An Evidence-Based Trust Assessment Framework for Critical Infrastructure Decision Making

Yujue Wang, Carl Hauser

To cite this version:

Yujue Wang, Carl Hauser. An Evidence-Based Trust Assessment Framework for Critical Infrastructure Decision Making. 5th International Conference Critical Infrastructure Protection (ICCIP), Mar 2011, Hanover, NH, United States. pp.125-135, 10.1007/978-3-642-24864-1_9 . hal-01571778

HAL Id: hal-01571778
https://hal.inria.fr/hal-01571778

Submitted on 3 Aug 2017

Chapter 9

AN EVIDENCE-BASED TRUST ASSESSMENT FRAMEWORK FOR CRITICAL INFRASTRUCTURE DECISION MAKING

Yujue Wang and Carl Hauser

Abstract The availability and reliability of large critical infrastructures depend on decisions made by hundreds or thousands of interdependent entities and, by extension, on the information that the entities exchange with each other. In the electric power grid, the widespread deployment of devices that measure and report readings many times per second offers new opportunities for automated control, which is accompanied by the need to automatically assess the trustworthiness of the received information. This paper proposes a Bayesian estimation model for calculating the trustworthiness of entities in the electric power grid and making trust-relevant decisions. The model quantifies uncertainties and also helps minimize the risk in decision making.

Keywords: Trust assessment, Bayesian estimation, electric power grid

1. Introduction

The U.S. electric power grid is on the cusp of a tremendous expansion in the amount of sensor data available to support its operations. For decades, the power grid has been operated using supervisory control and data acquisition (SCADA) systems that poll each sensor once every two or four seconds – a situation that some in the industry have characterized as "flying blind." The widespread deployment of sensing systems called phasor measurement units (PMUs) that provide accurately time-stamped data 30, 60 or more times per second is near at hand. By the end of 2013, utilities, with the assistance of the American Recovery and Reinvestment Act of 2009 (ARRA), will have increased the number of these devices on the grid to nearly 1,000, roughly an order of magnitude increase over what exists today. Data from PMUs and other high-rate sensing devices will support new control schemes for the reliable and efficient operation of the power grid as larger fractions of electric power demand are met by intermittent sources such as wind and solar, and as controllable loads such as electric vehicle rechargers increase.

As power grid operations come to increasingly rely on these new control schemes, the availability and integrity of the data, but to some degree confidentiality as well, are of great concern. Good security practices and technologies such as those required by the NERC CIP standards [15] will be more essential to reliable grid operations than they are today. However, uncertainty is inherent to large-scale systems such as the power grid due to measurement errors (e.g., sensor reading errors) and the stochastic nature of physical processes (e.g., weather conditions).

Uncertainties associated with PMUs arise from several factors: (i) PMUs are deployed under the control of various entities, throughout the transmission and distribution systems, that have different management policies and configurations; (ii) a number of cyber attacks on PMUs can impact electric power systems; and (iii) with vastly more data available, it becomes possible to use a subset of data from the most trustworthy sources. For example, when PMU data authentication is performed using a public-key infrastructure, the reliability of the authentication is ultimately limited by the uncertainty of the binding between a particular public key and the authenticated entity. While one might wish that there were no uncertainty, it is, in fact, quite likely that some of the bindings in a large-scale system are incorrectly known at least some of the time by some entities, whether due to error or malicious manipulation.

If uncertainty is indeed unavoidable, the reliability of the system comes down to blind faith – we know that security is uncertain but we have to trust in it because it is all we have – or to decision processes that explicitly and appropriately take into account the uncertainties associated with security. This paper focuses on the latter viewpoint.

Since the power grid must be controlled in real time in an ever-changing security threat environment, we are interested in decision models that can be fully automated rather than models that rely on human insight. Because Bayesian decision theory fits well with our desire for a computational solution, our approach uses a Bayesian perspective [19].

