An evidence-based Bayesian trust assessment framework for critical-infrastructure decision processing

Yujue Wang and Carl Hauser

January 20, 2011

School of EECS, Washington State University, Pullman, Washington 99163
yujue.wang@email.wsu.edu, hauser@eecs.wsu.edu

1 Introduction

Currently the U.S. power grid is on the cusp of a tremendous expansion in the amount of sensor data that is available to support its operations. For decades the power grid has been operated using supervisory control and data access systems that poll each sensor once every two or four seconds—a situation that some in the industry have characterized as “flying blind.” Now, widespread deployment of sensing systems called phasor measurement units (PMUs) that provide accurately time-stamped data 30, 60, or more times each second is near at hand. By the end of the year 2013, utilities, with assistance of the American Recovery and Reinvestment Act of 2009 (ARRA) will have increased the number of these devices on the grid to nearly 1000, roughly an order of magnitude increase over what exists today. Data from PMUs and other high-rate sensing devices will be used to support new control schemes in support of 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 new control schemes using these data, the security of the data and their delivery, especially availability and integrity, but to some degree confidentiality as well, is of great concern. The security challenges will become even more difficult as the number of sensors increases and they become more widely deployed under the control of various entities throughout the transmission and distribution systems, extending even to micro-grids in the future. Good security practices and technologies such as those required by the NERC CIP standards will be even more essential to reliable grid operations than they are today.

Our thesis in this work is that regardless of the quality of the conventional security mechanisms used in such a system, the scale of the system and operational realities associated with large numbers of sensors and people spread over a wide geographic area and under diverse management means that conventional security mechanisms can provide only uncertain security.

For example, when 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 about this, it is in fact quite likely in a large-scale system that some of the bindings are incorrectly known at least some of the time by some entities, whether due to mistake or malicious manipulation.

If this thesis is true, then the reliability of the system will either come down to blind faith—*we know the 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. In the remainder of this paper we describe our work thus far 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 ones that rely on insights of humans. Starting with this goal—computational decision making in the face of uncertainty—leads to the vast and expanding literature on decision theories that are used in business strategy and operations, military planning, etc. 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 (the trustor's) reliance on and belief in the performance of another party (the trustee); for example, trust in a PKI certifier to correctly bind a public key to some other entity. The reliance or belief often must occur without certainty or may be in the form of a prediction about the future (itself a source of uncertainty). Trust, however, need not be blind: trustors can use evidence, for example in the form of past experience with a trustee, reputation information, or contracts and laws that impose penalties for non-performance, to form their trust judgments. We believe that if critical infrastructures are to be resilient against attacks it is essential that operational decision making processes *appropriately* take into account evidence about the trustworthiness of their input data. As we will show in the next section, using evidence appropriately means that it is considered in light of the particular decision being made: there is no single approach to judging trust that is universally appropriate.

The contributions of this paper are, first, establishing the need for a systematic way of dealing in critical infrastructure control systems with uncertainty related to trust and, second, an initial theoretical framework, based on Bayesian decision theory, for incorporating trust-related evidence to address this need. The framework suggests a number of data acquisition needs that would be required for its use, a point to which we return in section 5.

1.1 A motivating analogy

The credit reporting and scoring system for consumer credit (Fig. 1) provides an interesting analogy for evidence-based decision making in the

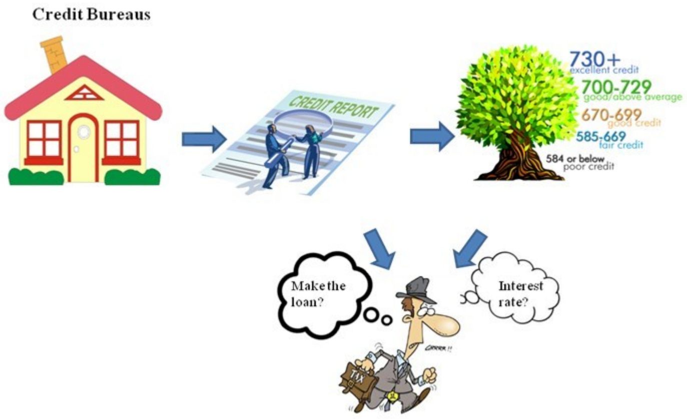

Figure 1: Credit Reporting System

presence of risk and illustrates some of the issues. Credit bureaus collect information from various sources and provide credit reports that detail individual consumers' past behavior in borrowing and bill paying. Some companies further analyze the information in credit reports from multiple sources to produce a single, numerical credit score based on statistical analysis of a person's credit reports. The credit score is claimed to statistically represent the creditworthiness of an individual.

Now consider the decisions lenders make in analyzing a loan application: they have to decide whether or not to make the loan and on what terms. If the loan is made, a lender stands to make a profit if the borrower pays it back, or a loss if the borrower defaults on 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 expected loss (maximizes expected return) by assessing the probability of different future borrower behaviors. To do this they turn to the credit report or credit score as well as information about employment, income, and stability of residence contained in the loan application.

There are several important things to point out in this analogy.

- First, different lenders will 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 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: for 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 whole portfolio of loans. Power grid operational decisions' consequences cannot be easily aggregated, so in this domain the decision processes will emphasize analysis of individual decisions.

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

1.2 Preliminaries

The distributed control system for a large-scale critical infrastructure (such as the power grid) can be described abstractly as consisting of a collection of controllers, a collection of data sources, and a collection of actuators. Actuators and controllers share the essential characteristics, for our purposes, of dealing with uncertainty of security so we will focus in what follows on controllers and information sources. In the power grid, for example, controllers are things like protective relays, automatic generator controls, remedial action schemes, etc. Data sources include sensors, human operators, and outputs of controllers. 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 those data, with the decisions ultimately being reflected in an action that changes the physical state of the grid in some way.

Because of noise in sensor outputs, in today's system inputs are assumed probabilistically related to the actual state of the sensed world by considering that each measurement consists of the actual state plus a normally-distributed noise term. Failures in the system can lead to bad inputs (highly improbable in the normally-distributed-noise model) which can often be detected and excluded by bad-data detection algorithms that exploit redundancy present in the inputs. Several recent papers have addressed ways that input data streams might be intentionally attacked invisibly to the bad data detectors in use today [15, 6].

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 commanded actions. The uncertainties associated with these aspects as well as with outcomes are modeled probabilistically though with much greater flexibility than afforded by the normally-distributed-noise approach currently used, and with explicit incorporation of uncertain results in the form of loss functions.

2 The Bayesian decision model

Decision theory studies the values and uncertainties related to making rational and optimal decisions [12]. Statistical theory has been widely applied to decision theory and is a common tool for decision making problems [17]. Our method is based on the Bayesian statistical paradigm which can quantify the uncertainties of decisions using personal probability [16]. A systematic introduction to Bayesian decision theory can be found in [19].

As previously noted, uncertainty is inherent in complex systems and 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 the system should have the ability to incorporate evidence about the trustworthiness of other entities and be inclined to rely on more-trustworthy peers. To begin formalizing this viewpoint 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, together forming the trust evidence. Focusing on a single entity $\mathcal{A}$, at a certain time point, it could collect the current evidence about a certain entity $\mathcal{B}$ which can be denoted as $x_i = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p) \in \mathbb{R}^p$. Over a period of time it will collected a number of x_is denoted $x = (x_1, x_2, \cdots x_n)$. Based on x, $\mathcal{A}$ will make a decision $d \in \mathcal{D}$ (where $\mathcal{D}$ is the decision space) on $\mathcal{B}$ in light $\mathcal{A}$'s estimate of the value of θ $(0 \leq \theta \leq 1)$ from the parameter space Θ which is called the trustworthiness to be placed on $\mathcal{B}$. Essentially, θ is probability that $\mathcal{B}$ is trustworthy.

In the current model, the decision-making process is considered as a choice of action made by the decision maker among a set of alternatives according to their possible consequences. In the 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 get accurate values of the evidence due to the complexity and uncertainty of the system. Probabilistic modeling is a natural choice both 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, $x_i \sim f_i\,(x_i|\theta, x_1, \cdots, x_{i-1})$ 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 the distribution can simply be denoted as

$$x \sim f\,(x|\theta)$$

The *likelihood function* l defined as

$$l\,(\theta|x) = f\,(x|\theta)$$

is equal to f but 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 inference of θ is contained in the likelihood function $l\,(\theta|x)$ and the value of θ can be inferred from x. Decisions can be made based on the inferred value of θ. To combine these processes, when the likelihood function $l\,(\theta|x)$ is fixed, a function from $\mathcal{X}$ to $\mathcal{D}$ can be obtained as $\delta\,(x)$ which is called the decision rule as it relates to trust. (Keep in mind 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).

In the remainer of this section we describe the elements involved in a Bayesian determination of decision rule $\delta\,(x)$, namely *prior distributions* and *loss functions,* and then state the derived rule.

2.1 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 on the trustworthiness value θ of a trustor regarding a trustee *before* receiving evidence is modeled using a probability distribution $\pi(\theta)$ on Θ, called the *prior distribution.* Subjectivity of trust is naturally modelled by different prior distributions.

2.2 The loss function

While it is easy to talk about making "good" decisions, the model requires a precise formalization of the notion of goodness. All of the possible choices in a decision should be ordered or quantified. Decision theory uses the *loss function* for this purpose. A loss function is any function $L \geq 0$ from $\Theta \times \mathcal{D}$ to $\mathbb{R}^p$ and represents the penalty $L(\theta, d)$ associated with the decision d when the parameter takes the value θ. In our situation, the penalty $L(\theta, d)$ 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 [7]. So it is important for the model to reflect such uncertainty. A simple way to obtain the loss is to integrate over all of the possible values of θ. What's more, instead of focusing on evaluating one decision, our goal is to assess a decision rule $\delta(x)$ which is the allocation of a decision to each outcome $x \sim f(x|\theta)$, so the loss function $L(\theta, \delta(x))$ should also be integrated on $\mathcal{X}$ which is the whole space of x.

Given the prior distribution, $\pi(\theta)$, and the distribution of x, $f(x|\theta)$, θ should be integrated in proportion to $\pi(\theta)$ and x in proportion to $f(x|\theta)$. So 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.

2.3 The Bayesian estimator

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

A common choice for Bayes paradigm is the minimax rule which chooses the $\widetilde{\delta}$ that satisfies $\underset{\theta}{sup}\, r\left(\theta, \tilde{\delta}\right) = \underset{\delta}{inf}\underset{\theta}{sup}\, r\left(\theta, \delta\right)$. Moreover, the minimax rule also fits for our original intention which is to make decisions that reduce the risk of the 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 uncertainties and minimizes the risk in decision-making, which is a crucial to make rational decisions, but also smoothly incorporates trustors' prior information about the trustees' trustworthiness. This is essential when the decision process is viewed in the context of long term operation of the system: trustors continuously acquire new evidence that must be combined with their prior information when making new decisions.

3 A simple example

In this section, we give an example of the decision-making model. We examine the simplified decision-making case with the goal of inferring the trustworthiness value of a trustee based on the observation x: so $\mathcal{D} = \Theta$.

The evidence aggregator of the trustor collects values of the related attributes $E = (E_1, E_2, \ldots, E_p)$ and stores these values in the corresponding vector $x_i = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)$. Within a short time, T, this evidence aggregator will collect the n vectors like x_i and form $x = (x_1, x_2, \cdots x_n)$. Since 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 modeling, the values of the attributes are conditional on the trustworthiness value θ, so the distribution can be denoted as $f\left(x|\theta\right)$.

As we said 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 attributed to many factors. For instance, the difference might be attributed to former experience of the trustors: positive experience, which means that the trustor made many correct decisions on trustworthy entities, will make the trustors more risk-preferred. Conversely, negative experience, which means that trustors made wrong decisions and trusted the wrong entities, will make the trustors more cautious. For one-dimensional evidence, this particular kind of subjectivity can be modelled 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 the number of past positive experiences. The prior information of trustors can be modeled as:

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

where $\pi\left(\theta\right)$ is the probability that trustor will decide to trust the trustee. Increasing α makes the trustor more risk-averse and increasing β makes the trustor more risk-preferred.

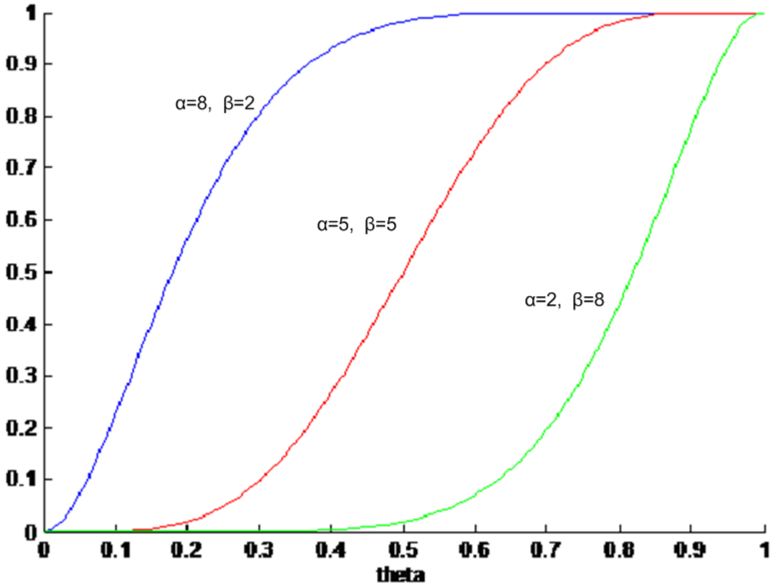

Figure 2: CDF of three different prior distributions

As Figure 2 shows, the decision maker with $\alpha = 8$ and $\beta = 2$ (top line in the diagram) will tend to make negative trust decisions since the probability that it allocates trustworthiness values under 0.5 is high. The decision maker with $\alpha = 2$ and $\beta = 8$ (bottom line in the diagram) is more likely to make positive decisions.

For this simplified example, since we just want to estimate the value of θ, we select a commonly-used simple loss function—the quadratic loss function:

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

The risk function would be

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

for which the computed estimator is

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

4 Related work

Trust in the information security area is drawing increasing attention. In 1996, Rasmussen and Jansson stated the relationship between security and social control and classified security mechanisms as: *soft security* such as trust and reputation systems and *hard security* like authentication and access control [18]. Actually, typical security mechanisms include some aspects of trust, but they make explicit "trust assumptions" [8]. In order to overcome some drawbacks of the current security mechanisms such as the inadequacy of authentication [4], a more general concept of “trustworthiness” should be managed [1].

Trust management is largely associated with inference or decision making. Related evidence should be collected first and delivered to the trust management system as input for the decision making model. Several trust management systems such as PolicyMaker [4], KeyNote [5], and REFEREE [9] were designed to collect security credentials and test the compliance of the credential with security policies. Also, some trustworthiness computing models [11] collect trustors’ former experience as evidence and make predictions based on this former experience. Some models collect evidence from other entities–these are essentially reputation systems [14, 13]. Generally, however, current trust management systems or trustworthiness computing models [10, 2] set their goal as determining a numerical trustworthiness value for a trustee or making a binary decision about whether a trustee is trustworthy or not. We go beyond this viewpoint by looking at trust decision making as coupled to succeeding decision processes.

5 Conclusions and future work

We have described a framwork for incorporating trust into the decision making processes associated with control of large-scale critical infrastructure systems. Our framework is based on the Bayesian paradigm. The risk function, prior distribution and the distribution of evidence are three components of the Bayesian paradigm. We used the prior distribution to model subjectivity of trustors and showed how it could be combined with newly-acquired evidence and the derived Bayes risk function to obtain a decision rule by minimizing the risk function.

Though the mathematical structure of the framework is straightforward, its practical applicability depends on gaining experience with the kinds of data that are available in critical infrastructure systems and what those data say about trustworthiness. It is not clear for example what a particular ratio of good/bad past experience means for a particular decision, but the framework tells us what to do with such data when it is collected.

6 Acknowledgements

This research was funded in part by Department of Energy Award Number DE-OE0000097 (TCIPG) and by matching funds provided by Washington State University.

Disclaimer: "This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof."

References

[1] A. Abdul-Rahman and S. Hailes. A distributed trust model. In *Proceedings of the 1997 workshop on New security paradigms*, pages 48–60. ACM, 1998.

[2] Alfarez Abdul-Rahman and Stephen Hailes. Supporting trust in virtual communities. *Hawaii International Conference on System Sciences*, 6:6007, 2000.

[3] J.O. Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985.

[4] M. Blaze, J. Feigenbaum, J. Ioannidis, and A. Keromytis. The role of trust management in distributed systems security. *Secure Internet Programming*, pages 185–210, 1999.

[5] M. Blaze, J. Feigenbaum, and A. Keromytis. KeyNote: Trust management for public-key infrastructures. In *Security Protocols*, pages 625–625. Springer, 1999.

[6] Rakesh B. Bobba, Katherine M. Rogers, Qiyan Wang, Himanshu Khurana, Klara Nahrstedt, and Thomas J. Overbye. Detecting false data injection attacks on DC state estimation. In *First Workshop on Secure Control Systems (SCS 2010)*, April 2010.

[7] Elizabeth Chang, Patricia Thomson, Tharam Dillon, and Farookh Hussain. The fuzzy and dynamic nature of trust. In Sokratis Katsikas, Javier Lopez, and Günther Pernul, editors, *Trust, Privacy and Security in Digital Business*, volume 3592 of *Lecture Notes in Computer Science*, pages 161–174. Springer Berlin / Heidelberg, 2005. 10.1007/11537878_17.

[8] B. Christianson and W. Harbison. Why isn't trust transitive? In *Security Protocols*, pages 171–176. Springer, 1997.

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

[10] William Conner, Arun Iyengar, Thomas Mikalsen, Isabelle Rouvellou, and Klara Nahrstedt. A trust management framework for service-oriented environments. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 891–900, New York, NY, USA, 2009. ACM.

[11] M.K. Denko, T. Sun, and I. Woungang. Trust management in ubiquitous computing: A Bayesian approach. *Computer Communications*, 2010.

[12] Simon French. *Decision theory: an introduction to the mathematics of rationality*. Halsted Press, New York, NY, USA, 1986.

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

[14] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*, WWW '03, pages 640–651, New York, NY, USA, 2003. ACM.

[15] Yao Liu, Michael K. Reiter, and Peng Ning. False data injection attacks against state estimation in electric power grids. In *Proceedings of the 16th ACM conference on Computer and communications security*, CCS '09, pages 21–32, New York, NY, USA, 2009. ACM.

[16] A. O'Hagan. Bayesian statistics: principles and benefits. *Wageningen UR Frontis Series*, 3(0):31, 2004.

[17] J.W. Pratt, H. Raiffa, and R. Schlaifer. *Introduction to statistical decision theory*. The MIT Press, 1995.

[18] L. Rasmusson and S. Jansson. Simulated social control for secure Internet commerce. In *Proceedings of the 1996 workshop on New security paradigms*, pages 18–25. ACM, 1996.

[19] C.P. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Verlag, 2007.