International Journal of Innovative
Computing, Information and Control
Volume 8, Number 11, November 2012

 ISSN 1349-4198
pp. 7925–7942

RISK-BASED DECISION MAKING FOR PUBLIC KEY INFRASTRUCTURES USING FUZZY LOGIC

Carlos Gañán, Jose L. Muñoz, Oscar Esparza
Jorge Mata-Díaz and Juanjo Alins

Department of Telematics Engineering
Universitat Politècnica de Catalunya
C. Jordi Girona 31, Barcelona 08034, Spain
{carlos.ganan; jose.munoz; oesparza; jmata; juanjo}@entel.upc.edu

Received July 2011; revised February 2012

Abstract. *Public key infrastructures (PKIs) are complex systems that are responsible for giving users enough information to make reasonable trust judgments about one another. The validation of public keys is hence of great importance. In general, validation of public keys is achieved by public-key certificates. However, in some circumstances certificates have to be revoked. Certificate revocation is generally achieved using Certificate Revocation Lists (CRLs) but the use of CRLs implicitly entails the risk of trusting a certificate that is not included in the list. This can be because the actual status of the certificate is unknown to the CA, or because the CRL is not updated. In this context, in this article we propose a fuzzy risk-based decision making system to assist any network user to make easier its decision-making process when using CRLs.*
Keywords: Risk management, Public key infrastructure, Revocation, Fuzzy logic

1. **Introduction.** From its earliest days in academia, the Internet was designed with the assumption that it was something akin to a "private club", with the main goal being the free exchange of ideas and information. With the birth of the World Wide Web in the early 1990s, the Internet was opened up to anyone, and it soon became clear that something would need to be done to allow users, public and technical alike, to confirm that they were communicating with correctly identified parties. Public Key Infrastructure (PKI) was developed to solve this problem. Moreover, PKI is envisioned as the security solution to provide integrity, authentication and non-repudiation to state-of-the-art networks such as vehicular networks, wireless sensor networks or mobile ad-hoc networks.

PKI is an information technology infrastructure that enables network users to securely and privately exchange information through the use of a public and a private key pair that is obtained and shared through a trusted authority. The public key infrastructure provides for a digital certificate that can identify an individual or an organization. Digital certificates are the means of accurately and reliably distributing public keys to users needing to encrypt messages or to verify digital signatures. Certificates are signed by certification authorities (CAs) and they are issued with a planned lifetime, which is defined through a validity start time and an explicit expiration date. Once issued, a certificate becomes valid when its validity start time is reached, and it is considered invalid after its expiration date. However, various circumstances may cause a certificate to become invalid prior to the expiration of the validity period. Such circumstances include the compromise or suspected compromise of the private key associated to the certificate, requiring to change the name of the subject of a certificate, and modifying the association type between subject and CA, for example, when an employee terminates employment with

an organization. Thus, the PKI has to collect and distribute information about revoked certificates. Currently deployed PKIs rely mostly on Certificate Revocation Lists (CRLs) for handling certificate revocation [1]. A CRL is a list identifying revoked certificates; it is signed by a CA and made available at public distribution points. The CRL has a validity period, and updated versions of the CRL are published before the previous CRL's validity period expires.

However, the use of CRLs to manage revocation in PKIs becomes a risk source. Each new CRL contains a large number of recent revoked certificates that differs from the previous CRL. The number of new revoked certificates will vary depending on the time elapsed since the previous CRL publication. These new revoked certificates are unknown to the user during the validity interval of the current valid CRL. It is during this validity interval when a user could be operating with a revoked certificate without knowing it. In this context, any user will be taking certain risk of operating with an unknown revoked certificate. Therefore, total security is unattainable, even under the unrealistic assumption that CRLs can be delivered to everyone instantaneously. A private key may be compromised long before the compromise is noticed and the certificates revoked. This cannot be handled by current revocation schemes, but it should be taken into consideration when analyzing the inherent risk in a PKI. This article is mainly motivated by the lack of current revocation schemes to manage this risk. This risk associated with the use of a PKI cannot be completely removed, but it can be analyzed and controlled. It is clear that different applications have different risk requirements and that different users have different preferences in the risk-cost balance. Therefore, a PKI aiming to support multiple applications should provide a revocation interface that is tunable. Users should be able to set different recency requirements based on their needs and resources. The aim of this article is to develop a new risk analysis method to identify and assess this risk in an acceptable way in which any risk information is processed and reliably applied to the users' decision-making process.

Previous works in the literature [2-6] acknowledged the existence of an operational risk when using a revocation mechanism such as CRLs. However, these works neither quantified this risk nor provided a means to deal with it. Authors in [7] calculated the probability of considering a certificate as valid when the real status known by the PKI is revoked. This work could be considered as the first step towards making network users aware of the hazards of operating under a PKI, though authors just provided a simple metric to measure the percentage of unknown revoked certificates. The main drawbacks of this proposal are that the standard CRL has to be extended to allow users to calculate this probability, and authors did not deal with the potential consequences of trusting an outdated CRL. To the best of our knowledge, our proposal is the first model that provides a risk indicator to help network users in their decision-making process, taking into account all the risk sources that are present in a PKI.

However, decision making is a tough process. It involves dealing with a lot of uncertainty and projecting what the outcome might be. Depending on the projection of the final outcome, a decision has to be made. So, in order to ease the decision-making process we propose the utilization of a fuzzy system. The user can ease its decision-making process by utilizing a fuzzy approach to risk-based decision making. This fuzzy approach will allow users to further strengthen their previous belief of proceeding in an interaction with a probable trusted user (i.e., a user whose certificates have not been revoked). The proposed risk-based decision making model combines the possibility of interacting with an

illegitimate user and its possible consequences and gives an output to the trusting user[1]. Fuzzy logic is used to model uncertainty, and similarly, the decision-making process deals with predicting the possible uncertain outcome and deciding the future course of action based on the predicted uncertain outcome.

Now, being more specific, we can precise that the goals of this article are to explore the inherent risk in conventional PKIs and to show how fuzzy logic risk analysis techniques can successfully help in the analysis of this risk. By focusing on these goals, we have modeled and characterized a risk-based decision-making system based on fuzzy logic. To that end, in the first place, we analyze the risk of operating with CRLs, determine the possibility of trusting a revoked certificate and analyze the possible consequences of an interaction with a potential illegitimate network user. Then, taking into account the information that users can obtain from the CRLs, we design a fuzzy inference system that gives as output the risk of operating with a particular CRL. Our proposed model can provide to the users an idea of how risky is to operate with their current CRL, and it can also help them to make risk-based decisions. Finally, a case study on risk analysis of a CRL issued by an actual CA (GoDaddy) is used to show the validity of the proposed model. The results of the risk assessment in the case study are represented as risk score, located in a defined range, and risk category with linguistic words, which indicate that by using the proposed methodology the risk associated with CRLs can be assessed effectively and efficiently.

Following this research method, we obtain a fuzzy system that models risk and assists users in their decision-making processes related to certificate revocation. Our solution takes into account a set of key risk factors to estimate the risk of operating when using a particular revocation service. In this respect, we have identified potential risk sources involved in the revocation system and we have characterized them using fuzzy logic. Based on the estimated risk, users can decide whether to interact or not with another PKI user. The results show that although this CA is issuing CRLs with a frequency of only one day, there is still an inherent risk that our model is able to measure.