The word "trust" is introduced here for its connotations of one party's (trustor's) reliance on and belief in the performance of another party (trustee). An example is the trust in a public-key infrastructure (PKI) certifier to correctly bind a public key to some other entity. The reliance or belief often must occur without certainty or they may be in the form of a prediction about the future (itself a source of uncertainty). Trust, though uncertain, need not be blind: trustors can use evidence, in the form of past experience with a trustee, reputation information, or contracts and laws that impose penalties for nonperformance, to form their trust judgments. We believe that if critical infrastructures are to be resilient against attacks, then operational decision making processes must appropriately take into account evidence about the trustworthi-

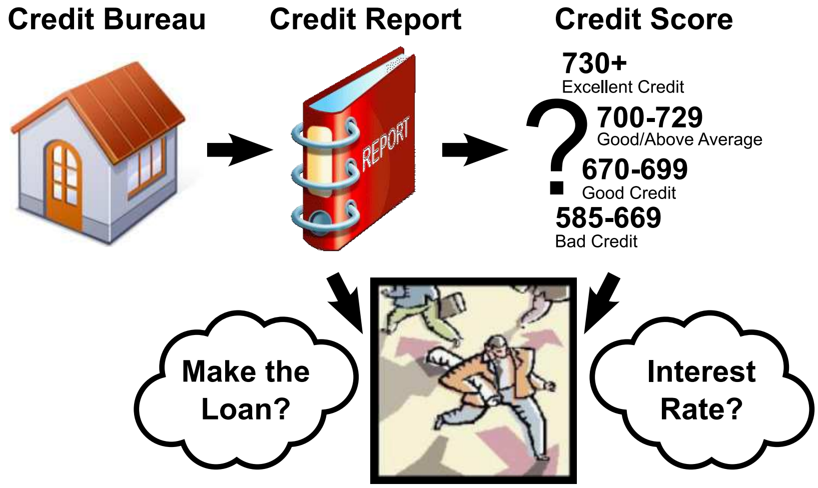

Figure 1. Credit reporting system.

ness of their input data. As we will show, using evidence appropriately means that it is considered in the light of the particular decision being made: there is no single approach to judging trust that is universally appropriate.

This paper establishes the need for a systematic method of dealing with uncertainty related to trust in the context of control systems. It presents a theoretical framework based on Bayesian decision theory that addresses this need by incorporating trust-related evidence.

2. Motivating Example

The consumer credit reporting and scoring system provides an interesting analogy for evidence-based decision making in the presence of risk. As shown in Figure 1, credit bureaus collect information from various sources and provide credit reports that detail individual consumers' past borrowing and bill paying behaviors. Some companies further analyze the information in credit reports from multiple sources to produce a single, numerical credit score based on the statistical analysis of an individual's credit reports. The credit score is claimed to statistically represent the creditworthiness of an individual.

Now consider the decisions that lenders make in analyzing a loan application. They have to decide whether or not to make the loan and the terms under which the loan should be made. If the loan is made, then the lender stands to make a profit if the borrower pays it back, or the lender makes a loss if the borrower defaults on the payment. A loss function describes the lender's payback for various future behaviors of the borrower. While the loss function is known, the future behavior of the borrower is, of course, uncertain at the time the loan is made. The lender thus seeks to make a decision that minimizes the expected loss (maximizes the expected return) by assessing the probabilities of different future borrower behaviors. This is done by considering the credit report or credit score as well as information about employment, income and stability of residence that is contained in the loan application.

There are several important points to note in this analogy.

- First, different lenders have different loss functions, and a single lender may have different loss functions for different kinds of loans: trust decisions are situational. In the power grid domain, a decision to turn off electric car charging at a time when the power supply is stressed carries different loss implications than a decision to shed load by turning off power to an entire region.

- Second, different lenders may assess the probability of various borrower behaviors differently based on the same credit report facts: trust decisions are subjective.

- Third, the analogy is imperfect. In the case of lending, risk pooling allows businesses to balance losses from some loans with profits from others, so decisions take into account not only an individual loan but a portfolio of loans. The consequences of decisions related to power grid operations cannot be easily aggregated, so the decision processes generally emphasize the analysis of individual decisions.

Thus, there are similarities and differences between the two domains. However, the structure is basically the same: the trustor collects evidence about a trustee and uses it to probabilistically predict the behavior of the trustee according to a model. The trustor may make decisions that, in hindsight, seem wrong, but the decisions are, nevertheless, the best that could be made at the time based on the available information.

3. Preliminaries

The distributed control system for a large-scale critical infrastructure (such as the electric power grid) can be described abstractly as consisting of a collection of controllers, a collection of data sources and a collection of actuators. Controller and actuators share the essential characteristics related to the uncertainty of security, so we focus primarily on controllers and information sources. In the power grid, for example, controllers include protective relays, automatic generator controls, remedial action schemes, etc. Data sources include sensors, human operators and controller outputs. Communication channels link data sources to controllers. The essential property of controllers is that they receive inputs from data sources and repeatedly make decisions based on the data, with the decisions ultimately being reflected in an action that changes the physical state of the grid.

