A Probabilistic Model for Evaluating the Operational Cost of PKI-based Financial Transactions

Agapios Platis, Costas Lambrinoudakis and Assimakis Leros

Department of Information and Communication Systems Engineering,
University of the Aegean, Karlovasi,
83200 Samos, Greece
{platis, clam, aleros}@aegean.gr

Abstract. The use of PKI in large scale environments suffers some inherent problems concerning the options to adopt for the optimal cost-centered operation of the system. In this paper a Markov based probability model has been applied and a performability indicator has been introduced for assisting the evaluation of the operational cost of the system in a decision support process. Considering the unavailability of the Certification Authority server, three different strategies have been evaluated for determining the optimal one.

1 Introduction

During the last decade the Public Key Infrastructure (PKI) has been widely used for the provision of security services, especially in application domains like e-commerce, financial transactions, e-health etc. A central role of a PKI is that of the Certification Authority (CA) that mainly deals with issuing key pairs (a private and a public key) for customers[14], or/and simply register a public key supplied by the registered entity. The private key must remain secret, under the control of its owner, while the public key must become available to anyone wishing to have some type of transactions with the owner of the private key. At this point the concept of a *digital certificate* is introduced, linking the public key of a customer with her/his identity. This linkage is certified and digitally signed by the CA and, therefore, trusted by any two parties utilizing the PKI for performing a transaction. However, any digital certificate has an expiration date and frequently unexpired certificates must, for some unexpected reason (the private key has been compromised, user credentials have changed etc), be invalidated (*revoked*)[5, 6]. In either case the certificate must not be used. It is therefore clear that a mechanism supporting an entity to confirm the validity of some other entity's certificate must exist.

Periodically-issued *Certificate Revocation Lists* (CRLs) are one common approach to revoking certificates; each such list specifies what unexpired certificates have been revoked, and when the next CRL will be issued. The issuing CA signs the CRL and someone wishing to check the validity of a certificate must download the "most recent" CRL from the CA. However, the "recency" of the CRL is a parameter that should reflect the requirements of each specific customer, implying that someone could be satisfied with weekly updated CRLs while someone else could require at most day-old evidence[15].

In this paper we present a probabilistic model that can be employed for evaluating the "cost" of different strategies as far as the certificate validation checks are concerned. Specifically, depending on the transaction characteristics, an entity may choose to trust the provided certificate and proceed with the transaction without confirming its validity –in this case the specific entity accepts the risk of a *security incident* to happen due to invalidated or/and expired certificates. Alternatively, for a different transaction the same entity may set as a prerequisite for carrying out the transaction the confirmation of the certificate's validity through the CA's CRL. However, the desired confirmation may not be possible due to *CRL (CA) unavailability*, a term used in this paper for describing one or more of the following cases:

- The CA server (and thus the CRL) is not accessible.
- The CRL is not accessible / available.
- The "recency" of the CRL does not fulfill the user requirements.

It is therefore not possible to proceed with the transaction *(failed transaction)* if the CRL-CA is unavailable (irrespective of the reason causing the unavailability) for a period exceeding a specified threshold.

In order to evaluate the different strategies, a Markov based Performability model has been used. Performability modeling or Markov Reward Models (MRM) have been initiated by Beaudry [1] and Meyer [8] and successfully used for evaluating the performance of gracefully degrading computer systems, cf.[2,3,4,10,17] and electrical systems, cf.[12], but also for evaluating the quality of Internet services, cf.[18,19], or website reorganization, cf.[13]. The major advantage of such models is that they combine reliability and performance measures, including to a greater extent cost related measures and external environmental parameters and thus allowing a more detailed modeling. Indeed, for a high availability system, a failure during peak hours has a much greater impact on users than a failure occurring when the system is not extensively used. Where classical availability evaluations cannot differentiate these events, the performability model can perceive and evaluate each occurrence individually. In the same way a classic Markov model will give important information about the probabilities of a security incident and a failed transaction. A security incident is, of course, a highly undesirable event, normally causing much more serious consequences than those caused by a failed transaction. On the other hand, a large amount of failed transactions may have a higher cumulative cost than a single or a few security incidents. By utilizing a performability model the above-mentioned parameters can be taken into account.

