SAPIENZA
UNIVERSITÀ DI ROMA

SAPIENZA UNIVERSITY OF ROME

MASTER'S DEGREE IN CYBERSECURITY

DEPARTMENT OF COMPUTER SCIENCE

Survey

Applications of statistics to cybersecurity

Author:
Marco Ruvolo

Professor:
Tommaso Gastaldi

Matricola number:
1883257

Course:
Statistics

December 7, 2022

Contents

List of Figures

List of Tables

1 Introduction

1.1 Aims and Objectives

The main objective of this paper is to perform a survey on some applications of statistics in cybersecurity.

In particular: in Chapter 2 the risk management process is described and a Monte Carlo Method-based approach for evaluating information security investments is examined; in Chapter 3 the IDS systems are introduced and, above all, statistics-based techniques for developing these systems are briefly explained.

Lastly, in Chapter 4 some considerations are made and conclusions are drawn.

Actually, this paper is rather a survey than a proper innovative research paper: indeed, most of the results presented below have been excerpted from several papers published within the last three decades.

The references to these latter can be found in Chapter 5.

1.2 Paper composition tools

This paper has been composed using the online LaTeX editor Overleaf and the Report Template(UoB) class.

2 Risk management

According to IBM [1], risk management is the process of identifying, assessing and controlling the risks to an organisation's capital and earnings: these risks could arise from different sources, including accidents, malicious attacks and natural disasters.

The impact of unforeseen negative events could be catastrophic and have significant consequence on (the assets of) your organisation. Therefore, a proper approach to risk management can help identify, manage and mitigate risks. The three fundamental steps (i.e., risk identification, risk analysis and assessment and risk mitigation and monitoring) of the risk management process are described below.

Risk identification is the process of identifying and assessing threats (e.g., assessing IT security threats such as malwares and malicious attacks).

Risk analysis consists of establishing the probability of occurence of each risk and the potential outcomes of each negative event.

Risk assessment evaluates each risk, comparing the magnitude of each of them and ranking them with reference to their significance and consequences.

Risk mitigation is the process of planning and developing methods and options to limit threats to your organisation objectives. Since this is a continuous process that evolves over time, risk monitoring can help in the coverage of known and unknown risks.

Moreover, there are five commonly accepted strategies for addressing risk:

- risk avoidance: mitigating the risk by avoiding activities that may negatively affect the organisation;
- risk reduction: accepting the risk, minimising its consequences (i.e., focusing on keeping the loss contained, preventing it from spreading);
- risk sharing: transferring the possibility of loss from the individual to the group;
- transferring risk: contractually transferring a risk to a third-party (e.g., insurance shifts the risk from the organisation to the insurance company);
- risk acceptance and retention: since it is virtually impossible to eliminate all risk (except by totally avoiding it), some risk will remain (i.e., residual risk).

2.1 Information security investments

Usually, information security investments compete for resources (e.g., budget) with other investment opportunities. Though, the financial decision makers should choose how to allocate financial resources and which investments to carry out. Although investing in new equipment or infrastructure to support an alternative business opportunity could seem profitable, the impact of possibly successful cyber attacks should be taken in consideration: the benefits of these investments could be made irrelevant by a successful attack [2].

As for any organisation, security managers need to measure their investments' cost-effectiveness, thus justifying their budget usage and providing arguments for their next budget claim. However, since security is an investment that tipically provides loss prevention rather than direct profit [3], measuring the effectiveness and the cost of information security activities is often difficult.

Among the financial metrics used to estimate risks, there is the ALE (Average Loss Expectancy), calculated as the sum of the products of annual consequences and frequencies of occurrence [4]. Since ALE dangerously equates likely but low impact events with unlikely but high impact events and since ALE-based risk models become overly complex when used to address all threats, assets and vulnerabilities, security risk measures require more robust approaches in identifying the true potential of threat exposure [5]: the usage of a stochastic approach (e.g., Monte Carlo Method) to evaluate risks is recommended [6].

2.2 Monte Carlo Simulation

Monte Carlo Simulation (also known as the Monte Carlo Method) is a mathematical technique used to estimate the possible outcomes of an uncertain event: it predicts a set of outcomes based on an estimated range of values (instead of fixed input values, like other forecasting methods).

