# INFORMATION SECURITY INCIDENTS SIMULATION FOR RISK ANALYSIS IN THE NATIONAL AUTONOMOUS UNIVERSITY OF MEXICO

**Israel Andrade Canales**[(a)]

[(a)]National Autonomous University of Mexico

[(a)]iandradec@unam.mx

**ABSTRACT**

The National Autonomous University of Mexico (UNAM) offers different types of services to support academic activities. All of these services use valuable information for the achievement of their objectives and goals; consequently, information is one of the most important assets that the University has. However, thousands of security incidents affect these assets every year; for instance, in 2012 the university network suffered about 16,000 incidents provoked by botnets, spam and brute force attacks. Until now, this problem has been confronted by qualitative risk analysis methodologies in order to select counter-measures that mitigate these dangerous events. Nevertheless, these approaches lack either an optimization point of view or accurate results. Because the institution needs to treat risk not only precisely but also plausibly in financial and technical terms, this paper tries to shed light on a mixed model that combines simulation and linear optimization for the prediction and treatment of security incidents.

Keywords: information security, risk analysis, simulation, optimization

## 1. INTRODUCTION

Nowadays, information is a valuable asset that is used by people and organizations for decision making, communicating ideas, offering services and creating a variety of products. Therefore, Information Technologies (IT) have been developed for processing, storing and transmitting information in a practical manner. However, different risks affect information seriously; for instance, software bugs that generate vulnerabilities, and risky user habits that damage it.

The National Autonomous University of Mexico (UNAM) uses information technologies to offer different types of services that support academic activities, but thousands of security incidents affect these assets every year. For instance, in 2012 the university network suffered about 16,000 incidents provoked by botnets, spam and brute force attacks (UNAM-CERT 2012).

However, this kind of incidents frequently occurs in world-wide organizations; for example, the survey Information Security Branches (Price Waterhouse Coopers 2012) shows that nine of ten large enterprises in the UK reported an information security incident whose impact amounted to somewhere between 110 and 250 thousand pounds.

Until now, this problem has been dealt with risk analysis, i.e. the methodical use of information in order to identify and evaluate risk (ISO 2009). In this sense, there are two manners to assess risks: qualitatively and quantitatively (Gollman 2011, Buchanan 2011).

Qualitative risk methodologies use techniques such as manual inspections, staff interviews and information provided by experts in accordance to structured methods like OCTAVE (Caralli 2007). Although these techniques allow to carry out risk analysis in a coherent, repeatable and documented way, they may be subjective because they lack mathematical models that can give more accurate information to identify and analyze risks.

On the other hand, quantitative risk methodologies are model-based techniques that use mathematical tools like decision trees (Sahinoglu 2005), and simulation, such as those proposed by Winkelvos et al. (Winkelvos 2011). These methodologies give more precise information about risks. Even though these models are more precise than the qualitatively ones for risk analysis, most of them lack an optimization focus that helps to minimize risks for decision making and risk treatment.

Since UNAM needs to treat risk not only precisely but also plausibly in financial and technical terms, this paper tries to shed light on a quantitative risk analysis with two principal purposes: (1) to analyze risks with a model-based technique; and then, (2) to design a feasible risk treatment plan. Therefore, the proposed model combines simulation and linear optimization for the prediction and treatment of risks based on incident reports. First, the methodology used to analyze risk with the model proposed is described. Next, details about the model are given. Then, the results obtained are discussed. And finally, the conclusions that can be drawn from this research are presented.

## 2. METHODOLOGY

To carry out the information security risk analysis proposed in this paper, a combined model of simulation and optimization was proposed. The simulation model was built to perform a risk analysis in some scenarios of

interest. On the other hand, the optimization model was formulated to treat efficiently the simulated risks in order to minimize the negative effects of information security incidents.

Figure 1 shows the connection between the simulation and the optimization model; the arrows represent the variables used to formulate each model, and the variables used to link both models; for example, the predicted incidents, the simulation output, are used as the objective function in the optimization model.