2 Scenario Description

This section describes the PKI architecture and the operational scenarios that have been chosen for modeling (Section 3). It is evident that different implementations, either in terms of the certificate revocation mechanisms or/and in terms of supported functionality, would cause differentiations in the model presented in this paper. However, the scenario that has been adopted can be characterized as *representative* of a typical general-purpose PKI implementation capable of supporting a wide range of applications.

Specifically, the existence of a Certification Authority (CA) is assumed, which in addition to the task of generating and distributing key pairs to customers it is responsible for maintaining a certificate revocation list (CRL) in order to allow anyone interested to check, prior to a transaction, the validity of someone else's certificate. For the purposes of the current paper, it is assumed that the CRL can only be checked if the CA server is available. This assumption is based on the fact that even if the customers maintain local copies of the CRL, prior to a high valued financial transaction they will always request an up-to-date CRL from the CA. Based on the above, a typical operational scenario is the following:

1. A financial institution X supports on-line transactions (money orders, investments etc) for users (customers) that are registered with the specific service and have obtained a pair of valid keys from a certification authority (CA).
2. A party Y (customer) wishing to perform a financial transaction, through the institution X, submits the appropriate request, digitally signed, to X. For the digital signature X is using her/his private key.
3. The institution X, before serving the request submitted by Y, must either assume that the public key of customer Y is valid or must verify its validity through the CRL maintained by the CA. Therefore the alternative actions of X are the following:
   - a) If the amount of a transaction is below some threshold value F, X decides not to check the validity of Y's certificate and proceeds serving the transaction
   - b) If the amount of the transaction exceeds the threshold value F, X must check the validity of Y's certificate by accessing the CA's CRL list.
4. For transactions that the validity of Y's certificate was required, the following alternative paths are possible:
   - a) The CRL maintained by the CA server is accessible and thus the validity check can be performed.
   - b) The CRL is, for some reason, unavailable. If this is the case then the institution X waits for a predetermined period of time for the CRL to become available and then proceed as in 4a.
   - c) If the CRL is unavailable (case 4b) and the waiting time exceeds a threshold, the institution X cancels the transaction (*failed transaction*).
5. For transactions that the validity of Y's certificate was not checked (case 3a), there is a possibility for a *security incident* to occur.
6. Even for transactions that the validity of Y's certificate has been checked, there is a possibility for a *security incident* to occur (although with significantly smaller possibility than that in case 5).

3 PKI Markov modeling

3.1 State transition diagram

The process is described as follows: transaction requests addressed to the financial institution X, arrive at a rate λ_1 which is modeled by a Poisson process, hence the inter-arrival times are exponentially distributed, β is the probability to have a transaction with an amount exceeding the threshold value F and therefore imposing the need to check the customer's certificate, γ_1 is the unavailability probability of CA's server (and thus unavailability probability of CRL's) and in this case the system waits till the server becomes available (the server restoration rate is μ_{rest}). If the waiting time exceeds a threshold value then the transaction fails with a rate μ_2 (which is the inverse of the mean delay of a failed transaction). The service rate is μ_1 and the security incident rate is μ. A security incident may occur if no access to the revocation list has been achieved with a rate λ_3, however a security incident can also occur even if the revocation list has been accessed but with a less important hazard rate λ_2.

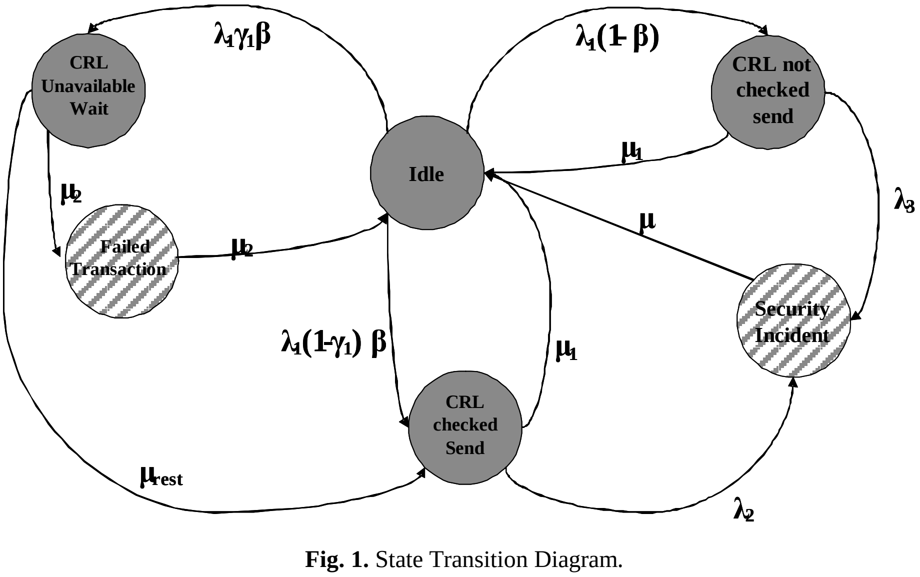

Fig. 1. State Transition Diagram.