A Monte Carlo Simulation builds a model of the possible outcomes by exploiting a probability distribution, for any variable that has inherent uncertainty and it, then, recomputes the results multiple times, each time using a different set of random numbers within the bounds: as the number of inputs increase, the number of forecast also grows, allowing you to project outcomes farther out in time with more accuracy. Upon the completion of a Monte Carlo Simulation, a range of possible outcomes with the respective probability of occurence is carried out.

The Monte Carlo techniques involves three basic steps:

1. set up the predictive model, identifying both the dependent variables to be predicted and the independent variables that will drive the prediction;
2. specify the probability distributions of the independent variables identified at step 1;
3. run simulations repeatedly, generating random values of the independent variables, until enough results are gathered to make up a representative sample of the possible combinations.

2.3 Monte Carlo Simulation as a risk management tool

Using a Monte Carlo Method-based approach in the analysis of information security investments is effective in evaluating the return on investments that defend against security attacks. This approach captures uncertainty in security modeling parameters (e.g., vulnerabilities, frequency of instrusion, damage estimates) and expresses its impact on the model's forecast (e.g., projected benefit).

In a Monte Carlo Simulation, the security model is treated as a function that takes a set of parameters and returns a set of (forecasted) result, as can be seen in figure 1.

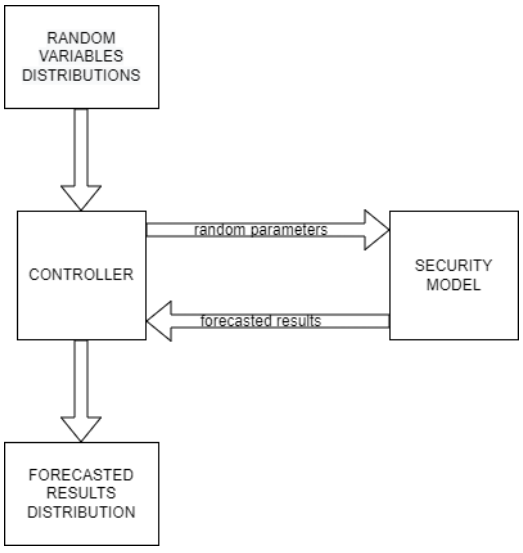

Figure 1: Schema of the Monte Carlo Simulation model

More precisely:

1. a set of random variables are identified (they represent the expert's estimates);
2. a probability distribution is specified for each of the random variables identified at step 1 (the distributions represent the expert's estimates);
3. a random value for each parameter, sampled according the distribution associated to the respective random variable, is given to the security model (i.e., selection);
4. the security model is executed within the values passed as parameters (i.e., execution);
5. the (forecasted) results are collected (i.e., collection).

Naturally, selection, execution and collection phases are repeated in many iterations (usually thousands) of the security model.

Before building a simulation, it is fundamental to ensure that each iteration of the simulation is supplied with independent parameters (i.e., the input random variables must be independent) in order to carry out meaningful results.

2.4 A case study

Monte Carlo Simulation model can be used within the context of IT to reduce the disparities of opinion in resource allocation (e.g., allocating optimal resources to specific security protective assets or other business productive assets).

A Monte Carlo Simulation can perform quantitative risk analysis by assigning a probability distribution to uncertain parameters and through random sampling of the distribution: thus, it is possible to determine all potential outcomes under those uncertainties [7].

Using a conceptual enterprise as a case study and verifiable historical cost of security breaches as parametric values, the model described shows why using conventional risk assessment approach as budgeting process can result in significant over/under allocation of resources for cyber capabilities.

2.4.1 Overview

Traditionally, organisations use risk assessment models to determine the optimal allocation of resources to cyber capabilities. Then, financial decisions are based on the organisation's threat tolerance and its score from the risk scoring matrix.

The risk scoring matrix is calculated on the assumption that an event will happen given a probability of occurrence and a measure of the impact or severity of an associated security breach. Information security budget is then allocated based on the resultant estimated risk score.

According to the risk scoring formula, the risk R associated to an asset is given as the product of the probability P of occurence of possibly negative events and the impact I of these latters on the given asset, namely:

$$R = P * I \tag{1}$$

The values of P and I for a given asset is assigned based on security experts' opinions, statistic from reports, corporate level assessments or records from past events and the resultant single value represents the risk score R for that particular asset.

In practice, in real word problems, it is difficult to follow this approach in order to optimise resource allocation decisions: the deterministic estimates carried out might not reflect the actual organisation's context. Furthermore, a risk-matrix based approach may lead security assessors to assume that (under some assumptions) certain events would be true while completely ignoring the possible occurence of least significant events.

