \times P(ARO) \times P(T1) \times P(T2) \times P(T3) \times P(T4) \times P(V1) \times P(V2) \times P(V3) \times P(V4)$$

$$\Rightarrow P(Risk \mid High) = (4/6) * (4/6) * (3/6) * (2/6) * (3/6) * (3/6) * (1) * (4/6) * (3/6) * (4/6) * (1) * (4/6) * (1) * (1) * (2/6) = 0.0914$$

$$\Rightarrow P(Risk \mid Medium) = 0$$
$$\Rightarrow P(Risk \mid Low) = 0$$

The probability of risk to be high is 0.0914 and for risk to be medium and low are 0. It means that risk is probable to be high.

IV. Recommendation for Future Research

One of the most obvious factors which affect on the risk is external factor. External factor is an event that is happening out side of the organisation and may be social, economical or political. The Intelligent agent could make a connection to other databases which track these events and adjust itself and the data set by making change in external attribute. It will help the organisations to adjust themselves with the best security controls all the times. The agent could also use Data Mining techniques to determine accuracy parameters.

APPENDIX

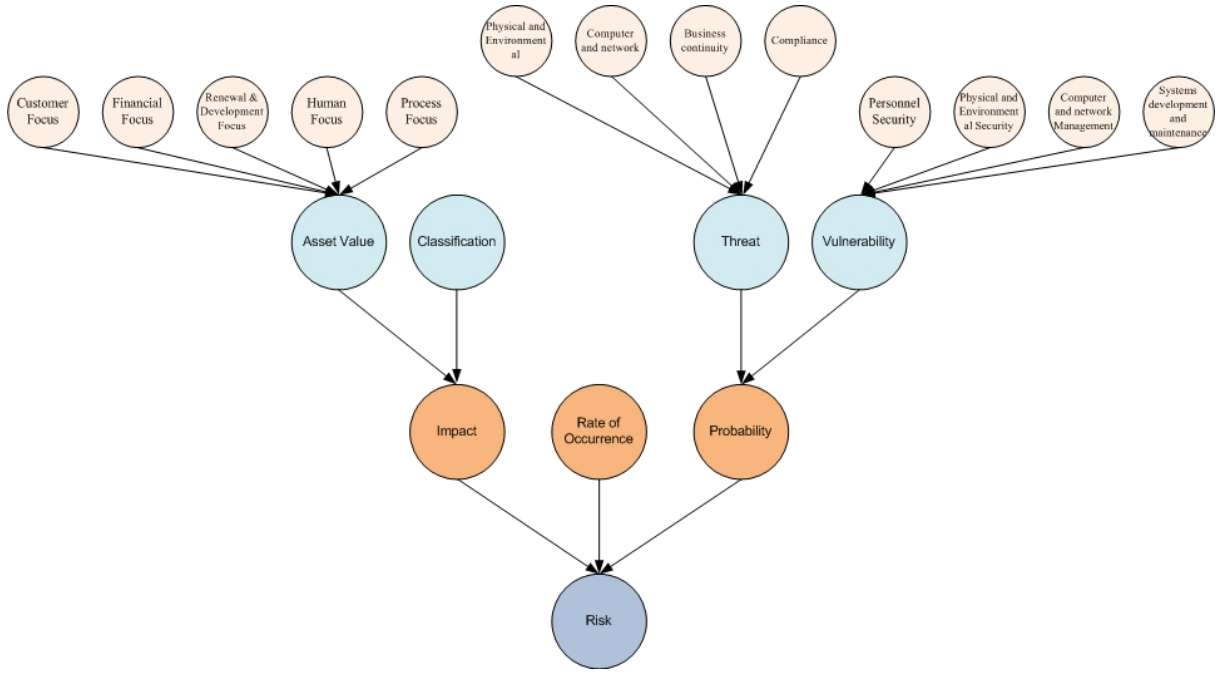

Figure 1: Information Security Risk Assessment – Bayesian Belief Network

Table 1: Company 'A' – Training Data

Asset Characteristics / Impact							Threat				Vulnerability				Probability	Risk
Asset		Asset Content					Threat Group									
	Classification	Financial Focus	Customer Focus	Process Focus	Renewal and Development Focus	Human Focus	Physical and Environmental	Computer and Network	Business Continuity	Compliance	Personal Security	Physical and Environmental Security	Computer and Network Management	System Development and Maintenance	Occurrence Rate	Risk Level
A	C	AC1	AC2	AC3	AC4	AC5	T1	T2	T3	T4	V1	V2	V3	V4	ARO	R
1	P.	Y	Y	Y	Y	N	Y	N	Y	N	Y	Y	Y	Y	1	High
2	P.	Y	N	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	1	High
3	P.	N	Y	N	Y	N	Y	Y	N	N	Y	N	Y	Y	2	High
4	P.	N	N	N	N	Y	Y	N	N	N	Y	N	Y	Y	3	High
5	S.	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	2	High
6	S.	Y	N	Y	N	Y	Y	N	N	N	Y	Y	Y	Y	2	High
7	S.	N	Y	N	Y	N	Y	N	N	N	Y	N	Y	Y	1	Medium
8	S.	N	N	N	N	Y	Y	N	N	N	Y	N	Y	Y	1	Medium
9	E.	Y	Y	Y	N	N	Y	N	N	N	Y	Y	Y	N	1	Medium