Let X_n be the Markov Chain modeling the operational behavior of the system with state space {Idle, CRL-U, CRL-C, CRL-NC, FT, SI} and transition rates matrix A.

Table 1. System States

Description of the System States		
Idle	Idle	System is idle, waiting for requests
CRL-U	CRL Unavailable - Wait	System is waiting for accessing the CRL
CRL-C	CRL checked - Send	The request is processed after checking the CRL
CRL-NC	CRL not checked - Send	The request is processed without prior check of the CRL
FT	Failed Transaction	The transaction has failed due to the fact that the CRL stays unavailable for a time period exceeding a threshold value
SI	Security Incident	The transaction data have been altered

Since the chain is irreducible and aperiodic, then the chain is ergodic and has a unique steady state probability distribution **π**.

The computation of the steady state probability distribution is obtained by solving a linear system **π.A**=**0** with the additional condition **π.1**=1, where **1** is a 6-dimensional column vector containing ones and **π** the 6-dimensional row vector containing the steady state probabilities of the system (steady state probability distribution), cf.[9].

The resolution of the previous system gives the following results concerning the steady state probabilities of the failed transaction and the security incident states.

$$\pi_{FT} = \frac{\dfrac{\lambda_1\gamma_1\beta}{\mu_{rest}+\mu_2}}{1+\dfrac{2\lambda_1\gamma_1\beta}{\mu_{rest}+\mu_2}+\dfrac{\lambda_1\beta}{\mu_1+\lambda_2}\left(1-\gamma_1+\dfrac{\mu_{rest}\gamma_1}{\mu_{rest}+\mu_2}\right)\left(1+\dfrac{\lambda_2}{\mu}\right)+\dfrac{\lambda_1(1\text{-}\beta)}{\mu_1+\lambda_3}\left(1+\dfrac{\lambda_3}{\mu}\right)} \qquad \textbf{(1)}$$

and

$$\pi_{SI} = \frac{\dfrac{\lambda_1}{\mu}\left[\dfrac{\lambda_2\beta}{\mu_1+\lambda_2}\left(1-\gamma_1+\dfrac{\mu_{rest}\gamma_1}{\mu_{rest}+\mu_2}\right)+\dfrac{\lambda_3(1\text{-}\beta)}{\mu_1+\lambda_3}\right]}{1+\dfrac{2\lambda_1\gamma_1\beta}{\mu_{rest}+\mu_2}+\dfrac{\lambda_1\beta}{\mu_1+\lambda_2}\left(1-\gamma_1+\dfrac{\mu_{rest}\gamma_1}{\mu_{rest}+\mu_2}\right)\left(1+\dfrac{\lambda_2}{\mu}\right)+\dfrac{\lambda_1(1\text{-}\beta)}{\mu_1+\lambda_3}\left(1+\dfrac{\lambda_3}{\mu}\right)} \qquad \textbf{(2)}$$

The above equations have been evaluated using the empirical hazard rates and probabilities listed in Table 2 below. It should be mentioned that empirical data have been used in order to demonstrate the applicability of the model.

Table 2. Hazard Rates and Probabilities of the System Parameters

Hazard Rates and Probabilities		
λ_1	Transaction request arrival rate	50 h^{-1}
λ_2	Security incident rate given that the CRL has been checked	0,000001 h^{-1}
λ_3	Security incident rate given that the CRL has not been checked	0,001 h^{-1}
β	Probability of transaction exceeding threshold F	Variable 0; 1; 0,4
γ_1	CRL's unavailability probability	0,001
μ_1	Service rate	500 h^{-1}
μ_2	1/ Mean delay of a failed transaction	100 h^{-1}
μ_{rest}	Restoration rate for the CA-customer link	1h^{-1}
μ	1/ Mean duration of a security incident	500 h^{-1}

It should be stressed at this point that the probability of a security incident to occur when the validity of the customer's certificate has been checked –CRL has been accessed successfully-- (λ_2), has been assumed to be much smaller than the respective probability of a security incident when the financial transaction is performed without prior validation (λ_3). Furthermore, based on existing statistical information, the CA server is expected to be unavailable, in average, one time every one thousand transactions (γ_1).