Hence, in order to explain how uncertainty affects security breach costs and resource allocation decisions to mitigate those risks, a case study is analysed below.

The scenario presented [8] refers to a bank and the following assumptions are made:

- there are five key assets (namely, DDoS Mitigation System, Personnel and third party contractors, Data Backup and Recovery System, Incident Response Solution, and Antivirus Software) that need to be safeguarded from security threats;
- resource allocation decisions are based on the impact of breaches to those assets and how they may impact banking operations.

Therefore, security breach costs estimation approaches are described below.

2.4.2 Deterministic estimation of security breach costs

This approach is based on the use of a conventional risk assessment models to determine appropriate resource allocation.

Each risk is ranked within a five level scale of likelihood and a scale of (financial) impact, as shown in tables 1 and 2.

LIKELIHOOD	DESCRIPTION	FREQUENCY OF OCCURRENCES
1	An incident is expected to occur in exceptional circumstances (e.g., once in 10 years)	Rare/Very Low
2	An incident may occur at some point (e.g., once in 3 years)	Possible/Low
3	An incident will occasionally recur (e.g., once in a year)	Probable/Medium
4	An incident will occur in most circumstances (e.g., once every 4 months)	Certain/High
5	An incident is certain to occur in most circumstances (e.g., once every month)	Frequent/Very High

Table 1: Risk likelihood

SEVERITY	DESCRIPTION	EXAMPLE OF BUSINESS IMPACT
1	None: no disruption of service	Financial loss $<$ 1000€
2	Minor	Financial loss $<$ 10'000€
5	Moderate	Financial loss $<$ 100'000€
10	Significant	Financial loss $<$ 1'000'000€
15	High	Financial loss $>$ 1'000'000€

Table 2: Risk severity

With reference to the risk scoring formula yet presented, the risk scoring matrix can be computed by taking into account the likelihood and severity values of each risk. In practice, risk scoring is carried out by multiplying the likelihood of each risk by the severity of that risk occurring.

LIKELIHOOD / SEVERITY	NONE	MINOR	MODERATE	SIGNIFICANT	HIGH
FREQUENT	5	10	25	50	75
CERTAIN	4	8	20	40	60
PROBABLE	3	6	15	30	45
POSSIBLE	2	4	10	20	30
RARE	1	2	5	10	15

Table 3: Risk scoring matrix

After the risk analysis phase, given an organisation risk threshold and the risk score number, the budget is allocated for countermeasures to mitigate risks in that context.

2.4.3 Probabilistic estimation of security breach costs

As the assets grow, probabilistic estimation approaches can be used in place of conventional deterministic ones, in order to address the huge amount of uncertainties associated with these latter.

Using Monte Carlo Method, the probabilistic costs of security breaches for each asset in a given scenario can be determined. In fact, a Monte Carlo Simulation works by sampling several scenarios from a probability distribution instead of static estimates (e.g., like in the deterministic approach just described). Probabilistic estimation assigns minimum and maximum cost boundaries for each security breach. The combined cost of all security breaches is then calculated as the total minimum and maximum cost of a security breach for each asset in order to project total resource allocation for the enterprise. In that case, it is possible to establish absolute

bounds for allocated resources to the entire enterprise. Monte Carlo may not be able to tell with certainty the exact cost of a breach, but it can describe the probability of cost associated with security breaches, to aid resource allocation. In comparison to the deterministic approach, the probabilistic estimate is also based on random variables, however, each estimate follows a particular distribution, independent and unaffected by other variables.

ASSET	SECURITY INCIDENT	C = COST OF BREACH
DDos Mitigation System	DoS/DDoS attack	53'477€
Personnel and third party contractors	Fraud, malicious insider	40'403€
Recovery System	Data loss, stolen devices	39'905€
Incident Response Solution	Cyber espionage	69'026€
Antivirus Software	Malicious code infection	31'572€

Table 4: Expert estimation of security breach costs

Consider the deterministic cost of breach for the DDoS Mitigation System as described in table 4. Under probabilistic estimation approach, *blurring* parameter can be used to suggest that in place of a fixed quantity like 53'477€, the minimum value in of 30'000€ and the maximum value of 65'000€ in a distribution could be included, as shown in table 5. Essentially, a fixed value is replaced with a probability distribution, which is a true representation of the state in the real world. Hence, the fixed quantity is now our most likely value, but it is not the only possible value in the distribution. The key to Monte Carlo simulation is that each variable is assigned a random value, and the total value is calculated thousands of times during the simulation. It, therefore, allows us to understand the risk that expectations may not match reality, hence, appropriate precautions can be taken [9].