The rest of this article is organized as follows. In Section 2, we briefly review the basics of fuzzy logic. In Section 3, we identify and characterize a fuzzy inference system that allows estimating the risk of operating with CRLs. Next, in Section 4, we present an empirical case study using data collected from one of the most extended CAs. Finally, in Section 5, we conclude and point out possible future directions.

2. **Fuzzy Logic.** Fuzzy logic aims to model human thinking and reasoning. The key advantage of fuzzy methods is how they reflect the human mind in its remarkable ability to store and process information that is imprecise, uncertain, and resistant to classification [8]. This kind of logic intends to equip computers with the ability to process special data and to work by making use of human experiences and insights. When human logic solves problems, it creates verbal rules such as "if <event realized> is this, the <result> is that" [9]. Fuzzy logic tries to adapt these verbal rules and the human capability to make decisions to computers [10]. It uses verbal variables and terms together with verbal rules [11]. Usually, verbal rules and terms used in human decision-making process are fuzzy rather than precise. Adapting human logic system to computers increases problem-solving capabilities of computers.

Verbal terms and variables are expressed mathematically as membership degrees and membership functions. Fuzzy decision-making mechanisms use symbolic verbal phrases instead of numeric values. Systems that use fuzzy logic are alternatives to the difficulty

[1]A trusting user is a network agent that makes an informed decision of whether to interact with another agent or not, by analyzing beforehand the possible level of risk that could be present in that interaction.

of mathematical modeling of complex non-linear problems and fuzzy logic meets mathematical modeling requirement of a system.

In this context, fuzzy logic emerges as an alternative to the classical logic where every proposition must either be "true" or "false". Instead, fuzzy logic asserts that things can be simultaneously "true" and "not true", with a certain membership degree to each class [12]. It is based on membership functions and linguistic parameters to express vagueness in security issues. Fuzzy logic has the power to handle the concept of "partial truth" to quantify uncertainties associated with linguistic variables. It allows defining a *degree of membership* of an element in a set by means of a membership function. For classical or *crisp* sets, the membership function only takes two values: 0 (non-membership) and 1 (membership). In fuzzy sets the membership function can take any value from the interval $[0, 1]$. The value 0 represents complete non-membership, the value 1 represents complete membership, and values in between are used to represent partial membership [13].

Summing up, fuzzy logic provides a way to use imprecise and uncertain information generated by the system and human judgments in a precise way. In the case of a PKI, as the revocation data available do not provide proper statistical treatment, fuzzy arithmetic is clearly suitable, since it works well for addressing poorly characterized parameters and linguistic variables. Using a fuzzy inference system, we will show that we are able to map a given input set of variables (e.g., CRL age or revocation causes) to an output (e.g., risk indicator). The mapping then provides a basis from which decisions can be made, or patterns discerned. The process of fuzzy inference involves membership functions, logical operations, and If-Then rules. There are two main types of fuzzy inference systems: Mamdani-type and Sugeno-type. These two types of inference systems vary somewhat in the way outputs are determined. In this article, we will use the Mamdani-type inference system due to its adequacy to the problem in question.

2.1. **Mamdani's fuzzy inference method.** Mamdani's fuzzy inference method is the most commonly used fuzzy methodology [14, 15]. Mamdani's method was among the first control systems built using fuzzy logic [16]. It was proposed by Mamdani as an attempt to control a steam engine and boiler combination by synthesizing a set of linguistic control rules obtained from experienced human operators [17]. Mamdani's effort was based on Zadeh's paper on fuzzy algorithms for complex systems and decision processes [18].

The Mamdani-style fuzzy inference process is performed in four steps (see Figure 1):

1. Fuzzification of the input variables.
2. Rule evaluation (inference).
3. Aggregation of the rule outputs (composition).
4. Defuzzification.

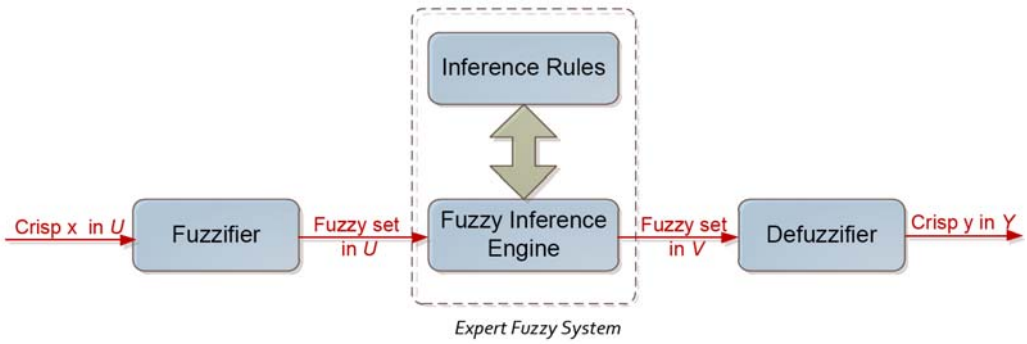

FIGURE 1. A fuzzy logic inference system

A fuzzification operator has the effect of transforming crisp data into fuzzy sets. In most of the cases fuzzy singletons are used as fuzzifiers:

$$fuzzifier(x_0) := \bar{x}_0, \tag{1}$$

where x_0 is a crisp input value from a process.

Suppose now that we have two input variables x and y. A fuzzy control rule

$$R_i : \text{if } x \text{ is } A_i \text{ and } y \text{ is } B_i \text{ then } z \text{ is } C_i,$$

is implemented by a *fuzzy implication* R_i and is defined as:

$$R_i(u, v, w) = [A_i(u) \text{ and } B_i(v)] \to C_i(w), \tag{2}$$

where the logical connective *and* is implemented by the minimum operator, i.e.,

$$\begin{aligned}[A_i(u) \text{ and } B_i(v)] \to C_i(w) &= [A_i(u) \times B_i(v)] \to C_i(w) \\ &= \min [A_i(u), B_i(v)] \to C_i(w).\end{aligned} \tag{3}$$

Fuzzy control rules are combined by using the sentence connective *also*. Since each fuzzy control rule is represented by a fuzzy relation, the overall behavior of a fuzzy system is characterized by these fuzzy relations.

In other words, a fuzzy system can be characterized by a single fuzzy relation which is the combination of the fuzzy relations in the rule set. The combination in question involves the sentence connective *also*. Symbolically, if we have the collection of rules:

R_1 : if x is A_1 and y is B_1 then z is C_1,
also
R_2 : if x is A_2 and y is B_2 then z is C_2,
also
...
R_n : if x is A_n and y is B_n then z is C_n.

The procedure for obtaining the fuzzy output of such a knowledge base consists of the following three steps:

- Finding the firing level of each of the rules;
- Finding the output of each of the rules;
- Aggregating the individual rule outputs to obtain the overall system output.

To infer the output z from the given process states x, y and fuzzy relations R_i, we apply the compositional rule of inference:

R_1 : if x is A_1 and y is B_1 then z is C_1,
also
R_2 : if x is A_2 and y is B_2 then z is C_2,
also
...
R_n : if x is A_n and y is B_n then z is C_n,
input x is $\bar{x}_0$ and y is $\bar{y}_0$

Consequence: z is C

where the consequence is computed by:

$$\text{consequence} = \mathbf{Agg}\ (\text{fact } \circ R_1, \ldots, \text{ fact } \circ R_n). \tag{4}$$

That is,

$$C = \mathbf{Agg}(\bar{x}_0 \times \bar{y}_0 \circ R_1, \ldots, \bar{x}_0 \times \bar{y}_0 \circ R_n), \tag{5}$$

taking into consideration that

$$\bar{x}_0(u) = 0, \quad u \neq x_0, \tag{6}$$

and

$$\bar{y}_0(u) = 0, \quad v \neq y_0. \tag{7}$$

The computation of the membership function of C is very simple:

$$C(w) = \mathbf{Agg}\ \{A_1(x_0) \times B_1(y_0) \to C_1(w), \ldots, A_n(x_0) \times B_n(y_0) \to C_n(w)\} \tag{8}$$

for all $w \in W$.