Because of the noise in sensor outputs, system inputs are assumed to be probabilistically related to the actual state of the sensed world by considering that each measurement corresponds to the actual state plus a normally-distributed noise term. System failures can lead to bad inputs (highly improbable in the model with normally-distributed noise), which can often be detected and excluded by bad-data detection algorithms that exploit redundancy present in the inputs. Several researchers have studied how input data streams might be intentionally attacked in a manner that is invisible to the bad-data detectors that are in use today (see, e.g., [14]).

The approach described in this paper is, at a high level, aimed at providing controllers with the ability to evaluate evidence from a variety of sources regarding the correctness of data received from sensors and the ability of actuators to carry out the commanded actions. The uncertainties associated with these aspects and the outcomes are modeled probabilistically, albeit with much greater flexibility than afforded by current approaches that assume normally-distributed noise, and with the explicit incorporation of uncertain results in the form of loss functions.

4. Bayesian Decision Model

Decision theory studies the values and uncertainties related to making rational and optimal decisions [11]. Statistical theory has been widely applied to decision making problems [17]. Our method is based on the Bayesian statistical paradigm, which quantifies the uncertainty of decisions using personal probability [16]. Interested readers are referred to [19] for a systematic introduction to Bayesian decision theory.

As previously noted, uncertainty is inherent in complex systems. Thus, risk, which is a state of uncertainty where some of the possibilities involve a loss, catastrophe or other undesirable outcome, is unavoidable. In order to reduce risk, every entity in a system should have the ability to incorporate evidence about the trustworthiness of other entities and be inclined to rely on entities that are (more) trustworthy.

To formalize this view, we assume that there are a number of trust-related attributes $E = (E_1, E_2, \ldots, E_p)$ concerning each entity in the system, which together form the trust evidence. By focusing on a single entity $\mathcal{A}$ at a certain point in time, it is possible to collect the current evidence about an entity $\mathcal{B}$, which is denoted as $x_i = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p) \in \mathbb{R}^p$. Over a period of time, a number of x_is, denoted by $x = (x_1, x_2, \cdots, x_n)$, would be collected. Based on x, $\mathcal{A}$ makes a decision $d \in \mathcal{D}$ (where $\mathcal{D}$ is the decision space) on $\mathcal{B}$ in the light of $\mathcal{A}$'s estimate of the value of θ $(0 \leq \theta \leq 1)$ from the parameter space Θ, which is the trustworthiness that is placed on $\mathcal{B}$. Essentially, θ is the probability that $\mathcal{B}$ is trustworthy.

In the proposed model, the decision making process is considered as a choice of action made by the decision maker from among a set of alternatives according to their possible consequences. In the electric power grid, these decisions are made under uncertainty, i.e., the decision maker can neither know the exact consequence of a chosen decision before it occurs nor obtain accurate values of the evidence due to system complexity and uncertainty. Probabilistic modeling is a natural choice for interpreting the evidence E and evaluating the consequences. The model should not only incorporate the available information in E, but also the uncertainty of this information. In the probabilistic model, $x_i\,(1 \leq i \leq n)$ follows a probability distribution f_i, $x_i \sim f_i\left(x_i|\theta, x_1, \cdots, x_{i-1}\right)$ on $\mathbb{R}^p$, where f_i is known but θ is unknown. If x is collected over a short enough period of time, it is reasonable to assume that $x_1, x_2, \cdots, x_n$ are independent repeated trials from identical distributions and that the distribution can be denoted as:

$$x \sim f\left(x|\theta\right).$$

The likelihood function l is defined as:

$$l\left(\theta|x\right) = f\left(x|\theta\right).$$

The likelihood function l is equal to f, but it emphasizes that θ is conditional on x and manifests that θ can be inferred from x. According to our assumptions and the likelihood principle [3], all available information to make an inference about θ is contained in the likelihood function $l\left(\theta|x\right)$. Decisions can then be made based on the inferred value of θ. To combine these processes, when the likelihood function $l\left(\theta|x\right)$ is fixed, a function from $\mathcal{X}$ to $\mathcal{D}$ can be obtained as $\delta\left(x\right)$, which is called the decision rule as it relates to trust. Note that trustworthiness assessment is only one aspect of the overall decision process. Decisions are made according to the inferred trustworthiness value, but trustworthiness evaluation is not the end goal.