ASSET	SECURITY INCIDENT	C_{min}	C_{ml}	C_{max}
DDos Mitigation System	DoS/DDoS attack	30'000€	53'477€	65'000€
Personnel and third party contractors	Fraud, malicious insider	20'000€	40'403€	50'000€
Recovery System	Data loss, stolen devices	25'000€	39'905€	45'000€
Incident Response Solution	Cyber espionage	35'000€	69'026€	75'000€
Antivirus Software	Malicious code infection	15'000€	31'572€	37'000€
Total		123'000€	234'383€	272'000€

Table 5: Model simulation parameters

It is difficult to compute values for multiple scenarios without some form of simulation, especially if factoring in multiple assets and security breach costs, as part of the budgetary allocation process, is needed.

2.5 Setup

There are two basic assumptions for this model:

- Key information asset points are determined by an organisation CIO and the security team;
- Minimum and maximum values of security breach costs are subject to expert elicitation, based on experience and previous security breach events.

The work described in this paper use some security breach cost parametric values obtained from verifiable information security breach reports [10] and [11]. Please note that limitations of the costing methodology outlined in these latter are not validated nor described neither in this paper nor in Fagade et al.'s one [8].

Before starting the simulation, security breach costs are identified and converted into a range of values using a probability distribution (i.e., for each security breach cost estimate, fixed values are replaced with a probability distribution).

Since the triangular distribution is one of the most used probability distributions to draw out expert opinion (especially in the case of limited or absence of historical data), the latter is used in this simulation. The triangular distribution defines, for each asset, uncertain security breach cost values as a minimum C_{min}, maximum C_{max} and most-likely C_{ml} range of values.

Here, actually, the approach followed holds C_{min} and C_{max} constant while randomly selecting C_{ml} according to a triangular distribution.

For this simulation, MATLAB [12] and Vose ModelRisk software [13] are used: both tools allow configurable simulations with a very large number of runs and can generate thousands of scenarios for each set of uncertain inputs.

The simulation runs generated are 50'000, the model output is a probabilistic range of values and scenarios associated with security breach costs, as well as the probability distribution associated with those values.

2.6 Simulation results

Results of Monte Carlo Simulation are shown in figures 2 and 3 and discussed, as reported in Fagade et al.'s paper [8], below.

For each iteration, the following steps are performed:

1. samples are taken from each of the breach cost (triangular) probability distribution;
2. the average random value (i.e., the total probabilistic estimation of security breach) is computed, at the end of the current iteration.

At the end of the simulation (here, 50'000 iterations are run), the output histogram represents the generated scenarios (during the simulation, different scenarios are generated asccording to the probability of those scenarios occurring) for security breach cost.

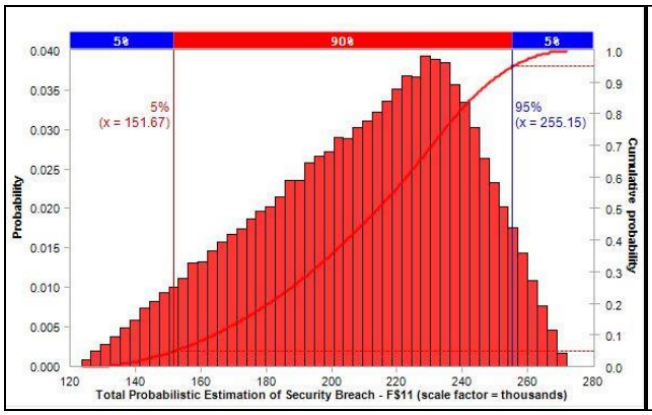

Figure 2: Simulation results in ModelRisk

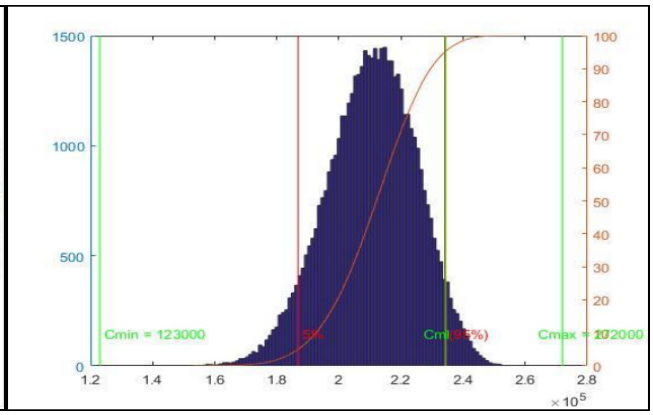

Figure 3: Simulation results in MATLAB

From the simulation results in figure 2, it can be seen that the upper 5% and the lower 5% represent extreme cases that are ignored by the simulation output.