In the particular case of a Mamdani inference system, the fuzzy implication is modeled by Mamdani's minimum operator and the sentence connective *also* is interpreted as ORing the propositions and defined by the `max` operator.

Thus, the procedure for obtaining the fuzzy output of such a knowledge base can be formulated as:

- The firing level of the i-th rule is determined by:

$$\alpha_i = \min(A_i(x_0), B_i(y_0)).$$

- The output of the i-th rule is calculated by:

$$C_1'(w) = \min(\alpha_i, C_i(w)), \quad \forall w \in W.$$

- The overall system output, C, is obtained from the individual rule outputs C_i' by:

$$C(w) = \max(C_i'(w)), \quad \forall w \in W.$$

The output of the inference process so far is a fuzzy set, specifying a possibility distribution of control action. In the on-line control, a nonfuzzy (crisp) control action is usually required. Consequently, one must defuzzify the fuzzy control action (output) inferred from the fuzzy control algorithm, namely:

$$z_0 = defuzzifier(C), \tag{9}$$

where z_0 is the nonfuzzy control output and $defuzzifier$ is the defuzzification operator.

Defuzzification is a process to select a representative element from the fuzzy output C inferred from the fuzzy control algorithm. The most often used defuzzification operators are Center-of-Area/Gravity, First-of-Maxima, Middle-of-Maxima, Max-Criterion and Height Defuzzification [19].

It is possible, and in many cases much more efficient, to use a single spike as the output membership function rather than a distributed fuzzy set. This is sometimes known as a singleton output membership function, and it can be thought of as a pre-defuzzified fuzzy set. This enhances the efficiency of the defuzzification process because it greatly simplifies the computation required by the more general Mamdani method, which finds the centroid of a two-dimensional function. Rather than integrating across the two-dimensional function to find the centroid, the weighted average of a few data points. Sugeno type systems support this type of model [20]. In general, Sugeno type systems can be used to model any inference system in which the output membership functions are either linear or constant.

3. **Risk Assessment Model for PKI.** A risk analysis should cover all aspects of risks in the revocation process in question and should also specify how the risks involved are to be minimized. Therefore, it should include sufficient particulars to demonstrate that hazards with the potential to cause the network failure can be identified and evaluated, and that the appropriate measures have been taken to reduce risks.

A typical risk assessment framework consists of four stages: risk identification, risk assessment, risk response, and risk monitor and review. The nature of revoking certificates has imposed substantial uncertainties and subjectivities in the risk analysis process. However, no risk assessment method has been proposed to quantify this risk. Fuzzy reasoning techniques have proven useful to handle ill-defined and complex problems arising in other environments to reach a reliable decision [21-24].

In the following, we define a Fuzzy Inference System based on the Mamdani model to capture the risk of operating with a PKI when CRLs are used in the revocation scheme. To do so, in first place, we analyze the risk factors involved in using CRLs and their consequences, and then, we describe each one of the components of the model, i.e., the fuzzifier, the rules of the fuzzy logic system and the defuzzifier. The details are described in the following sections.

3.1. **Analyzing the risk when operating with CRLs.** As discussed earlier, any network user has to make decisions whether to interact or not with a probable legitimate user. Our proposal consists in making users aware of the risk they are taking when assuming as comprehensive the information contained in their cached CRL. Then, users will be able to make an informed decision by analyzing the possible risk that could be present in their interaction. The risk analysis of the trusting user in its potential interaction with an illegitimate user can be done by:

1. Determining the possibility of operating with users that have their certificate revoked;
2. Determining the possible consequences of operating with an illegitimate user.

Hence, the trusting user should consider these two factors for each probable illegitimate user in order to determine the possible risk associated with this interaction. Based on the analysis, the user can make an informed decision of whether to interact or not with another user of the network.

3.1.1. *Determining the possibility of trusting a revoked certificate.* Before operating with another network user, each user has to check the legitimacy of that user. This is achieved by checking the status of her certificate. To check this status, users have to corroborate that the serial number of the user in question is not contained in the CRL. However, though this serial number is not included in the CRL, it could be revoked. As users operate with cached CRLs, certificates that have been revoked after the issuance instant of the CRL are unknown to the users. Thus, as the information contained in the CRL is not totally comprehensive, users have to determine the possibility of trusting a revoked certificate.

This possibility of trusting a revoked certificate captures the extent to which a user thinks that another user of the network is illegitimate. This possibility increases with the time elapsed since the issuance of the CRL. The trusting user can determine this possibility of interacting with a probable illegitimate user by analyzing the information contained in the CRL, i.e., the number of revoked certificates and the issuance and update time of the CRL. Thus, the trusting user should analyze the possibility of operating with users whose certificates have been revoked, and in accordance to the context and criteria of its future interaction with them.

With revocation, users could control this possibility by, for instance, setting freshness requirements for CRL acceptance. Smaller freshness requirements require lower communication costs but lead to a higher risk. Setting the right freshness requirement requires risk analysis and balancing the risk and the cost. It is clear that different applications have different risk requirements and that different users have different preferences in the risk-cost balance. Users should be able to control the possibility of trusting a revoked certificate by setting different CRL recency requirements based on their needs and resources.

3.1.2. *Determining the possible consequences of an interaction.* The trusting user in order to gauge the potential risk in an interaction should also determine the possible consequences of operating with the probable illegitimate user apart from determining the possibility of trusting a revoked certificate. For instance, in a peer-to-peer financial interaction, the possible loss that a trusting user could suffer is usually the financial loss in its resources that are involved in the interaction.

The consequences of operating with users whose certificate has been revoked will be modeled over a scale of 0-10 representing the loss incurred. The possible consequences will vary depending on the revocation cause of the certificates. The PKIX/X.509 certificate and CRL specification [25] defines nine reason codes for revocation of a public-key certificate (see Table 1).

Note that we have defined a weight value w_i for each of the possible revocation causes representing the aforementioned potential threat incurred during the interaction with an illegitimate user. This weighting will allow us to give more importance to those certificates which were revoked due to a key compromise or malicious use. This weighting is purely intuitive as there are some revocation causes that pose bigger threats to the users than other causes. For instance, the compromise of the private key of the CA is more dangerous and has potentially more disastrous consequences than a superseded certificate.

TABLE 1. Revocation codes, weight values w_i and description

Numerical Code	Revocation Code	w_i	Description
(1)	`keyCompromise`	9	Private key has been compromised.
(2)	`cACompromise`	10	Certificate authority has been compromised.
(3)	`affiliationChanged`	1	Subject's name or other information has changed.
(4)	`superseded`	0	Certificate has been superseded.
(5)	`cessationOfOperation`	1	Certificate is no longer needed.
(6)	`certificateHold`	3	Certificate has been put on hold.
(7)	`removeFromCRL`	0	Certificate was previously on hold and should be removed from the CRL.
(8)	`privilegeWithdrawn`	5	Privileges granted to the subject of the certificate have been withdrawn.
(9)	`aACompromise`	10	Attribute authority has been compromised.