Next, we describe the elements involved in a Bayesian determination of decision rule $\delta\left(x\right)$, namely prior distributions and loss functions, and proceed to specify the derived rule.

4.1 Modeling Prior Information

As previously noted, trust decisions are subjective: based on the very same evidence, different trustors may make different decisions. In the Bayesian model, the uncertainty of the trustworthiness value θ of a trustor regarding a trustee before receiving evidence is modeled using a prior probability distribution $\pi(\theta)$ on Θ. Subjectivity of trust is naturally modeled by different prior distributions.

4.2 Loss Function

While it is easy to talk about making "good" decisions, the model requires a precise formalization of the notion of goodness. All the possible choices in a decision should be ordered or quantified. Decision theory uses a loss function for this purpose. The loss function is a function $L \geq 0$ from $\Theta \times \mathcal{D}$ to $\mathbb{R}^p$, which represents the penalty $L\left(\theta, d\right)$ associated with the decision d when the parameter takes the value θ. In our case, the penalty $L\left(\theta, d\right)$ is the quantified consequence at the time the decision is made when the trustee's trustworthiness value is θ and the trustor chooses decision d. However, it is very hard to measure the trustworthiness value of a trustee in a complex system due to the dynamic and fuzzy nature of trust [6]. Therefore, it is important that the model can reflect this uncertainty. A simple way to obtain the loss is to integrate over all possible values of θ. Moreover, instead of focusing on one decision, our goal is to assess a decision rule $\delta(x)$, which is the allocation of a decision to each outcome $x \sim f(x|\theta)$. Thus, the loss function $L(\theta, \delta(x))$ should also be integrated over $\mathcal{X}$, the entire space of x.

Given the prior distribution $\pi(\theta)$ and the distribution $f(x|\theta)$ of x, θ should be integrated in proportion to $\pi(\theta)$ and x in proportion to $f(x|\theta)$. Thus, the loss function can be written as:

$$r(\pi, \delta) = \mathbb{E}^{\pi}[R(\theta, \delta)] = \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) f(x\theta)\, dx \pi(\theta)\, d\theta$$

where $r(\pi, \delta)$ is called the risk function of δ.

4.3 Bayesian Estimation

The goal of the decision making model is to derive an "optimal" decision rule that provides trustors with rational decisions about trustees based on observations (evidence) x. Optimality is implemented by minimizing the risk function $r(\pi, \delta)$. The decision maker follows the decision rule that gives the smallest risk. However, the trustworthiness value θ is often unknown, so a problem arises regarding the situation under which the risk function is minimized.

A common choice is the minimax rule, which chooses the $\widetilde{\delta}$ that satisfies the equation:

$$\underset{\theta}{sup}\, r\left(\theta, \tilde{\delta}\right) = \underset{\delta}{inf}\underset{\theta}{sup}\, r(\theta, \delta)\ \underset{\theta}{sup}\, r\left(\theta, \tilde{\delta}\right) = \underset{\delta}{inf}\underset{\theta}{sup}\, r(\theta, \delta).$$

The minimax rule also fits our original intention to make decisions that reduce the risk of trustors under uncertainty.

As an implementation of the likelihood principle, the Bayesian paradigm satisfies the decision-related requirements for trust assessment. It not only quantifies the uncertainties and minimizes the decision making risk, which is crucial when making rational decisions, but it also smoothly incorporates the trustors' prior information about the trustees' trustworthiness. This is essential when the decision process is viewed in the context of long-term system operation: trustors continuously acquire new evidence that must be combined with their prior information when making new decisions.

5. Illustrative Example

This section illustrates the application of the decision making model. It examines the simplified decision making case involving the inference of the trustworthiness value of a trustee based on the observation x, for which $\mathcal{D} = \Theta$.