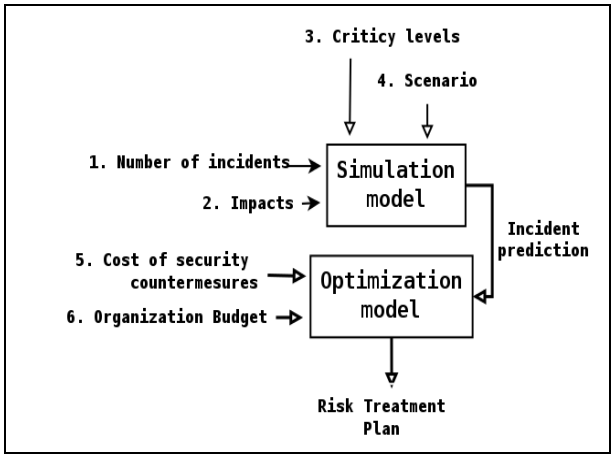


Figure 1: The Model Proposed

### 2.1. Simulation Model

The simulation model was used to analyze risks. This tool uses two random variables that represent the number of incidents and their impact. These variables were obtained through the data analysis of the information security incident reports of a university office.

#### 2.1.1. Variables

The first variable (number of incidents) expresses the frequency of four types of incidents classified according to their sources: external entities (H), e.g. hackers; configuration errors (C); policy violations (P); and lack maintenance (M). However, because some organization activities affect the probabilities of the security incidents, four main scenarios were analyzed. Table 1 shows the probabilities for each incident in the four considered scenarios: External Projects (1), "normal" days (2), auditing procedures (3) and public events (4).

For example, when the academic office organizes a public event such as a congress, it is more likely to suffer an external attack (0.42) than when the organization has a normal period of activities. On the other hand, when the university office administers external projects, it is more common for configuration errors in the equipment to occur than when the organization is running a public event.

Table 1: Incident Probability Table

| | Scenario | | | |
|---|---|---|---|---|
| **Incident** | **1** | **2** | **3** | **4** |
| **H** | 0.10 | 0.01 | 0.11 | 0.42 |
| **C** | 0.36 | 0.97 | 0.77 | 0.14 |
| **M** | 0.36 | 0.01 | 0.01 | 0.42 |
| **P** | 0.18 | 0.01 | 0.11 | 0.02 |
| ***Min*** | 2 | 0 | 2 | 3 |
| ***Max*** | 4 | 1 | 6 | 4 |

The second variable (Impacts) corresponds to the number of idle hours caused by each incident. This information is the impact of each incident on the organization. Other kind of impacts may be: the cost of each incident, the damage to reputation, etc.

To estimate the productivity hours lost, the incidents reported by the academic institution in 2011 were analyzed. A security incidents histogram was built and adjusted by a probability distribution function; for instance, Figure 2 shows how the histogram of external attacks incidents (H) was adjusted to a beta probability density function; the beta function expresses that the organization loses between 0.5 to 20 productivity hours in an external attack, but the most likely ranges are between 0.5 to 7, and between 14 to 20 hours. Table 2 presents the parameters of the probability density function for each incident.

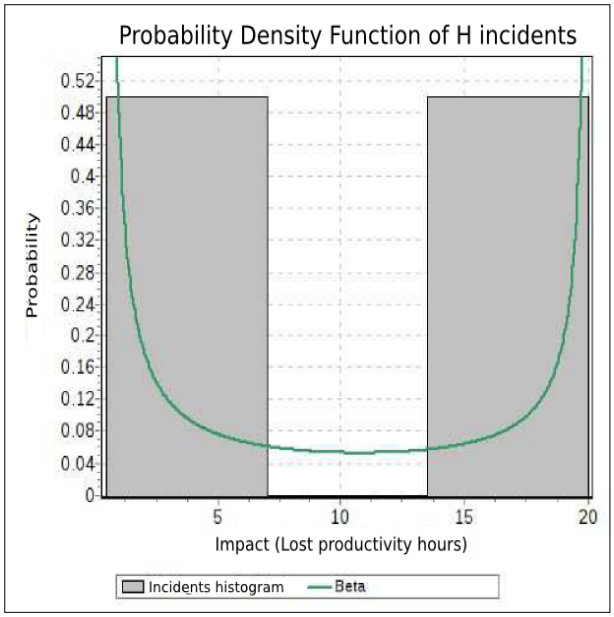


Figure 2: Probability Density Function of External Attacks Incidents (H)

Table 2: Impact Functions

| Incident | Function | Parameters |
|---|---|---|
| H | Beta | α = 0.063, β = 0.159<br>a = 0, b = 20 |
| C | Beta | α = 0.071, β = 0.337<br>a = 0.5, b = 24 |
| P | Uniform | Min = 0.3<br>max = 4.09 |
| M | Uniform | Min = 0<br>max = 2.62 |

#### 2.1.2. Constants

On the other hand, two constants were used to run the simulation: a critical level of incidents and the experiment scenario. The first constant indicates the ranges of hours for three level categories: high, medium and low. Table 3 shows the ranges defined in the simulation.