Once the user determines the possibility of trusting a revoked certificate and the potential consequences, she should combine those to determine the risk in order to assist decision making. As mentioned earlier, decision making is a tough process as it involves in dealing with a lot of uncertainty. The trusting user, in spite of determining the possibility of trusting a revoked certificate and the possible loss in its resources, might still be uncertain or undecided whether to interact or not with the particular user. To alleviate this problem, we propose the utilization of a fuzzy system which will help the trusting user in its decision-making process. We describe the fuzzy system in the next section.

3.2. **Developing a fuzzy risk based decision making system.** Once the possibility of trusting a revoked certificate and its consequences have been determined, we need a systematic approach to synthesize these constituents of risk into a given risk value for making an informed decision. To this end, we propose the use of a fuzzy approach. The main aim of the fuzzy decision making system is to assist the trusting user with the decision making process. To achieve that, we propose that the trusting user inputs the relative values of the probable trusted user to the fuzzy system, which in turn evaluates them according to the pre-defined rules. Based on the evaluations of the rules, an output is given to the trusting user. The output of the fuzzy system will be a risk value that depending on the user's attitude towards risk, will decide to proceed or not.

3.2.1. *Inputs of the fuzzy inference system.* The first stage to build the fuzzy risk inference system consists in identifying the factors that contribute to the risk. In this way, we identify the key risk factors (KRF) in the revocation process. These key factors are directly related to the aforementioned process of determining the possibility of having a revoked certificate and the consequences of operating with a user whose certificate has been revoked.

The herein proposed mapping technique consists in identifying the key risk factors, capable of reasonably signaling priorities and strategies for risk management purposes. These key risk factors will be those variables, either quantitative or qualitative, which together will serve the purpose of estimating the probability and severity of risk events at the task level.

As proposed in [26], the definition of the KRFs should observe five convenient features:

- *Relevancy*: variables should effectively capture a specific KRF;
- *Generality*: variables can be used across processes or tasks;
- *Non-redundancy*: avoid correlated KRFs;
- *Measurability*: variables should be quantifiable and verifiable;
- *Monitoring facility*: cost and simplicity of monitoring.

According to these features, the proposed KRFs are

1. **Number of revoked certificates** ($NumRev$): as users have cached CRLs which include the list of revoked certificates and their revoked date, users can know the number of revoked certificates per day;
2. **Revocation categories** ($RevCat$): CRLs can also include the revocation cause of each certificate;
3. **Age of the CRL** (CRL_{age}): using also the information contained in the CRL; users can calculate the time elapsed since the issuance of the CRL.

3.2.2. *Fuzzification procedure.* Afterwards, the fuzzification procedure has to be defined to capture KRFs and to be able to translate them into quantitative variables. The foundation of this procedure is the design of the fuzzy sets and the membership functions. The fuzzy set theory will make possible to obtain the imprecise and vague, yet valuable and irreplaceable, judgment of the people associated with the tasks and processes to be

evaluated. It would be clumsy and imprecise to ask for true or false, yes or no, 1 or 0 answers when dealing with variables such as expertise, impact or probability.

In order to translate the judgment of the people into a quantitative variable, the corresponding *membership* functions should be defined for each KRF. Membership functions can have different shapes. The most commonly used shapes are triangular, trapezoidal, Gaussian and bell shaped membership functions. Therefore, for each KRF we choose an appropriate membership function.

The first KRF, the number of revoked certificates, is described by Gaussian shaped membership functions in the following three fuzzy sets as depicted in Figure 2:

- *Low*: the number of revoked certificates per day is low;
- *Moderate*: the number of revoked certificates is neither low nor high;
- *High*: the number of revoked certificates per day is high.

In our fuzzy system, the range of the input $NumRev$ is within $[0, 20]$.

The second KRF, the revocation category, is described by a Generalized bell-shaped membership function. Three different categories are defined according to weight values defined in the Table 1 (see Figure 3).

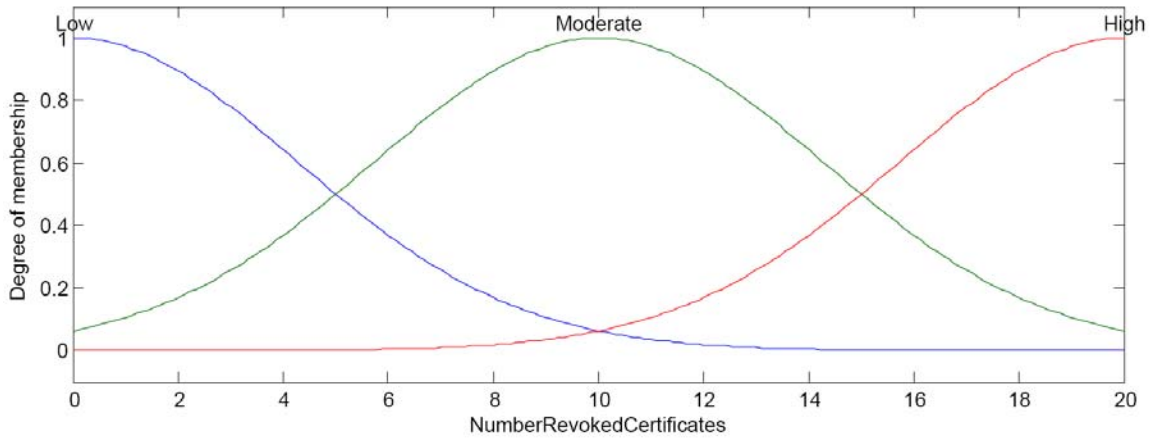

FIGURE 2. Number of revoked certificates as a fuzzy variable

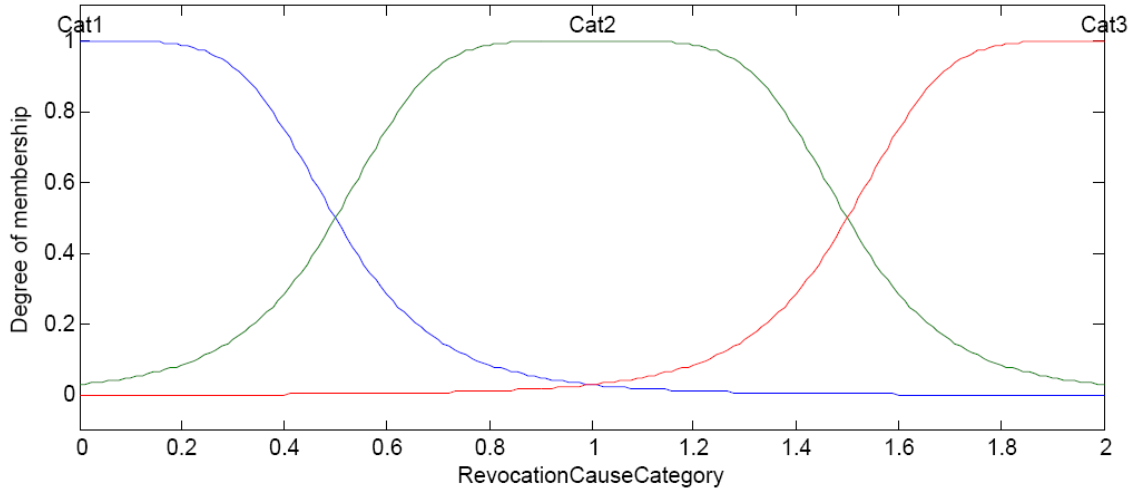

FIGURE 3. Revocation cause categories as a fuzzy variable

- *Category 1*: This category includes revoked certificates that represent a low risk to the network. Thus, this category includes revocation causes such as superseded, change of affiliation or cessation of operation;
- *Category 2*: This category includes revoked certificates that represent a moderate risk to the network. Thus, this category includes revocation causes such as privilege withdrawn or certificate on hold;
- *Category 3*: This category includes revoked certificates that represent a high risk to the network. Thus, this category includes revocation causes such as key compromise, CA compromise or Attribute Authority compromise.