10	E.	N	N	N	N	N	Y	N	N	N	Y	Y	Y	N	2	Medium
11	E.	N	Y	N	N	Y	N	Y	N	Y	Y	N	Y	Y	2	Medium
12	E.	N	N	N	N	Y	N	Y	N	N	Y	N	Y	N	1	Low
13	Pu.	Y	Y	N	N	N	N	N	N	N	Y	N	N	N	0	Low
14	Pu.	N	N	N	N	N	N	N	N	N	Y	N	N	N	1	Low
15	Pu.	N	N	N	N	N	N	N	N	N	Y	N	N	N	1	Low

Table 2: Company 'A' – Asset Classification Table

Classification	Definition	Examples
External (E.)	Security and handling requirements are given by another entity outside of company	- Data from a government program - Controlled information from a business partner
Private (P.)	If disclosed could cause serious harm to business	- Specifications or drawings of products - Business plans/strategies
Sensitive (S.)	If disclosed could cause moderate harm to business or personnel	- Salary information - Sales figures - Organization charts
Public (Pu.)	Data is not sensitive	- Company picnic plans - Sales literature

REFERENCES

[1] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti & S.K. Sadhukhan. (2006, 01, 07). e-Risk Management with Insurance : A framework using Copula aided Bayesian Belief Networks, *Proceedings of the 39th Hawaii International Conference on System Sciences*.

[2] CSI/FBI (2007, 12, 03). The 12th Annual Computer Crime and Security Survey, [Online]. Available: http://www.gocsi.com/forms/csi_survey.jhtml.

[3] X. Wang, H. Kiliccote, P.K. Khosla & C. Zhang. (2000). Agent-based Risk Learning for Computing Systems, *Proceedings of the Fourth International Conference on MultiAgent Systems*, pp. 459 - 460.

[4] A.K.T. Hui & D.B. Liu (2004, 01, 29). A Bayesian Belief Network Model And Tool To Evaluate Risk And Impact In Software Development Projects, *Proceedings of the 2004 Annual Symposium of Reliability and Maintainability,* pp. 297-301.

[5] B. Berger. (2003, 08, 20). Data-Centric Quantitative Computer Security Risk Assessment, [Online]. Available: http://www.sans.org/reading_room/whitepapers/auditing/1209.php.

[6] N. Bontis. (2001, 03). Assessing knowledge assets: a review of the models used to measure intellectual capital, *International Journal of Management Reviews*, vol. 39, no. 1, pp. 41-60.

[7] BSI (2002, 12, 17), *PD 3002:2002 - Guide to BS 7799 Risk Assessment*, London: British Standards Institution.