STUDIA UNIVERSITATIS

Revistă științifică a Universității de Stat din Moldova, 2007, nr.8

MODELING AND PREDICTING OF RISK FINANCIAL LOSS IN INFORMATION SYSTEMS

Alexandru OBJELEAN

Moldova State University

Este prezentat un model de predicție a riscului pierderii informației în sisteme informatice bazate pe Modele Markov Ascunse. Modelul Markov descrie un lanț de stări, care prezintă un volum de pierderi financiare în caz de o intervenție nelegitimă la informația autorizată. Acest model permite utilizatorului sistemului informatic să determine probabilitatea pierderilor financiare ale organizației în cazul când încercările accesului nesancționat sunt staționare.

Introduction

There are many pressing issues within computer security, and risk analysis is one of the most critical. Computer security risk analysis methods are in their infancy, and as such, there are no techniques based on rigorous, defensible and generally accepted standards of measurement currently available. Decision makers lack credible data about risk, and there is no accepted methodology to forecast the consequences of computer security choices. The creation of these techniques is crucial if executive management is to have the information they need to make decisions on where to best allocate corporate resources. To complicate matters, at times security capabilities may compete with or impede functional capabilities. As a result, funds are being dispersed in a manner that may or may not result in improved security in the organization.

In this paper we present a model based on Markovian decision processes to predict the security losses of an organization, and processes that can use to calibrate such a model.

1. The Markov Model

In this paper we use a Markov Model based approach to help us model the financial impact of cyber attacks on systems in response to attacks over specific periods. The objective of this model is to provide organizations with a quantitative tool to help management estimate financial loss expectancy due to cyber attacks over a given period. It will allow organizations to make better decisions regarding security expenditures by providing the ability to estimate losses caused by security incidents. In this study, we focus on virus attacks, which are more costly than all computer security breaches [1]. The model can be expanded to include other types of attacks as data is collected for these attacks. Markov models exhibit the characteristics necessary for modeling security losses, namely, dynamic, stochastic (probabilistic), and time dependent [2]. These models are useful in describing sequential data whose components exhibit strong dependencies and are helpful in finding patterns that appear over a period. Markov chains are memory less, that is, future states depend only on the current state and not on any previous states. In other words, although a sequence of events is based on the probability of events preceding it, the probability of future states depends only on the current state.

Markov theory allows sensitivity calculations (“what if” questions) to be easily carried out and gives insight into changes in systems over time. It can look at a sequence of events and analyze the tendency of one event to be followed by another. Using this analysis, you can generate a new sequence of random but related events, which will look similar to the original. Markov techniques decrease what seems like an enormous task by reducing the problem from one of mathematical computation to that of state modeling [2]. Markov Models are currently being used in fields such as risk analysis, speech recognition, and artificial intelligence. They have been used to estimate loss expectancy rates in the accounting world due to bad debt and credit risk [3, 4, 5]. This study will perform similar loss expectancy rate estimates, but in this case, the losses are due to computer security breaches.

The formal definition of a Markov Model states [6]:

- $A = \{a_{ij} = P(q_j \text{ at } t+1 \mid q_i \text{ at } t)\}$, where P(a | b) is the conditional probability of a given b, $t \geq 1$ is time, and $q_i \in Q$.
  Informally, A is the probability that the next state is q_j given that the current state is q_i.
- $B = \{b_{ik} = P(o_k \mid q_i)\}$, where $o_k \in O$.
  Informally, B is the probability that the output is o_k given that the current state is q_i.
- $\Pi = \{p_i = P(q_i \text{ at } t=1.\}$

A simple Markov formula: $a_{kl} = P(\prod i = l \mid \prod_{i-1} = k)$.

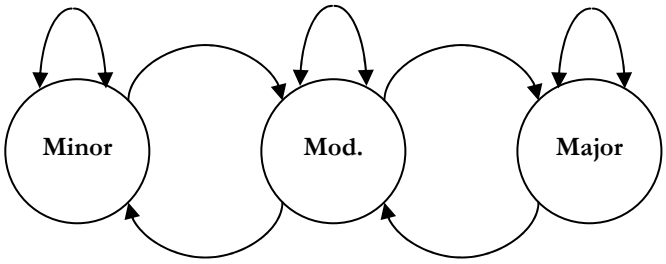
A generalized Markov transition matrix for N states:

$$P^{(n)} = \begin{array}{c|cccc} \text{State} & 0 & 1 & \cdots & M \\ \hline 0 & P00^{(n)} & & \cdots & P0M^{(n)} \\ 1 & \bullet & & & \bullet \\ \bullet & \bullet & & & \bullet \\ \bullet & \bullet & & & \bullet \\ M & PM0^{(n)} & & \cdots & PMM^{(n)} \end{array}, \text{ for } n = 0, 1, 2, \ldots.$$

A Markov transition matrix for three states, such as for our model, would look like:

State	Minor 0	Moderate 1	Major 2	
0 Minor	*P*00	*P*01	*P*02	= 1
1 Moderate	*P*10	*P*11	*P*12	= 1
2 Major	*P*20	*P*21	P22	= 1

This is a three by three square matrix. The transition probability entries measure the likelihood that an entity with dollar losses that place them in a particular category will move to another category during the applicable period of time.



Example of state transition diagram for a three-state Markov model

Let us consider total dollar losses from virus attack at time *i*. The losses can be classified then, or at any subsequent time, into one of three categories, minor, moderate or major. L_0 represents entities with dollar losses that place them in the minor category, L_1 are entities with dollar losses in the moderate category, and L_2 are entities with dollar losses in the major category. L_{jk} represents an entity in category *k* at time *i* + 1, which came from category *j* at time *i*. The transition probability, P_{jk} is defined as the probability of an entity with losses in classification *j* at time *i* transitioning to classification *k* at time *i* + 1. In terms of the matrix entries, L_{jk}, the transition probabilities P_{jk} are defined:

$$Pjk = \frac{Bjk}{\sum_{0}^{n} Bjk} (\mathrm{k} = 0, 1, \bullet\bullet, n)$$

Markov models require three types of information:
1) the various possible states,
2) the initial beginning state,
3) the transition probabilities.

The initial state is a vector and the transition probabilities a matrix. To build this model we will define a finite Markov chain with three mutually exclusive and collectively exhaustive states (one (minor loss), two (moderate loss) and three (major loss)), and use the data we collect to populate the transition matrix. We can then define a system state at a particular point in time, and predict the probabilities that the system will move from one state to another or remain where it is. This will help us predict the probability that a corporation will transition to other states of financial loss or remain the same over a defined period of time (six months, or a year for example) from virus attacks. This model will help to answer questions such as the following: If a company begins in state one, what is the probability it will be in the same state after six months, twelve months or two years? What is the probability it will transition to state two in six months or eighteen months? What is the probability it will transition to state three in twelve months or two years? If a company begins in state three, what is the probability it will be in the same state after six months, twelve month or two years? What is the probability it will transition to state two in six months or eighteen months? What is the probability it will transition to state one in twelve months or two years? It will also be interesting to determine if the data shows that this model has a "steady state distribution". In other words, can it be assumed that at some point the model reaches equilibrium where the numbers in each state stays the same?

For example, using hypothetical data, the model might look like this:

State	Condition
1	Minor or <1% of gross revenue lost from virus attacks
2	Moderate or 1.1 – 2.5 % of gross revenue lost from virus attacks
3	Major or 2.51 – 5 % of gross revenue lost from virus attacks

After collecting data from our survey and secondary sources, such as Symantec, the initial state vector might look like this:

State	1 2 3
Values	[.2, .4, .3]

Assuming no operational or system changes, the transition matrix that would look like this:

State	1	2	3
6 months	.1	.6	.2
12 months	0	.5	.3
18 months	0	.2	.5
24 months	0	0	0

2. Collection of Data

There are several issues related to collection of data especially since organizations are reluctant to share this information for privacy reasons, and out of fear that public disclosure of a security incident might have an adverse impact on their reputation. This is compounded by the fact that to compute the tangible and intangible financial losses incurred as a result of a virus attack is an arduous, labor-intensive task. Finally, thus far, few managers have attempted to ascertain the reduction in financial loss due to investment in security controls, so this may be the most difficult part of all. That is the reason we will turn to vendors such as Symantec to provide this information as an alternative to collecting our own data.

3. Conclusion

A model based on Markov decision process has been developed to estimate the future losses in an organization based on specific attacks. Such a model will enhance management's decision-making process by providing a quantitative risk assessment tool to predict the probability of various levels of financial loss

due to virus attacks given certain assumptions over a defined period of time. We believe that this model will provide valuable insight into better methods to justify expenditures in security and result in improved awareness and communication of security issues within the entity. This quantitative tool will contribute to a consistent and objective basis for decision-making and enhance the productivity of the security team and aid in the development of standard baseline security processes. It can help to answer many questions, such as the following. Is it necessary to spend money to update anti-virus software? Should the organization purchase a firewall? How should management balance the cost of investing zero dollars in security if they knew there was a forty percent chance that they would incur a financial loss due to virus attacks that would equal to one percent of their revenues during the next year; This model could also help the insurance industry determine cyber security insurance premiums. The use of more accurate historical data and a quantitative risk model, both of which are accepted methods used to issue more traditional insurance policies, would give insurance companies the ability to predict financial loss for a particular organization. This could allow them to set premiums based on this prediction. By implementing an effective risk management program, organizations can progress to view security and controls as business enablers. Security can be viewed as a component of business operations by mitigating risks in a cost-effective manner. The drawbacks of this type of risk analysis, is associated with the unreliability and inaccuracy of the data. Probability can rarely be precise, in some cases, might promote a false sense of precision and complacency. A disadvantage of Markov Modeling is that of a simplified model of a complex decision-making process.

Repherences:

1. Dimitrakos T., Rithie B., Rapis D., Stølen K. Model Based Security Risk Analysis for Web Applications: http://rac.alionscience.com/pdf/3Q2003.pdf.
2. Nong Ye. A Markov Chain Model of Temporal Behavior for Anomaly Detection, Proceedings of the 2000 IEEE Workshop on Information Assurance and Security, United States Military Academy, West Point, NY 2000, June 6-7.
3. Jonsson Erland. A Quantitative Model of the Security Intrusion Process Based on Attacker Behavior, IEEE Transactions on Software Engineering, 23(4), 1997.
4. The Applicability of Markov Analysis Methods to Reliability, Maintainability, and Safety, START – Related Topics in Assurance Related Technologies, Volume 10, Number 2, http://rac.alionscience.com/pdf/3Q2003.pdf.
5. Skora Richard. Modeling Credit Risk, Financial Engineering News. http://www.fenews.com/fen16/creditrisk.html.
6. http://www.nist.gov/dads/HTML/hiddenMarkovModel.html.

Prezentat la 13.09.2007