The second constant indicates a series of activities that the organization may perform in the analyzed period. This information indicates which probability function to use in the model.

Table 3: Critical Levels of Impact

| Levels | Low | Medium | High |
|---|---|---|---|
| Range (hrs) | 0-5 | 5-10 | 10+ |

#### 2.1.3. Simulation

Finally, a Monte-Carlo technique was used as a numeric method to relate the behavior of all the variables and constants in order to run the simulation. The Monte-Carlo simulation was programmed in the statistical language *R* (R Core Team 2012). Hundreds of simulation cycles were necessary to obtain stable results. These results acted as input data of the optimization model.

### 2.2. Optimization Model

The optimization model was formulated in order perform the risk treatment plan after the risk analysis; in other words, it was used to obtain a combination of activities that minimize the information security risks. This model takes into account that each activity has a cost in financial and work time terms. In this sense, it allowed us to obtain a security treatment plan, which is totally feasible for the organization.

The optimization model was derived from an Integer Binary Program which specifies a list of variables that represents the activities to be implemented, *i.e.* the countermeasures such as firewalls or information policies designed by the information security team; this approach has been suggested by Caulkins J. (Caulkins 2007) and Garvey (Garvey 2009) . The formal model is described as follows.

$$Min: \sum_{j=1}^{m}\sum_{i=1}^{n} -r_{ij}x_{ij} \qquad (1)$$

*s.t*

$$\sum_{j=1}^{m}\sum_{i=1}^{n} c_{ij}x_{ij} \leq B$$
$$x \in \{0,1\}, c \geq 0, B \geq 0 \quad i,j \in \mathbb{N} \qquad (2)$$

The objective function (Equation 1) represents the risks ***r*** to be mitigated by the activity ***x*** in order to minimize the total risk. In the equation, ***n*** is the number of levels registered in Table 3, and ***m*** represents the number of different kinds of incidents, *i.e*. H, C, M and P.

Equation 2 expresses the restriction of cost ***c*** of each activity ***x***, which must be less or equal to the organization's budget ***B***. Table 4 indicates those restrictions and the budget estimated for the academic office.

Once these equations have been solved, the results can be used as a decision-making aid to establish a Risk Treatment Plan that the organization can follow in order to obtain a reasonable state of security.

Table 4: Restrictions Used in the Optimization Model

| Restrictions | | |
|---|---|---|
| **Countermeasure** | **Time of Implementation** | **Cost** |
| **1** | 1 | 1 |
| **2** | 5 | 3 |
| **3** | 10 | 10 |
| 4 | 1 | 5 |
| 5 | 5 | 10 |
| 6 | 15 | 50 |
| 7 | 2 | 10 |
| 8 | 5 | 50 |
| 9 | 10 | 100 |
| 10 | 1 | 10 |
| 11 | 5 | 20 |
| 12 | 10 | 50 |
| Budget | 32 | 150 |

## 3. RESULTS AND DISCUSSION

In this section, we show and discuss the results of the risk analysis. First we present the simulation results; then, we present a brief comparison between the simulation results and the information reported in 2012 by UNAM. Next, we show how the results of the simulation were transformed to formulate the optimization model. And finally, we report the results of the optimization model that represents a feasible risk treatment plan.

### 3.1. Simulation results

The simulation reported 21 incidents according to the organization's activities report in 2012. These incidents were distributed as follows (Table 5), 11 configuration errors; 2 external attacks; 3 lack-of-maintenance related incidents; and 5 policy violation incidents. To reach stable results, a hundred simulations were run. Figure 3

shows a graphic with the number of incidents obtained in these tests.