In our fuzzy system the range of the input *RevCat* is within $[0, 2]$.

Finally, we represent the last KRF, the CRL age, by means of a triangular shaped membership function (see Figure 4).

- *New*: Describes CRLs that have recently being updated. In a scale of 0-34 hours, this category ranges from 0 to 9.6 hours.
- *Old*: Describes CRLs that were updated some days ago. In a scale of 0-34 hours, this category ranges from 2.4 to 21.6 hours.
- *Very Old*: Describes CRLs that have not been updated for several weeks. In a scale of 0-34 hours, this category ranges from 14.4 to 33.6 hours.

The output of risk is defined in the following five classes as shown in Figure 5:

1. *Unacceptability High*: High probability that trusting a potentially revoked certificate will cause important damages in the network and potential losses.
2. *High*: High probability of trusting a revoked certificate and cause moderate damages in the network.
3. *Moderate*: Probability of trusting a revoked certificate is intermediate and the potential consequences are limited.
4. *Low*: Risk of operating with an illegitimate user is low.
5. *Negligible*: There is almost no risk in trusting the information contained in the CRL as comprehensive.

3.2.3. *Rules for the fuzzy logic system.* According to the Mamdani approach, we need some rules to process the inputs to let the fuzzy system to conclude at an output. Linguistic rules in the fuzzy system consists of two parts, an antecedent (between the IF and

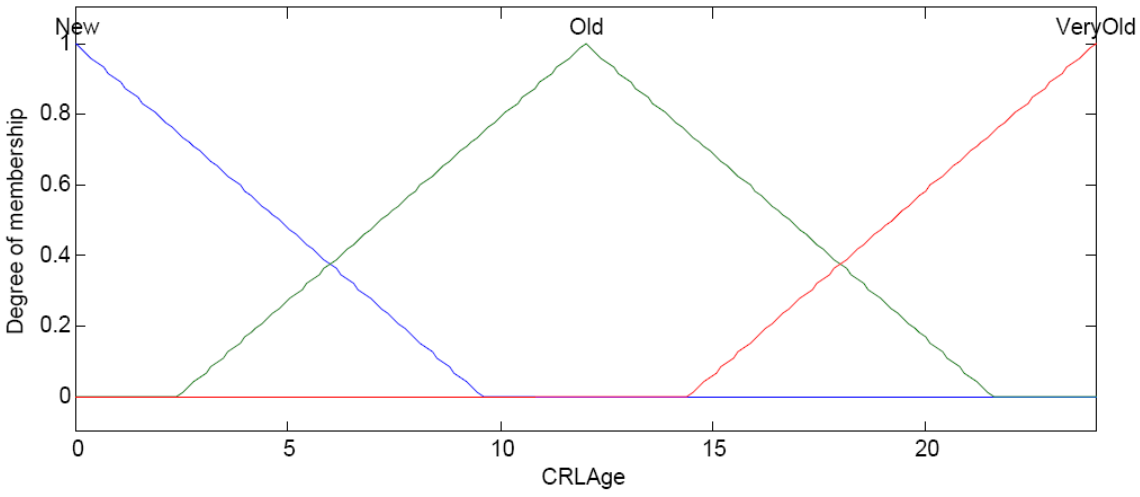

FIGURE 4. CRL age as a fuzzy variable

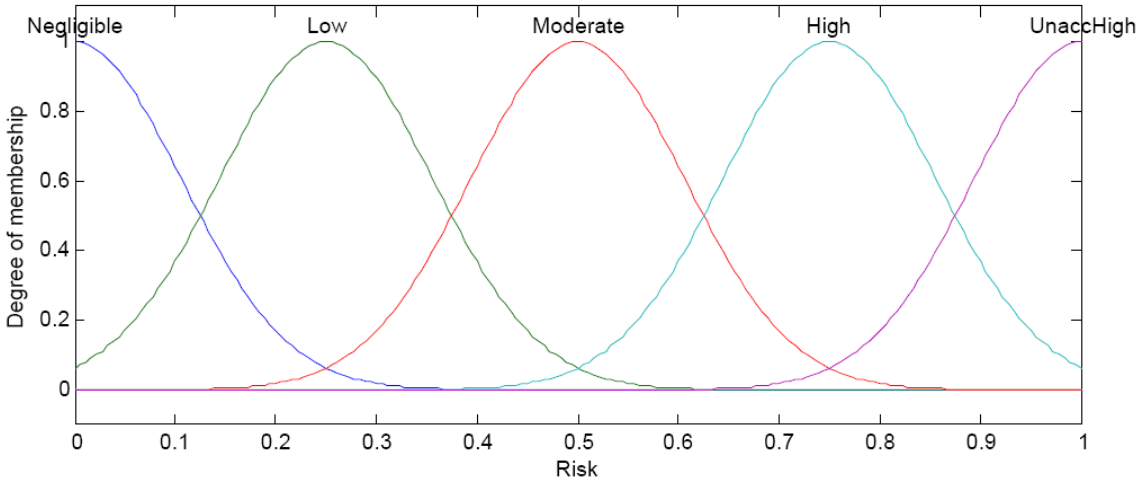

FIGURE 5. Risk as a fuzzy variable

THEN) and consequent (following THEN). There are 3 inputs to our fuzzy system and 3 fuzzy sets. Hence, the total number of rules is: $3^3 = 27$. However, some of these rules are correlated so they can be simplified. There is not a method for establishing the optimal number of inference rules, but achieving an intuitive, smooth and continuous solution space for every combination of KRFs is a fair rule of thumb. Following this guideline, we define eleven inference rules. These rules relate inputs and outputs as:

R_1: If ($NumRev$ is Low) and (CRL_{age} is New) then (Risk is Negligible)
R_2: If ($NumRev$ is High) and (CRL_{age} is New) then (Risk is Low)
R_3: If ($NumRev$ is Low) and (CRL_{age} is Old) and ($RevCat$ is Cat1) then (Risk is Low)
R_4: If ($NumRev$ is Low) and (CRL_{age} is Old) and ($RevCat$ is Cat2) then (Risk is Moderate)
R_4: If ($NumRev$ is Low) and (CRL_{age} is Old) and ($RevCat$ is Cat3) then (Risk is High)
R_6: If ($NumRev$ is Moderate) and (CRL_{age} is Old) then (Risk is High)
R_7: If ($NumRev$ is High) and (CRL_{age} is Old) then (Risk is High)
R_8: If (CRL_{age} is VeryOld) and ($RevCat$ is Cat3) then (Risk is UnaccHigh)
R_9: If (CRL_{age} is VeryOld) and ($RevCat$ is Catl) then (Risk is Moderate)
R_{10}: If (CRL_{age} is VeryOld) and ($RevCat$ is Cat2) then (Risk is High)
R_{11}: If ($NumRev$ is Moderate) and (CRL_{age} is New) then (Risk is Low)

This set of inference rules (or knowledge base) has the objective of deconstructing expert's knowledge and encoding it in a form that the fuzzy logic inference system is capable of mimicking human's reasoning capabilities to solve complex systems. Therefore, an expert (or group of experts) analyzes the KRFs, their different linkages and their relation to the linguistic variables in the output space, resulting in a list or set of educated inference rules that will solve simultaneously any combination of inputs and calculate the expected OR Indicator.

3.2.4. *Defuzzification.* Having specified the input space, the output space, and the rule base, the method for estimating the expected risk is to be defined. Cox in [27] highlights centroid's consistency and well-balanced approach, its sensitiveness to the height and width of the total fuzzy region and the smooth changes in the expected value of the output across observations. Additionally, Cox affirms that it behaves in a manner similar to Bayesian estimates, that is, it selects a value that is supported by the knowledge

accumulated from each executed proposition. Taking into account these advantages and because it is the most used method [27, 28], centroid or center of gravity method is used to defuzzify.

3.3. **Results.** Based on the set of inference rules the fuzzy inference system is capable of inferring all the attainable risk results for any KRFs combination. These results are best presented as a surface plot. Figure 6 is a three-dimensional depiction of the set of rules as a check on consistency. A number of inference methodologies exists for combining inputs and outputs. The method used in the proposed model is that of Mamdani [17] as it is considered the most appropriate.

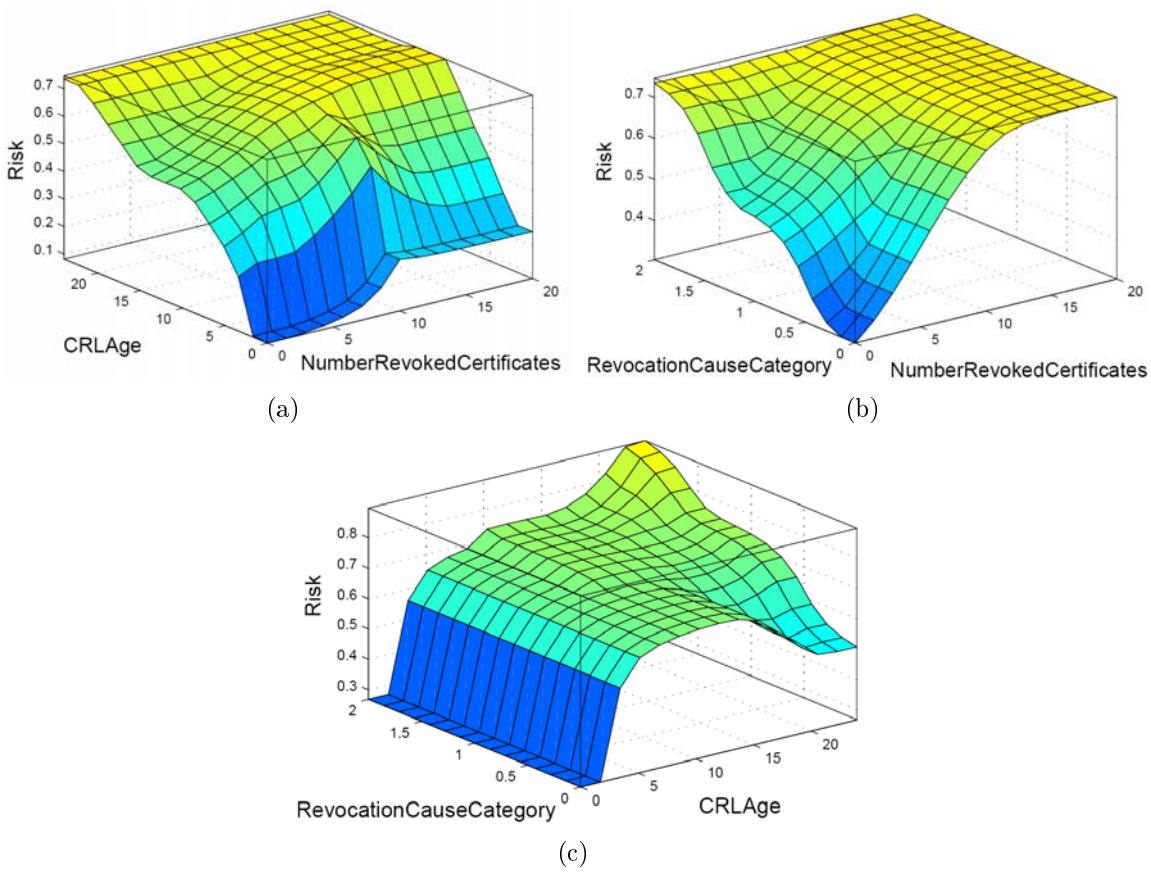

FIGURE 6. Risk indicator as a combination of (a) the CRL age and the number of revoked certificates, (b) the revocation cause categories and the number of revoked certificates, (c) the CRL age and the revocation cause categories

Figure 6, somewhat similar to a probability/severity chart, displays the nonlinear relation between the inputs and the risk indicator, where each combination of these KRFs results in a unique position on the surface. Intuitively, if a revocation event happening has a low (high) impact on the network and a low (high) probability, the risk yields a low (high) outcome, where intermediate results are also considered according to the knowledge base.

4. **A Case Study: GoDaddy.** Finally, to corroborate the benefits of the presented model, we analyze the case of a company that issues digital certificates. The selected company is GoDaddy, which is currently operating in the SSL market and it is the trusted

provider of Internet infrastructure services for the networked world that leads the global SSL marketplace with a 20.52% share [29]. Notice that though the collected data belongs exclusively to a single certification authority, the analysis of this case study is clearly representative because of the market share of the analyzed CA. Thus, the results obtained here can be extended to any other CA operating in the same market.

Using GoDaddy's Signing Certificate Revocation List [30], we analyzed a large sample of revoked certificates. Note that using this type of CRL, we are covering just one type of certificate: Code-Signing certificates. GoDaddy offers code-signing certificates for use by software developers and software vendors. The purpose of such a certificate is to sign code that users download off the Internet. By signing the code, users can be assured that the code has not been tampered with or corrupted since it was digitally signed with the private key of the software developer. In the online world, where people are not only becoming increasingly aware of security issues, but also worry about viruses and worms, signing the code provides a certain assurance to users that they are getting the software that they're expecting to get.

From each GoDaddy's CRL, we can obtain the three KRFs that our fuzzy inference model needs. To do so, we have to obtain the following parameters:

- Last Update instant of the CRL.
- Next Update instant of the CRL.
- Serial Number of each revoked certificate.
- Revocation Date of each revoked certificate.
- Revocation Code of each revoked certificate.

For the first KRF, we use the validity period of each CRL. This allows us to determine the range of the CRL Age. In the case of GoDaddy, CRLs are issued every 24 hours. Therefore, a CRL will be considered $VeryOld$ after this period of time.

For the second KRF, we need to determine the number of revoked certificates per day. For this purpose, we need the revocation date of each certificate and its serial number. With this information, we can analyze the time evolution of the number of revoked certificates per day (see Figure 7) and we can tally the number of revocation that occurred every day from 2009 to 2011.

Finally, the third KRF is also obtained from the CRL by means of the revocation code. Using the revocation code of each revoked certificate, we can calculate the impact of the revocation causes using the weights established in Table 1. In this context, Figure 8 shows the revocation causes of the certificates contained in GoDaddy's CRL. This analysis covered more than 300,000 certificates. It is worth noting, that the main cause

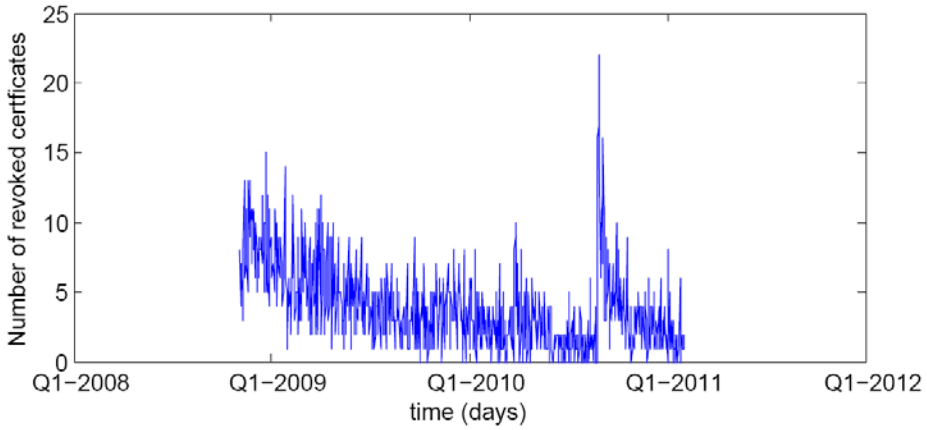

FIGURE 7. Number of revoked certificates evolution

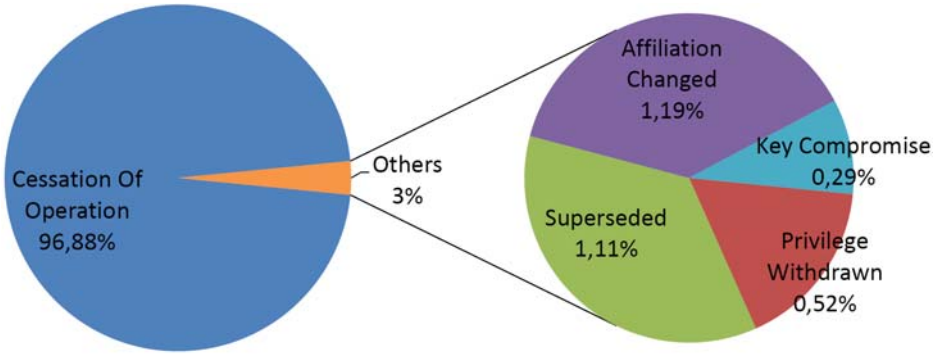

FIGURE 8. Revocation causes of code-signing certificates issued by GoDaddy

TABLE 2. Risk analysis score for ten days

Day	Number Revoked Cert	CRL Age	Revocation Category	Risk
24/01/2009	14	9 hours	Cat 1	0.686
22/04/2009	1	12 hours	Cat 2	0.523
21/05/2009	5	8 hours	Cat 2	0.567
18/05/2009	6	1 hour	Cat 3	0.112
25/08/2010	16	2 hours	Cat 2	0.253
27/08/2010	20	12 hours	Cat 2	0.748
16/09/2010	1	18 hours	Cat 1	0.424
28/09/2010	10	22 hours	Cat 3	0.892
22/10/2010	1	0.5 hours	Cat 1	0.0824
08/11/2010	5	10 hours	Cat 1	0.500

of revocation is the cessation of operation, i.e., the certificate is no longer needed for its original purpose. The rest of revocation causes are highly improbable compared with the main cause. Therefore, when a certificate is revoked by GoDaddy it is highly probable that the revocation is due to a cessation of operation.

Using these data, we choose 10 different days at different hours and we calculate the risk indicator using the proposed fuzzy inference system. Table 2 shows the results obtained. As expected, days with more revoked certificates involve higher risks. In a similar way, as the CRL becomes older the risk also increases. Finally, the category of the revoked certificates also affects significantly in the risk. However, as shown in Figure 8 the most common category is $Cat1$, as the predominant revocation cause has a weight $w_i = 1$.

For computing the previous risk outputs, we have used the Mamdani Method (as shown in Figure 9). For example, for 24/01/2009, we have obtained a scalar value of 0.686. This value is the result of using defuzzification with the centroid method. This method allows us to obtain a crisp output from the fuzzy output. As can be observed, in this particular case, the output value falls in the lower part of risk class defined as *Moderate*.

The previous results show that our risk analysis method is an effective and efficient way to assess the risk associated with the revocation system of GoDaddy and by extension to other SSL providers.

5. **Conclusions.** PKI requires a revocation mechanism to remove illegitimate users. This is commonly achieved using certificate revocation lists (CRLs). However, using CRLs presents a great challenge to users, as they have to make critical decisions based on the information contained in these lists but the information available is not always complete,

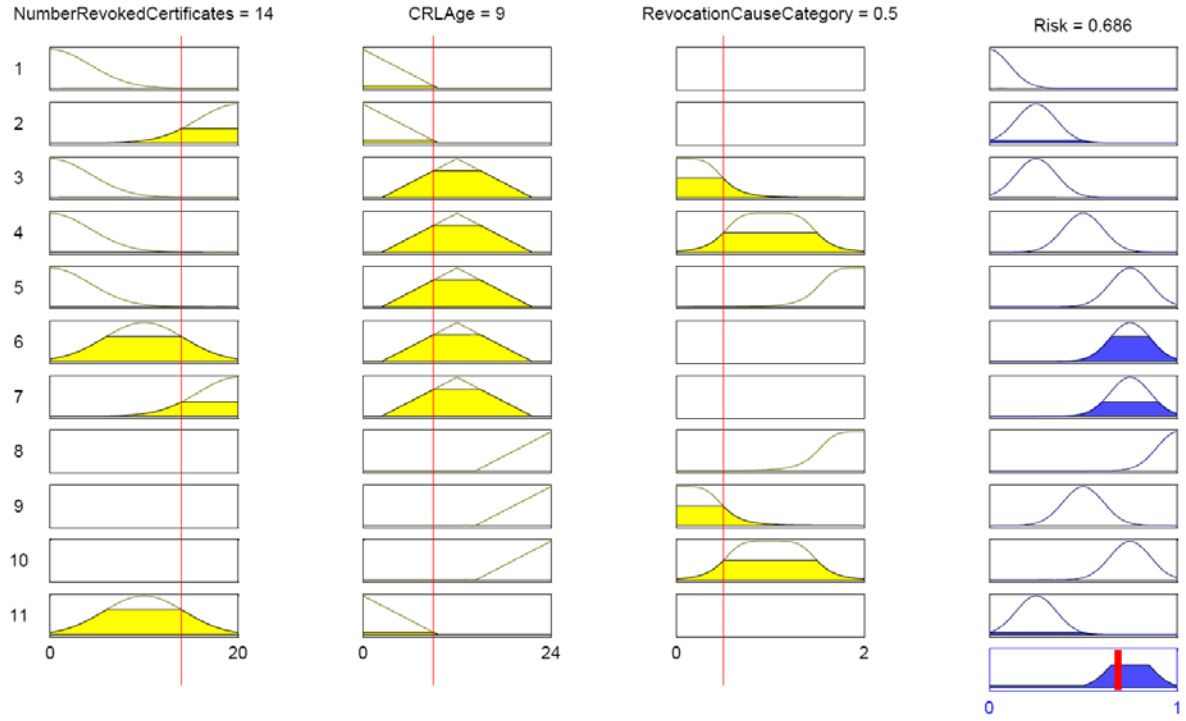

FIGURE 9. Mamdani method for computing the risk output (Jan 24 09:25:44 2009 GMT)

precise or updated. The key finding of this article is a systematic methodology to build a fuzzy system that models risk and assists the user in the decision making process related to certificate revocation. Our system not only considers the possibility of taking as valid a certificate that has been revoked but also other key risk factors (KRF). In this respect, we have identified potential risk sources involved in the revocation system and we have characterized them using fuzzy logic. The inputs given to the fuzzy system can be inferred from a standard CRL and as CRLs are accessible to any PKI user, in practice, everybody can take advantage of our fuzzy system. The output of our system is a measure of the risk of operating with a particular CRL at a given instant. Based on this output, the user can either decide whether to interact or not with another PKI user. Finally, a real certification authority (GoDaddy) has been analyzed using our fuzzy system. The results show that although this CA is issuing CRLs with a frequency of only one day, there is still an inherent risk that our model is able to measure.

Both academia and industry could benefit from our proposal. On the one hand, using the proposed fuzzy inference system, CAs could set policies to manage the issuance of CRLs to control the inherent risk of revocation within the PKI. CAs need to maintain an equilibrium between the costs of issuing a CRL, which increase significantly when a CRL is released frequently, and the risk, which tends to skyrocket if a CRL is not released in a timely manner. On the other hand, researchers could use our model to measure the risk of new revocation mechanisms that operate in novel environments such as vehicular networks or sensor networks.

As a final remark, we have to mention that our study has some limitations but probably each limitation provides an opportunity for further research. In first place, we assume that the CRL contains only one type of certificate. In practice, it could be the case that a CRL includes different types of certificates. Depending on the type of certificate, the involved risk changes. Thus, it could be interesting to analyze how the type of

certificate affects the risk-based decision making. Secondly, we have only analyzed the specific case of a PKI using CRLs as revocation mechanism. However, there are other revocation mechanisms such as the Online Certificate Status Checking Protocol (OCSP) that also involve risk when deployed. Analyzing the differences regarding the operational risk among the different revocation mechanisms could be another interesting area for future research.

Acknowledgment. This work is funded by the Spanish Ministry of Science and Education under the projects CONSOLIDER-ARES (CSD2007-00004) and TEC2011-26452 "SERVET", and by the Government of Catalonia under grant 2009 SGR 1362.

REFERENCES

[1] R. Housley, W. Polk, W. Ford and D. Solo, Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile, *RFC 3280, Internet Engineering Task Force*, 2002.

[2] N. Li and J. Feigenbaum, Nonmonotonicity, user interfaces, and risk assessment in certificate revocation (position paper), *Proc. of the 5th International Conference on Financial Cryptography, Lecture Notes in Computer Science*, vol.2339, pp.166-177, 2002.

[3] B. Fox and B. LaMacchia, Certificate revocation: Mechanics and meaning, *International Conference on Financial Cryptography, Lecture Notes in Computer Science*, vol.1465, pp.158-164, 1998.

[4] R. L. Rivest, Can we eliminate certification revocation lists? *International Conference on Financial Cryptography, Lecture Notes in Computer Science*, vol.1465, pp.178-183, 1998.

[5] P. McDaniel and A. Rubin, A response to can we eliminate certificate revocation lists, *Proc. of the 4th International Conference on Financial Cryptography, Lecture Notes in Computer Science*, vol.1962, pp.245-258, 2000.

[6] C.-I Hsu, C. Chiu and M. S.-H. Ho, The prediction of PKI security performance using PSO and bayesian classifier, *ICIC Express Letters*, vol.3, no.4(A), pp.1031-1036, 2009.

[7] J. L. Muñoz, O. Esparza, C. Gañán and J. Parra-Arnau, PKIX certificate status in hybrid MANETs, *WISTP, Lecture Notes in Computer Science*, vol.5746, pp.153-166, 2009.

[8] L. A. Zadeh, Fuzzy sets, *Information and Control*, pp.338-353, 1965.

[9] L. A. Zadeh, The calculus of fuzzy if/then rules, *Fuzzy Days'92*, pp.84-94, 1992.

[10] L. A. Zadeh, Commonsense knowledge representation based on fuzzy logic, *IEEE Computer*, pp.61-65, 1983.

[11] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, *Information Science*, pp.301-357, 1975.

[12] L. A. Zadeh, The role of fuzzy logic in the management of uncertainty in expert systems, *Fuzzy Sets Syst.*, vol.11, pp.197-198, 1983.

[13] T. J. Ross, *Fuzzy Logic with Engineering Applications*, John Wiley & Sons, 2010.

[14] F. Eshragh and E. H. Mamdani, A general approach to linguistic approximation, *International Journal of Man-Machine Studies*, vol.11, pp.501-519, 1979.

[15] E. H. Mamdani and S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, *Int. J. Hum.-Comput. Stud.*, vol.51, pp.135-147, 1999.

[16] E. H. Mamdani, Application of fuzzy logic to approximate reasoning using linguistic synthesis, *Proc. of the 6th International Symposium on Multiple-Valued Logic*, pp.196-202, 1976.

[17] E. H. Mamdani, Application of fuzzy algorithms for control of simple dynamic plant, *Proc. of the Institution of Electrical Engineers*, vol.121, no.12, pp.1585-1588, 1974.

[18] L. A. Zadeh, Fuzzy logic and its application to approximate reasoning, *IFIP Congress'74*, pp.591-594, 1974.

[19] R. Fullér, Introduction to neuro-fuzzy systems, *Advances in Soft Computing*, 2000.

[20] M. Sugeno, *Industrial Applications of Fuzzy Control*, Elsevier Science Inc., 1985.

[21] Y. Liao, C. Ma and C. Zhang, A new fuzzy risk assessment method for the network security based on fuzzy similarity measure, *The 6th World Congress on Intelligent Control and Automation*, vol.2, pp.8486-8490, 2006.

[22] D. Zhao, J Wang and J. Ma, Fuzzy risk assessment of the network security, *International Conference on Machine Learning and Cybernetics*, pp.4400-4405, 2006.

[23] A. S. Sendi, M. Jabbarifar, M. Shajari and M. Dagenais, Femra: Fuzzy expert model for risk assessment, *The 5th International Conference on Internet Monitoring and Protection*, pp.48-53, 2010.
[24] C. Hu and C. Lv, Method of risk assessment based on classified security protection and fuzzy neural network, *Asia-Pacific Conference on Wearable Computing Systems*, pp.379-382, 2010.
[25] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley and W. Polk, Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile, *RFC 5280, Internet Engineering Task Force*, 2008.
[26] S. Scandizzo, Risk mapping and key risk indicators in operational risk management, *Economic Notes*, vol.34, no.2, pp.231-256, 2005.
[27] E. Cox, *The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems*, Academic Press Professional, Inc., 1994.
[28] S. N. Sivanandam, S. Sumathi and S. N. Deepa, *Introduction to Fuzzy Logic Using MATLAB*, Springer-Verlag, New York, 2006.
[29] WhichSSL, *SSL Market Share*, http://www.whichssl.com/ssl-market-share.html, 2010.
[30] *Legal Repository from GoDaddy*, http://certs.godaddy.com/anonymous/repository.seam.