From the model simulation parameters in table 5, it can be seen that the total resource allocation could be within the range of 123'000€ and 272'000€, but the realistic chance of resource allocation nearing these extreme values is very unlikely, hence the model ignored them.

Furthermore, it can be seen that 90% of the simulation iterations fall under a value less than the upper bound estimated total values: hence, we can say that 90% of the total allocation will meet the initial estimate

(this is not a guarantee, but it allows us to adjust IT security budget to match the cost of potential breaches and also understand the risk that resource allocation may not meet initial estimates).

Further analysis of the result in figure 2 shows that, given all the iterations of the simulation, the absolute minimum value of 149'794€ is much higher than the original deterministic lower bound value of 123'000€.

Similarly, the absolute maximum probabilistic value of 253'000€, after iterations, is much lower than the deterministic value of 272'000€ (with only 5% chance of the allocation going over the upper boundary).

The most likely point estimate is around the value of 290'000€; from the location of the peak of the distribution, it can be seen that this value is rather more realistic than the deterministic value of 234'383€ (this does not rule out the possibility that the cost of impact could be significantly higher, possibly twice as high in terms of cumulative percentage).

Now, results are compared with another simulation in MATLAB, as shown in figure 3, using the same input parametric values.

The invariant that holds in both states of the models is that extreme values are ignored in the output of both simulations. While both models follow a similar distribution, it can be seen that not only did both simulations ignore lower and upper bound values, but also shows higher C_{min} and lower C_{max} than the deterministic values.

3 IDS and anomaly detection

An intrusion can be defined as an unauthorised activitiy that causes damage to an information system: essentially, any attack that could pose a possible threat to the information confidentiality, integrity or availability is an intrusion.

An IDS (Intrusion Detection System) is a software or hardware system that identifies malicious actions on information systems in order to allow for system security to be maintained. The goal of an IDS is to identify malicious network traffic and computer usage, which cannot be identified by a traditional firewall, in order that high protection against actions that compromise the availability, integrity, or confidentiality of computer systems is achieved. IDS systems can be categorised into two groups: SIDS (Signature-based Intrusion Detection System) and AIDS (Anomaly-based Intrusion Detection System).

SIDS and AIDS systems will be described, according to Khraisat et al.'s work on IDS systems [14], in the following subsection.

3.1 Signature-based intrusion detection systems

SIDS systems are based on pattern matching techniques to find a known attack: when an intrusion signature matches with the signature of a previous intrusion that already exists in the signature database, an alarm signal is triggered (host's logs are inspected to find sequences of commands or actions which have previously been identified as malware). The main idea consists of building a database of intrusion signatures (signatures are created as state machines, formal language string patterns or semantic conditions) and comparing the current set of activities against the existing signatures and raising an alarm if a match is found.

SIDS systems usually give an excellent detection accuracy for previously known intrusions [15] while, on the other hand, they have difficulty in detecting zero-day attacks since no matching signatures exist in the database until the signatures of the new attacks are extracted and stored. The traditional technique used in SIDS systems consists of examining network packets and trying matching against a signature database. Unfortunately, this technique is unable to identify attacks that span several packets: due to the advanced sophistication of modern malwares, it may be necessary to extract signature information over multiple packets (this requires the IDS systems to recall the contents of earlier packets).

The increasing rate of zero-day attacks[16] has rendered SIDS techniques progressively less effective because no prior signature exists for any such attacks. Moreover, this traditional paradigm can be further undermined by polymorphic variants of malwares and the rising amount of targeted attacks. A potential solution to this problem would be to use AIDS techniques, which operate by profiling what is an acceptable behavior rather than what is anomalous, as described in the next subsection.

3.2 Anomaly-based intrusion detection system

AIDS is a system capable to overcome the limitation of SIDS: given the assumption that malicious behavior differs from typical user behavior, any significant deviation between the observed behavior and the model is regarded as an anomaly, which can be interpreted as an intrusion (so, the behaviors of abnormal users which are dissimilar to standard behaviors are classified as intrusions).