The calculation of the probability of *Failed Transactions* and *Security Incidents* has been performed with three different values of (β), thus simulating the operational scenarios described in section 2. Specifically:

- **Strategy 1**: β=1, validation of customer's certificate is required prior to any transaction (CRL must be checked --- section 2, case 3b with a threshold value, F, for the amount of the transaction, set to zero).
- **Strategy 2**: β=0, all transactions are served without prior validation of customer's certificate (CRL is not checked --- section 2, case 3a with a threshold value, F, for the amount of the transaction, set to an extremely large value).
- **Strategy 3**: β=0.4, a combination of strategies 1 and 2, expecting 40% of the requested transactions to require validation of the customer's certificate (implying that the amount of the transaction exceeds the threshold value F), while the remaining 60% to be served without prior validation (implying that the amount of the transaction is below the threshold value F).

The calculated probabilities for Failed Transactions and Security Incidents, for each strategy, are depicted in Fig. 2 and 3 respectively.

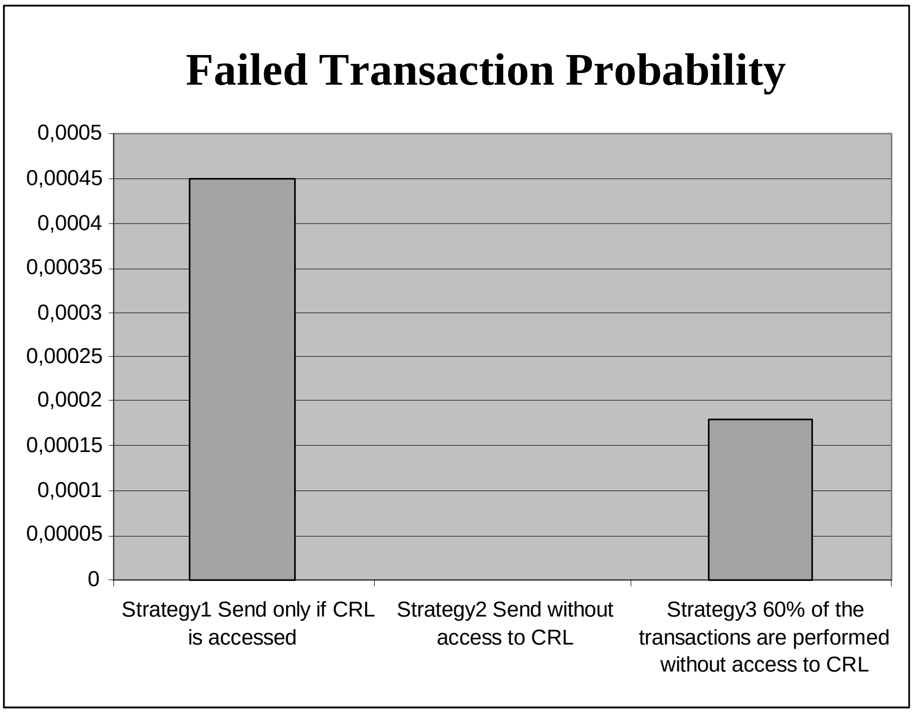

Fig. 2. Failed Transaction Probability for the three strategies.

It can be noticed that for strategy 1 the probability of security incidents to occur is extremely small although a large number of transactions may fail if the CA server is unavailable. On the other hand if the transactions are performed without prior validation of the customer's certificate (strategy 2) there are no failed transactions but the possibility of a security incident increases significantly. In the case of strategy 3, according which the decision on whether the CRL will be accessed or not depends on the amount of the transaction and the threshold value F set by the financial institution, it is evident that it is possible to face both failed transactions and security incidents. However the probability of failed transactions is much less that the respective probability for strategy 1, while the probability of security incidents is less than the respective probability for strategy 2.

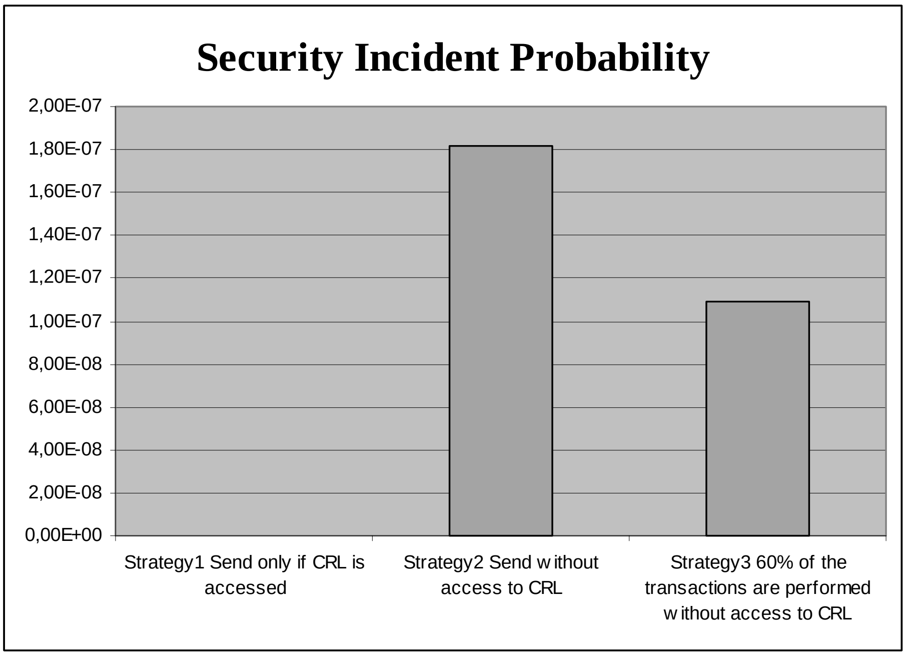

Fig. 3. Security Incident Probability as a Function of the Operational Strategy Adopted.

4 Expected Cost due to Security Incidents and Failed Transactions (ECSIFT indicator)

The steady state probabilities allow the evaluation of two undesirable events: the event of a *failed transaction* due to the unavailability of the CA server for verifying certificates' validity, which is a frequent event, and the event of a *security incident* which is a more exceptional event. On the other hand a failed transaction has a less significant cost compared to a security incident with a generalized impact.

The Markov modeling allows the evaluation of the probability of these events, although a more complete model taking into account additional parameters such as the derivation of the incident cost would be more appropriate for the evaluation of the operational safety of a PKI-based application.

For this purpose, the following probabilistic indicator is defined: The Expected Cost due to Security Incidents and Failed Transactions (ECSIFT) per unit of time.

If C_{SI} is the cost of a security incident and C_{FT} the cost of a failed transaction, the total probabilistic cost at time n is given as follows:

$$\mathrm{C_n} = \mathrm{C_{SI}}\, 1_{\{\mathrm{X_n}=SI\}} + \mathrm{C_{LT}}\, 1_{\{\mathrm{X_n}=FT\}} \qquad (3)$$

where X_n is the Markov chain modeling the system and $1_{\{.\}}$ the indicator random variable.

In steady state, the probability of this event is given by $\lim_{n\to\infty} E[C_n]$, hence the expected cost is given as follows:

$$\mathrm{ECSIFT} = C_{SI}\pi_{SI} + C_{FT}\ \pi_{FT} \qquad (4)$$

Note that this probabilistic indicator can be derived from the general formulation of the performability indicator in [11].

The indicator ECSIFT (Expected Cost due to Security Incidents and Failed Transactions) for all three strategies is depicted in Fig. 4. For each strategy the ECSIFT has been calculated for two values of γ1 (probability of CA server's unavailability), namely: γ1=0.001 and γ1=0.002.

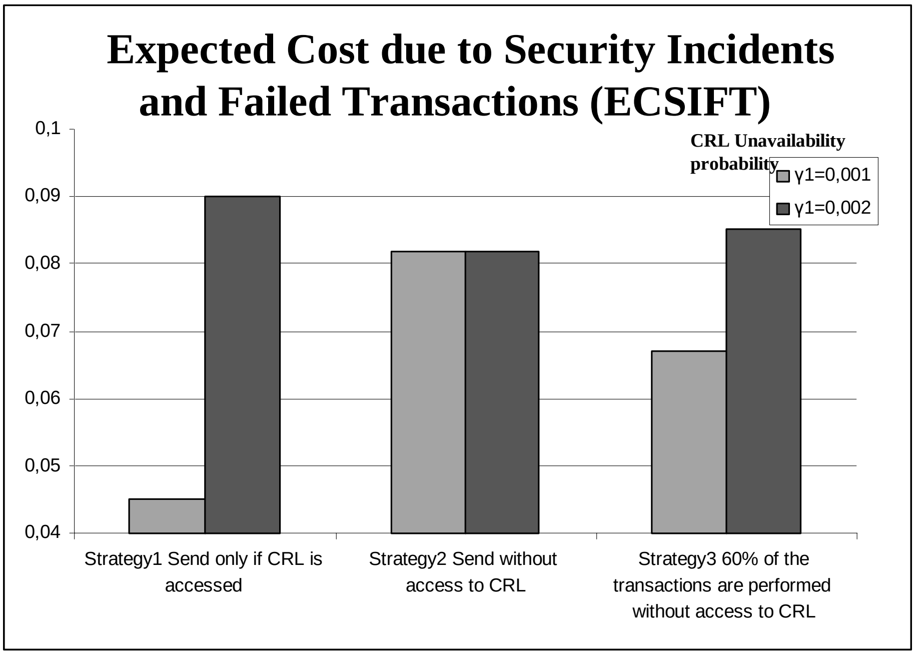

Fig. 4. ECSIFT Indicator for a CA Server's Unavailability Probability γ1=0,001 and γ1=0,002.

5 Discussion

Trying to comment the expected cost due to security incidents and failed transactions (depicted in Fig.4), it can be noticed that if the probability of the CA server's unavailability is one time every 1000 transactions (γ_1=0.001) then strategy 1 seems to be the best approach while strategy 2 the most expensive one. Taking into account the fact that the cost of a security incident is assumed to be orders of magnitude bigger than that of a failed transaction, the ECSIFT figures for γ=0.001 are, somehow, the expected ones, since with strategy 1 there is an extremely low number of security incidents although a considerable number of transactions may fail due to unavailability of the CA server. With strategy 2 the situation is exactly the opposite since there are no failed transactions but there is a much bigger probability for security incidents.

However, as demonstrated in Fig. 4, a small differentiation in the probability of the CA server's unavailability may significantly alter the entire picture. More specifically if $\gamma 1$ becomes 0.002, implying that the number of failed transactions --as a result of the inability of the interested party to access the CRL-- will increase, then strategy 2 becomes the best choice, while strategy 1 the most expensive. This is because strategy 2 is not affected by the unavailability of the CA server as opposed to strategy 1 that will now face a significantly bigger number of failed transactions.

The main objective of this paper is to demonstrate the applicability and the practical advantages of the proposed probabilistic Markov model for PKI-based applications, which can be employed, even in a real-time fashion. Indeed, parameters such as the ratio of transactions exceeding a threshold cost, or the CRL unavailability can be directly obtained from a database tracing these specific time-varying parameters. The model is therefore supplied with updated input parameters allowing a dynamic decision-aided process concerning adoption of the less expensive operational strategy.

If more accurate results are required then more elaborated probabilistic models can be used. For instance, a strict assumption for using the Markov model is that the inter-arrival times of transaction requests are exponentially distributed, which however is not always the case for the sojourn time in the state where the CRL is unavailable. In such cases a semi-Markov model, cf.[7] is more suitable, allowing the consideration of any distribution for the sojourn time in the states.

Additionally, if a periodic behavior of some parameters is observed a non-homogeneous Markov model can be utilized. For instance, periodically, the number of requests for transactions that involve high amounts may reach a peak during some specific hours of a day; similarly the CRL unavailability probability may be variable during a day due to an overloaded server during these specific hours. In this case the ECSIFT indicator can be modeled with a cyclic non-homogeneous Markov chain, cf.[11], permitting a more accurate cost evaluation.

Some difficulties however subsist: the evaluation of some specific parameters such as the different costs:

- The cost of a failed transaction (C_{FT})
- The cost of a security incident (C_{SI})

Indeed, these costs are not easy to compute mainly because they depend on the application's environment but also due to lack of information. In this paper empirical data have been used in order to highlight the interest of the methodology, although *real* data would give more accurate results.

6 Conclusions and future work

The use of a PKI in large scale environments highlighted some inherent problems concerning the options to adopt for the optimal cost-centered operation of the system.

A first approach is to always validate the digital certificates presented by the customers, in order to avoid, as much as possible, a *security incident*, even if this option may result in several *failed transactions* due to unavailability of the CA and thus of the CRL. A second option is to bypass the certificate validation step, aiming to minimize the number of *failed transactions* despite the higher probability of a *security incident* to occur. A third option is to access the CRL only for transaction amounts exceeding a certain threshold amount. The need to choose the optimal cost-centered solution led us to study the use of a Markov based probability model and specifically the introduction of a performability indicator, namely the ECSIFT, in order to evaluate the cost of each option in a decision support process. The results, despite the fact that there were obtained from empirical data, illustrated the need to use such models in environments with multiple operational and reliability parameters.

Our future work will be focused in identifying the cost-centered optimal transaction amount. For transactions exceeding this amount, the CRL check will be mandatory while in the remaining cases the CRL check will be optional.

References

1. Beaudry, M.: Performance related reliability for computer systems. IEEE Transactions on Computers, C-27, (1978) 540-547.
2. Ciardo, G., Marie, R., Sericola, B., Trivedi, K.S.: Performability analysis using semi-Markov reward processes. IEEE Transactions on Computers, C-39, (1990) 1251-1264.
3. Goyal, A., Tantawi A.N.: Evaluation of Performability for degradable computer systems. IEEE Transactions on Computers, C-36, (1987) 738-744.
4. Iyer, B.R., Donatiello L., Heidelberger P.: Analysis of Performability models of fault-tolerant systems. IEEE Transactions on Computers, C-35, (1986) 902-907.
5. Iliadis I., Spinellis D., Katsikas S., Preneel B., A Taxonomy of Certificate Status Information Mechanisms. Proceedings of Information Security Solutions Europe ISSE 2000, Barcelona, Spain, (2000).
6. Iliadis J., Gritzalis S., Spinellis D., Danny de Cock., Preneel B., Gritzalis D.: Towards a framework for evaluating certificate status information mechanisms. Computer Communications, 26(16), (2003).
7. Limnios, N., Oprisan, G.: Semi-Markov process and reliability, Birkhäuser (2001).
8. Meyer, J.F.: On evaluating the performability of degradable computing systems. IEEE Transactions on Computers, C-29, (1980) 720-731.
9. Pagès A., Gondran, M. : Fiabilité des Systèmes. Eyrolles, 1980.
10. Pattipati, K.R., Li, Y., Blom, H.A.P.: A unified framework for the performability evaluation of fault-tolerant computer systems. IEEE Transactions on Computers, C-42, (1993) 312-326.
11. Platis, A.: An extension of the performability measure and application in system reliability. Int. J. of Computational and Numerical Analysis and Applications 2 (1) (2002) 87-101.
12. Platis, A., Limnios, N., Le Du, M.: Performability of electrical power systems modeled by non-homogeneous Markov chains. IEEE Transactions on Reliability C-45 (1996) 605-610.
13. Platis, A, Tselios P., Vouros, G.: Attractability: an indicator for optimizing the design of Web sites. Datamining II, WIT Press (2000).
14. Rensburg, A., Solms, S., A reference framework for Certification Authorities / Trusted Third Parties, in L. Yngström and J. Carlsen (Eds.), Proceedings, IFIP 13th International Information Security Conference, Chapman & Hall, (1996).
15. Rivest R., Can We Eliminate Certificate Revocation Lists?. Proceedings of Financial Cryptography (1998).

16. Smith, R.M., Trivedi, K.S., Ramesh, A.V.: Performability analysis: measures, an algorithm and a case study. IEEE Transactions on Computers, C-37, (1988) 406-417.
17. Tsopelas, P., Platis, A.: Performability Indicators for the traffic analysis of wide area networks. Reliability Engineering and System Safety, 82 (2003), 1-9.
18. Van Moorsel, A.: Metrics for the Internet age: Quality of Experience and Quality of Business. HP Technical Report HPL-2001-179 (2001).
19. Wolter, K., Van Moorsel, A.: The relationship between Quality of Service and Business Metrics: monitoring, notification and optimization, HP Technical Report HPL-2001-96, (2001).