Table 5: Results of the Simulation

| Simulation results | | |
|---|---|---|
| **Incident** | **Number of Incidents** | **Impact (Idle hours)** |
| **C** | 11 | 49.4 |
| **H** | 2 | 10.4 |
| **M** | 3 | 4 |
| P | 5 | 10 |
| **Sum** | **21** | **73.8** |

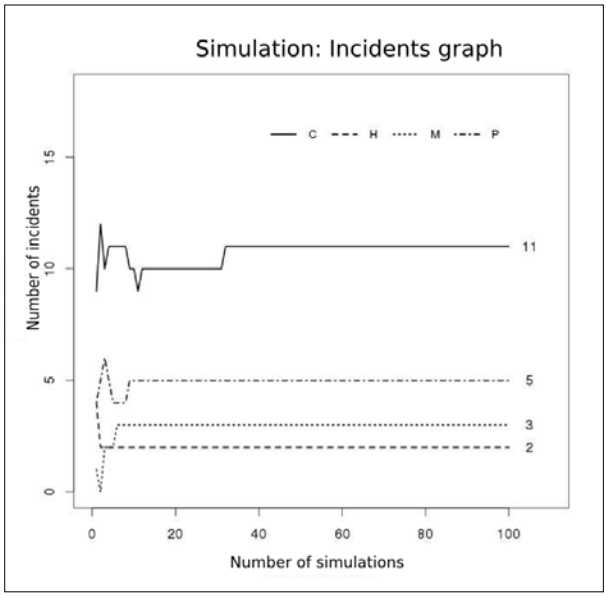


Figure 3: Number of Incidents Obtained by the Simulation Tests

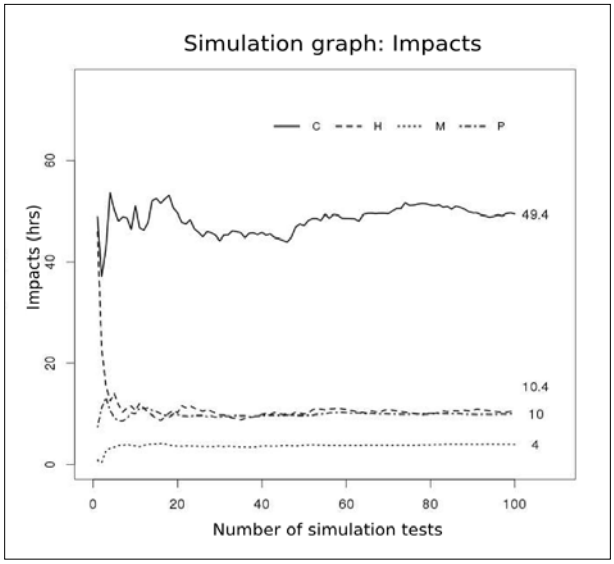


Figure 4: Productive Hours Lost Due to Information Security Incidents Obtained by the Simulation

In addition, the simulation reported 73.8 productive hours lost due to incidents. Figure 4 shows how the impact of each kind of incident, a random variable in the simulation, fluctuated at the beginning of the tests, and then leveled out at the end of the simulations.

### 3.2. Comparison between simulated and reported results

The simulation shows results consistent with the number and type of incidents reported in 2012 by the academic institution (Table 6). Figure 5 highlights the comparison among the incidents simulated and the incidents reported in 2012. As can be seen, the number of incidents was very similar to the actual number reported by the institution.

Table 6: Results Reported by the Organization in 2012

| Simulation results | | |
|---|---|---|
| **Incident** | **Number of incidents** | **Impact (Idle hours)** |
| **C** | 12 | 52.2 |
| **H** | 3 | 2 |
| **M** | 2 | 6 |
| P | 6 | 9.9 |
| **Sum** | **23** | **70.1** |

Furthermore, Figure 6 exhibits a comparison of impacts per incident between the simulation and the incidents reported in 2012. It is important to notice that incidents occurred due to external attacks (H) were considerably fewer than the ones reported in 2012. The probability function does not imitate the real-system variable because external attacks are significantly random in the problem analyzed.