The development of an AIDS consists of two phases:

- the training phase: the normal traffic profile is used to learn a model of normal behavior;
- the testing phase: a new data set is used to establish the system's capacity to generalise to previously unseen intrusions.

In recent years, several AIDS approaches have been proposed for improving detection accuracy and reducing false alarms. Thus, AIDS systems can be classified into different categories based on the method used for training (i.e., for creating a normal model of the behavior of a computer system):

- statistical-based: the collection and examinination of every data record in a set of items and the building of a statistical model of normal user behavior are involved;
- knowledge-based: it tries to identify the requested actions from existing system data such as protocol specifications and network traffic instances;
- machine learning-based: complex pattern-matching capabilities are acquired from training data.

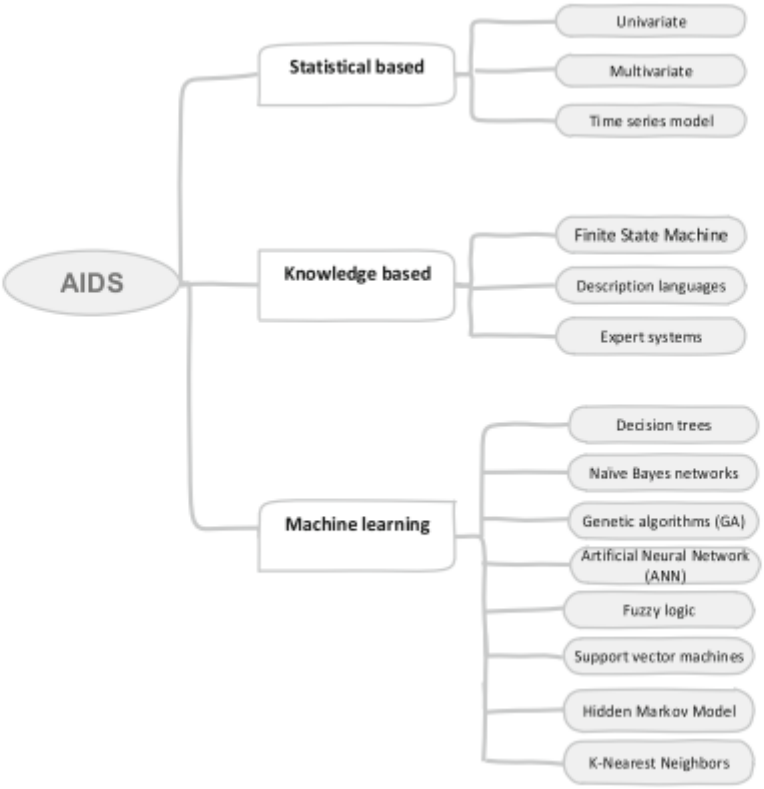

Figure 4: Classification of AIDS methods

The main advantage of AIDS is the ability to identify zero-day attacks due to the fact that recognising the abnormal user activity does not rely on a signature database: AIDS triggers a danger signal when the examined behavior differs from the usual behavior.

Furthermore, AIDS systems have various benefits:

1. they have the capability to discover internal malicious activities (e.g., if an intruder starts making transactions in a stolen account that are unidentified in the typical user activity, it creates an alarm);
2. it is very difficult for a cybercriminal to recognise what is a normal user behavior without producing an alert as the system is constructed from customised profiles.

To sum up, SIDS systems can only identify well-known intrusions whereas AIDS systems can detect zero-day attacks.

However, AIDS systems can result in a high false positive rate because anomalies may just be new normal activities rather than genuine intrusions.

3.3 Statistics-based techniques

A statistics-based IDS builds a distribution model for normal behaviour profile, then detects low probability events and flags them as potential intrusions.

Statistical AIDS systems essentially take into account the statistical metrics such as the median, mean, mode and standard deviation of packets: in other words, rather than inspecting data traffic, each packet is monitored.

Moreover, statistical AIDS are employed to identify any type of differences in the current behavior from normal behavior and are tipically based on one of the following models:

- univariate (i.e., the data has only one variable): this technique is used when a statistical normal profile is created for only one measure of behaviours in computer systems; univariate IDS look for abnormalities in each individual metric.
- multivariate (i.e., the data has more than one variable): it is based on relationships among two or more measures in order to understand the relationships between variables; this model would be valuable if experimental data show that better classification can be achieved from combinations of correlated measures rather than analysing them separately; however, the main challenge for multivariate statistical IDs is that it is difficult to estimate distributions for high-dimensional data;
- time series model: a time series is a series of observations made over a certain time interval; a new observation is abnormal if its probability of occurring at that time is too low (the feasibility of this technique was validated through simulated experiments).