The evidence aggregator of the trustor collects values of the related attributes $E = (E_1, E_2, \ldots, E_p)$ and stores the values in the corresponding vector $x_i =$

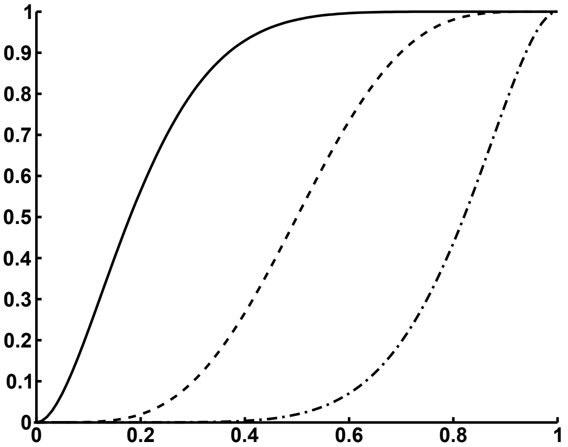

Figure 2. Cumulative distribution functions of three prior distributions.

$(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)$. Within a short time T, the evidence aggregator collects the n vectors forming $x = (x_1, x_2, \cdots x_n)$. Since the time T is short, we assume that $x_1, x_2, \cdots x_n$ are independent repeated trials from identical distributions f. According to the probabilistic model, the values of the attributes are conditional on the trustworthiness value θ, so the distribution can be denoted as $f(x|\theta)$.

As mentioned before, trust is subjective. For example, risk-averse trustors may tend to make negative decisions and risk-preferred trustors may tend to make positive decisions. The differences among trustors could be attributed to many factors. For instance, the difference might be due to the previous experience of trustors. A positive experience on the part of a trustor means that the trustor has made many correct decisions regarding trustworthy entities and makes the trustor more risk-preferred. Conversely, negative experience, which means that a trustor has made wrong decisions and trusted the wrong entities, makes the trustor more cautious. For one-dimensional evidence, this type of subjectivity can be modeled using a Beta-distribution with parameters α and β as the prior distribution of trustors. Let α be the number of past negative experiences and β be the number of past positive experiences. The prior information of trustors can be modeled as:

$$\pi(\theta) = Beta(\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt}$$

where $\pi(\theta)$ is the probability that the trustor decides to trust the trustee. Increasing the value of α makes the trustor more risk averse and increasing β makes the trustor more risk preferred.

As shown in Figure 2, a decision maker with $\alpha = 8$ and $\beta = 2$ (top line) tends to make negative trust decisions since the probability is high that trustworthiness values under 0.5 are allocated. On the other hand, a decision maker with $\alpha = 2$ and $\beta = 8$ (bottom line) is more likely to make positive decisions.

In this simplified example, since it is only necessary to estimate the value of θ, we select a commonly-used simple loss function, called the "quadratic loss function." This function is given by:

$$L(\theta, \delta) = (\theta - \delta)^2 .$$

The risk function is:

$$r(\theta, \delta) = \int_{\Theta} \int_{\mathcal{X}} (\theta - \delta)^2 f(x|\theta)\, dx \pi(\theta)\, d\theta.$$

The corresponding the computed estimator is:

$$\delta(x) = \frac{\int_{\Theta} \theta f(x|\theta)\, \pi(\theta)\, d\theta}{\int_{\Theta} f(x|\theta)\, \pi(\theta)\, d\theta}.$$

6. Related Work

The issue of trust is drawing increasing attention in the information security community. In 1996, Rasmussen and Jansson [18] examined the relationship between security and social control, and classified security mechanisms as: "soft security" such as trust and reputation systems, and "hard security" such as authentication and access control. Most security mechanisms include some aspects of trust, but they make implicit "trust assumptions" [7]. In order to overcome the drawbacks of current security mechanisms such as the inadequacy of authentication [5], a more general concept of trustworthiness should be engaged [1].

Trust management is largely associated with inference and decision making. Related evidence should be collected first and delivered to the trust management system as an input to the decision making model. Several trust management systems, such as PolicyMaker [5], KeyNote [4] and REFEREE [8], have been designed to collect security credentials and test the compliance of the credentials with security policies. Also, some trustworthiness computing models (e.g., [10]) collect trustors' prior experience as evidence and make predictions based on the experience. Some models collect evidence from other entities – these are essentially reputation systems [12, 13]. Generally, however, trust management systems and trustworthiness computing models [2, 9] attempt to determine a numerical trustworthiness value for a trustee or make a binary decision about whether or not a trustee is trustworthy. Our approach goes beyond this view by focusing on trust decision making as coupled with succeeding decision processes.

7. Conclusions

The Bayesian paradigm provides an elegant framework for incorporating trust into decision making processes associated with the control of large-scale critical infrastructure systems. The risk function, prior distribution and the distribution of evidence are three key components of the Bayesian paradigm. A prior distribution, which models the subjectivity of trustors, is combined with newly-acquired evidence and the derived Bayes risk function to obtain a decision rule by minimizing the risk function.

Although the mathematical structure of the framework is straightforward, its applicability depends on gaining experience with the kinds of data that are available in critical infrastructure systems and what the data say about trustworthiness. While it may not be clear, for example, what a particular ratio of good/bad past experience means for a particular decision, the framework shows what should be done with such data when it is collected.

Note that the views and opinions in this paper are those of the authors and do not necessarily reflect those of the United States Government or any agency thereof.

Acknowledgements

This research was partially funded by Department of Energy Award No. DE-OE0000097 (TCIPG) and by matching funds provided by Washington State University.

References

[1] A. Abdul-Rahman and S. Hailes, A distributed trust model, *Proceedings of the New Security Paradigms Workshop*, pp. 48–60, 1997.

[2] A. Abdul-Rahman and S. Hailes, Supporting trust in virtual communities, *Proceedings of the Thirty-Third Annual Hawaii International Conference on System Sciences*, vol. 6, 2000.

[3] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer, New York, 1985.

[4] M. Blaze, J. Feigenbaum and A. Keromytis, KeyNote: Trust management for public-key infrastructures, *Proceedings of the Sixth International Workshop on Security Protocols*, pp. 59–63, 1998.

[5] M. Blaze, J. Feigenbaum, and A. Keromytis, The role of trust management in distributed systems security, in *Secure Internet Programming (LNCS 1603)*, J. Vitek and C. Jensen (Eds.), Springer, Berlin, Germany, pp. 185–210, 1999.

[6] E. Chang, P. Thomson, T. Dillon and F. Hussain, The fuzzy and dynamic nature of trust, *Proceedings of the Second International Conference on Trust, Privacy and Security in Digital Business*, pp. 161–174, 2005.

[7] B. Christianson and W. Harbison, Why isn't trust transitive? *Proceedings of the International Workshop on Security Protocols*, pp. 171–176, 1997.

[8] Y. Chu, J. Feigenbaum, B. LaMacchia, P. Resnick and M. Strauss, REFEREE: Trust management for web applications, *Computer Networks and ISDN Systems*, vol. 29(8–13), pp. 953–964, 1997.

[9] W. Conner, A. Iyengar, T. Mikalsen, I. Rouvellou and K. Nahrstedt, A trust management framework for service-oriented environments, *Proceedings of the Eighteenth International Conference on the World Wide Web*, pp. 891–900, 2009.

[10] M. Denko, T. Sun and I. Woungang, Trust management in ubiquitous computing: A Bayesian approach, *Computer Communications*, vol. 34(3), pp. 398–406, 2011.

[11] S. French, *Decision Theory: An Introduction to the Mathematics of Rationality*, Ellis Horwood, New York, 1986.

[12] A. Josang, R. Ismail and C. Boyd, A survey of trust and reputation systems for online service provision, *Decision Support Systems*, vol. 43(2), pp. 618–644, 2007.

[13] S. Kamvar, M. Schlosser and H. Garcia-Molina, The Eigentrust algorithm for reputation management in P2P networks, *Proceedings of the Twelfth International Conference on the World Wide Web*, pp. 640–651, 2003.

[14] Y. Liu, P. Ning and M. Reiter, False data injection attacks against state estimation in electric power grids, *Proceedings of the Sixteenth ACM Conference on Computer and Communications Security*, pp. 21–32, 2009.

[15] North American Electric Reliability Corporation, Cyber Security Standards CIP-002-4 through CIP-009-4, Washington, DC (www.nerc.com/page.php?cid=2—20), 2011.

[16] A. O'Hagan, Bayesian statistics: Principles and benefits, in *Bayesian Statistics and Quality Modeling in the Agro-Food Production Chain, Volume 3*, M. van Boekel, A. Stein and A. van Bruggen (Eds.), Springer, Berlin, Germany, pp. 31–45, 2004.

[17] J. Pratt, H. Raiffa and R. Schlaifer, *Introduction to Statistical Decision Theory*, MIT Press, Cambridge, Massachusetts, 1995.

[18] L. Rasmusson and S. Jansson, Simulated social control for secure Internet commerce, *Proceedings of the New Security Paradigms Workshop*, pp. 18–25, 1996.

[19] C. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer, New York, 2007.