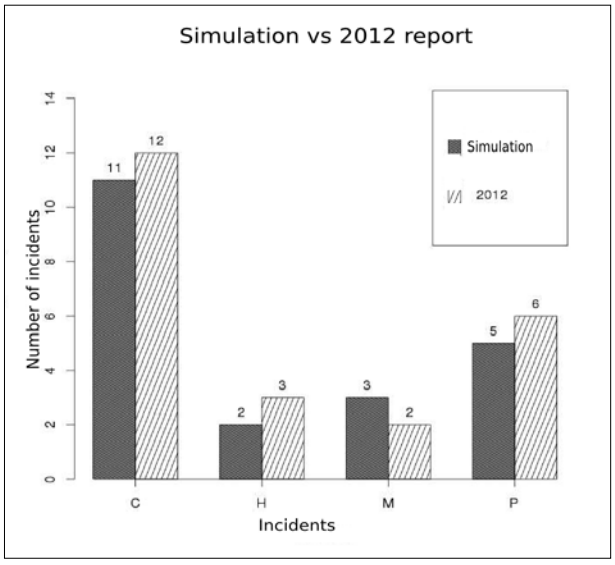


Figure 5: Comparison between Incidents Obtained by Simulation and Incidents Obtained by 2012 Reports

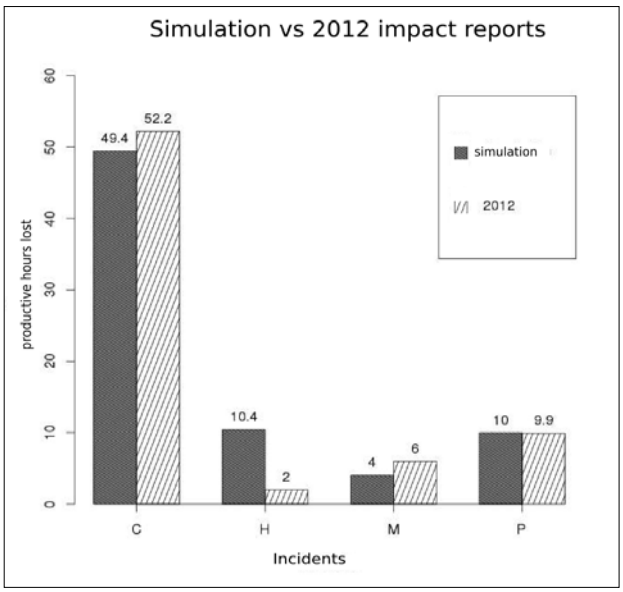


Figure 6: Comparison Between Incident Impacts Obtained by Simulation and Incidents Impacts Obtained by 2012 Reports

### 3.3. Simulation results as optimization model input data

Because the second objective of this paper was to formulate a risk treatment plan based on the simulation results, we grouped the incidents obtained according to the impacts critical levels (Table 3). These groups were used to establish the objective function (Equation 1) described in Section 2.2

Table 7 highlights the groups of incidents derived by the simulation and the impact critical levels established by the organization; moreover, Equation 3 indicates how this information is used to formulate the objective function.

Table 7: Incident Groups According to the Impact Critical Level

| | Incidents | | | |
|---|---|---|---|---|
| **Critical level** | **C** | **H** | **M** | **P** |
| **Low** | 6.3 | 1.1 | 4 | 10 |
| **Medium** | 3.7 | 0.8 | 0 | 0 |
| **High** | 39.4 | 8.5 | 0 | 0 |
| Total | 49.4 | 10.4 | 4 | 10 |

On the other hand, the restrictions shown in Equation 4 and Equation 5 were derived from the information described in Table 4.

$$Min: -6.3x_1 - 3.7x_2 - 39.4x_3 - 1.1x_4 - 0.8x_5 - 8.5x_6 - 4x_7 - 10x_{10} \quad (3)$$

*subject to*

$$x_1 + 5x_2 + 10x_3 + x_4 + 5x_5 + 15x_6 + 2x_7 + 5x_8 + 10x_9 + \ldots + x_{10} + 5x_{11} + 10x_{12} \leq 32 \quad (4)$$

$$x_1 + 3x_2 + 10x_3 + 5x_4 + 10x_5 + 50x_6 + 10x_7 + 50x_8 + \ldots + 100_9 + 10x_{10} + 20x_{11} + 50x_{12} \leq 150 \quad (5)$$

### 3.4. Results of the optimization model

The integer program, used to obtain a risk treatment plan, was solved through the *lpsolve software* (Berkelaar 2004). This software could be integrated with the simulation code to automate either the simulation tests or the optimization model.

The optimization model gave a risk treatment plan that allows to mitigate 69 of the 74 productive hours lost. Table 8 shows what activities should be implemented in order to reach these results. However, because both models are formulated with random data, the results should be used as a decision-making aid tool.

Table 8: Risk Treatment Plan Obtained from the Optimization Model

| **Risk treatment plan** | | | | |
|---|---|---|---|---|
| **Activity** | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
| **Plan** | 1 | 0 | 1 | 1 |
| **Activity** | $x_5$ | $x_6$ | $x_7$ | $x_8$ |
| **Plan** | 0 | 1 | 1 | 0 |
| **Activity** | $x_9$ | $x_{10}$ | $x_{11}$ | $x_{12}$ |
| **plan** | 0 | 1 | 0 | 0 |

## 4. CONCLUSIONS

Information security is a relative new field that can be explored through models like simulation and analytical models. In the approximation presented in this paper, the combination of the two models allowed both to analyze information security risks and treat them efficiently. This can be useful in an organization that deals with some restrictions on security investment.

On the other hand, the comparison between the results obtained and the information reported in the case studied suggests that systematic incidents such as policy violations and human errors caused by ambiguous procedures can be estimated successfully. Nevertheless, the same validation highlighted that some random events, like hacker attacks, are less precisely estimated. However, the two main objectives presented in this paper, not only to perform a risk analysis, but also to treat the impacts of incidents, were reached.

Therefore, this case of study showed that simulation and linear optimization are a powerful technique for a better decision-making in information security.

## ACKNOWLEDGMENTS

This research was supported by UNAM-PAPIIT grant IN116012, and the author would like to thank to Dra. Idalia Flores de la Mota and Dr. Juan Manuel Estrada Medina for their valuable and constructive suggestions during the planning and development of this research work, to Martha Mora Barreto for the editing and proofreading help, and to "The Computer Emergency

Response Team (UNAM-CERT)" for making available the data used in the case of study.

## REFERENCES

Berkelaar, M., Eikland, K. and Notebaert, P., 2004. *Open source (Mixed-Integer) Linear Programming system*. Poole, GNU lpsolve. Available from: http://lpsolve.sourceforge.net [accessed 1 July 2012]

Buchanan, W. J., 2011. *Introduction to Security and Network Forensics*. 1st ed. The U.S.: CRC Press.

Callari, R.A., Stevens, J.F., Young, L.R. and Wilson W.R., 2007. *Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process*. SEI, Carnegie Mellon University. Available from: http://*www.cert.org/archive/pdf/07tr012.pdf* [accessed 1 July 2012]

Caulkins, J. P., Hough, E. D., Mead, N. R., Osman, H., 2007. Optimizing Investments in Security Countermeasures: A Practical Tool for Fixed Budgets. *IEEE Security and Privacy*, 5 (5), 57-60.

Garvey, P. R., 2009. *Analytical methods for risk management: a systems engineering perspective*. 1st ed. Massachusetts, U.S.A: CRC Press.

Gollman, D., 2011. *Computer Security*. 3th ed. The U.K.: Wiley and sons.

Price Waterhouse Coopers, 2012. *Information Security Branches 2012*. Price Weterhouse Coopers. Available from: http://www.pwc.co.uk/en_UK/uk/assets/pdf/olpap p/uk-information-security-breaches-survey-technical-report.pdf [accessed 1 November 2012]

R Core Team, 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/ [accessed 1 July 2012]

Sahinoglu M., 2005. Security Meter: A Practical Decision-Tree Model to Quantify Risk. *IEEE Security and Privacy*, 3 (3), 18-24.

UNAM-CERT, 2012. Estadísticas. Universidad Nacional Autónoma de México. Available from: http://www.cert.unam.mx/estadisticas.dsc [accessed 1 November 2012]

Winkelvos, T., Rudolph, C. and Repp, J., 2011. A property based security risk analysis through weighted simulation. *Information Security South Africa (ISSA)*, 1 (8), 15–17.

## ISRAEL ANDRADE CANALES

Israel studied Computer Engineering at Facultad de Estudios Superiores Aragón, UNAM. He worked 3 years at the Computer Emergency Response Team (UNAM-CERT) of the National Autonomous University of Mexico in the information security auditing and risk analysis area. He is currently studying a master in Operations Research, and his line of research is optimization in information security.