4 Conclusions

In general, predictive models allow us to make more useful and less erroneous decisions: making important decisions without diligent consideration to uncertainties in the budgeting process can lead to unrealistic values, but forecasting with accuracy, on how much damage a successful security breach can cause, is a real challenge for risk managers, especially when multiple assets and associated threat exposure are considered [8].

In conclusion, using probabilistic simulation, simplifies the complexity of cost estimation processes. Indeed, the application of Monte Carlo Simulation to information security investment decisions, allows us to visualise different probabilistic outcomes in view of what might go wrong: given the best case, the worst case and the most likely case scenarios.

The Monte Carlo Method allows us to understand the outcome of scenarios and help in identifying unexpected pattern without necessarily exposing information assets to real threats.

It is expected that predictive models will help management making more effective decisions: if there is an effective understanding of what might go wrong, decision makers can utilise this probabilistic model to implement appropriate risk mitigation strategies and budget allocation for security investment, as discussed in Chapter 2.

Moreover, since cybercriminals are becoming really sophisticated and motivated, it becomes increasingly important for computer systems to be protected using advanced intrusion detection systems which are capable of detecting modern malwares [14], as discussed in Chapter 3.

5 References

[1] *What is risk management? — IBM*. URL: https://www.ibm.com/topics/risk-management.

[2] J. R. Conrad. "Analyzing the Risks of Information Security Investments with Monte-Carlo Simulations". In: (2005). URL: https://infosecon.net/workshop/pdf/13.pdf.

[3] *Introduction to Return on Security Investment — ENISA*. URL: https://www.enisa.europa.eu/publications/introduction-to-return-on-security-investment.

[4] National Bureau of Standards. *Guideline for The Analysis Local Area Network Security*. Tech. rep. Washington DC: U.S. Government Printing Office, 1994.

[5] *Cyber Security And Monte Carlo Simulation*. 2020. URL: https://sectara.com/news/cyber-security-and-monte-carlo-simulation/.

[6] K. J. Soo Hoo. "How Much is Enough? A Risk-Management Approach to Computer Security". PhD thesis. School of Engineering, Stanford University, Stanford, CA, 2000.

[7] D. Vose. "Monte-Carlo Risk Analysis Modelling". In: *CRC Press Inc.* In Vlasta Molak ed. Fundamentals of Risk Analysis and Risk Management (1997), pp. 57–78.

[8] T. Fagade, Maraslis K., and Tryfonas T. "Towards effective cybersecurity resource allocation: the Monte Carlo predictive modelling approach". In: (2017). URL: https://www.inderscienceonline.com/doi/abs/10.1504/IJCIS.2017.088235.

[9] *Risk Analysis using Monte Carlo Simulation*. 2016. URL: http://www.riskamp.com/files/Risk%20Analysis%20using%20Monte%20Carlo%20Simulation.pdf.

[10] *Ponemon Institute 2015 Cost of Data Breach Study: Global Analysis*. 2015. URL: https://www.ponemon.org/local/upload/file/2015%20Global%20CODB%20FINAL%203%20copy.pdf.

[11] *Kaspersky Lab (2015) Damage Control: The Cost of Security Breaches, IT Security Risks Special Report Series*. 2015. URL: https://media.kaspersky.com/pdf/it-risks-survey-report-cost-of-security-breaches.pdf.

[12] *MATLAB - MathWorks - MATLAB & Simulink*. URL: https://www.mathworks.com/products/matlab.html.

[13] *ModelRisk - Vose Software*. URL: https://www.vosesoftware.com/products/modelrisk/.

[14] A. Khraisat et al. "Survey of intrusion detection systems: techniques, datasets and challenges". In: (2019). URL: https://cybersecurity.springeropen.com/articles/10.1186/s42400-019-0038-7.

[15] C. Kreibich and J. Crowcroft. "Honeycomb: creating intrusion detection signatures using honeypots". In: *SIGCOMM Comput Commun Rev 34(1)* (2004), pp. 51–56.

[16] *Internet security threat report 2017*. 2017. URL: https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf.