Evaluating Intrusion Detection Systems for Energy Diversion Attacks

by

Abhishek Rajkumar Sethi

B.Tech., Indian Institute of Technology Bombay, India (2013)

Submitted to the School of Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Computation for Design and Optimization

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016

Author . . . .
School of Engineering
August 5, 2016

Certified by . . . .
Saurabh Amin
Robert N. Noyce Career Development Assistant Professor
Thesis Supervisor

Accepted by . . . .
Youssef Marzouk
Associate Professor of Aeronautics and Astronautics
Co-Director, Computation for Design and Optimization

Evaluating Intrusion Detection Systems for Energy Diversion Attacks

by

Abhishek Rajkumar Sethi

Submitted to the School of Engineering
on August 5, 2016, in partial fulfillment of the
requirements for the degree of
Master of Science in Computation for Design and Optimization

Abstract

The widespread deployment of smart meters and ICT technologies is enabling continuous collection of high resolution data about consumption behavior and health of grid infrastructure. This has also spurred innovations in technological solutions using analytics/machine learning methods that aim to improve efficiency of grid operations, implement targeted demand management programs, and reduce distribution losses. One one hand, the technological innovations can potentially lead large-scale adoption of analytics driven tools for predictive maintenance and anomaly detection systems in electricity industry. On the other hand, private profit-maximizing firms (distribution utilities) need accurate assessment of the value of these tools to justify investment in collection and processing of significant amount of data and buy/implement analytics tools that exploit this data to provide actionable information (e.g. prediction of component failures, alerts regarding fraudulent customer behavior, etc.)

In this thesis, the focus on the value assessment of intrusion/fraud detection systems, and study the tradeoff faced by distribution utilities in terms of gain from fraud investigations (and deterrence of fraudulent customer) versus cost of investigation and false alarms triggered due to probabilistic nature of IDS. Our main contribution is a Bayesian inspection game framework, which models the interactions between a profit-maximizing distribution utility and a population of strategic customers. In our framework, a fraction of customers are fraudulent - they consume same average quantity of electricity but report less by strategically manipulating their consumption data. We consider two sources of information incompleteness: first, the distribution utility does not know the identity of fraudulent customers but only knows the fraction of these consumers, and second, the distribution utility does not know the actual theft level but only knows its distribution.

We first consider situation in which only the first source of information incompleteness is present, i.e., the distribution utility has complete information about the actual theft level. We present two simultaneous game models, which have same assumption about customer preferences and fraud, but differ in the way in which the distribution utility operates the IDS. In the first model, the distribution utility probabilistically chooses to use IDS with a default (fixed) configuration. In the second model, the distribution utility can configure/tune the IDS to achieve an optimal operating point (i.e. combination of detection probability and false alarm rate). Throughout, we assume that the theft level is greater than cost of attack. Our results show that for, the game with default IDS configuration, the distribution utility does not use the IDS in equilibrium if the fraction of fraudulent customers is less than a critical fraction. Also the distribution utility realizes a positive "value of IDS" only if one or both have the following conditions hold: (a) the ratio of detection probability and false alarm probability is greater than a critical ratio, (b) the fraction of fraudulent customers is greater than the critical fraction. For the tunable IDS game, we show that the distribution utility always uses an optimal configuration with non-zero false alarm probability. Furthermore, the distribution utility does not tune the false alarm probability when the fraction of fraudulent customers is greater than a critical fraction. In contrast to the game with fixed IDS, in the game of tunable IDS, the distribution utility realizes a positive value from IDS, and the value increases in fraction of fraudulent customers.

Next, we consider the situation in which both sources of information incompleteness are present. Specifically, we present a sequential game in which the distribution utility first chooses the optimal configuration of the IDS based on its knowledge of theft level distribution (Stage 1), and then optimally uses the configured IDS in a simultaneous interaction with the customers (Stage 2). This sequential game naturally enables estimation of the "value of information" about theft level, which represents the additional monetary benefit the distribution utility can obtain if the exact value of average theft level is available in choosing optimal IDS configuration in Stage 1. Our results suggest that the optimal configuration under lack of full information on theft level lies between the optimal configurations corresponding to the high and low theft levels. Interestingly enough, our analysis also suggests that for certain technical (yet realistic) conditions on the ROC curve that characterizes achievable detection probability and false alarm probability configurations, the value of information about certain combination of theft levels can attain negligibly small values.

Thesis Supervisor: Saurabh Amin
Title: Robert N. Noyce Career Development Assistant Professor

Acknowledgments

It is a pleasure to acknowledge the contributions of many people who have made this thesis possible. I am very grateful to my research supervisor Professor Saurabh Amin for his continuous support of my study and research, and for sharing his enthusiasm and knowledge with me. I thank him for teaching me game theory and microeconomics, for having faith in my ability, and for instilling in me a passion to pursue knowledge.

I would like to acknowledge the financial support of FORCES (Foundations Of Resilient Cyber-Physical Systems), which receives support from the National Science Foundation (NSF award numbers CNS-1238959, CNS-1238962, CNS-1239054,CNS-1239166), the AFRL LABLET-Science of Secure and Resilient Cyber-Physical Systems (Contract ID: FA8750-14-2-0180, SUB 2784-018400), and the NSF CAREER award #1453126.

My friends Devendra Shelar and Tuhin Sarkar deserve special thanks for being there for me when I needed them the most. Furthermore I want to thank Srivatsa, Nirav, Akshay, Siya, Manxi, Srivatsan, Parnika, Krithika, Anil, Ashwin, Saurabh, Ankit, Ananth, Abhishek, Rohit, Dinesh, Zaid, Li Jin, Mathieu, Jeff, Derek and Lina for being splendid friends.

I am grateful to my parents, Mr. Raj Kumar Sethi and Mrs. Nita Sethi, and sister Ms. Rashmi Sethi for their support, love and understanding. I also want to thank my aunt Mrs Khanna and all my cousins for their continuous support. Lastly, and most importantly, I wish to thank my grandfather (Late) Shri Laxman Das Sethi. Nothing I can say will do justice to how I respect and feel for his love and support. To him, I dedicate this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Historically, widespread energy theft is characteristic for developing countries. Indeed, according to a World Bank report [3], the theft of electricity reaches up to 50in some jurisdictions of developing countries. For example, according to [34], India loses more revenue to theft than any other country in the world with one of the states(Maharashtra) alone losing $2.8 billion per year. Overall the country has transmission and distribution(T&D) losses of up to 23% making the electricity distribution business financially unsustainable.

Electricity Theft can have several manifestations. For example, it can occur via availing unauthorized/unrecorded supply by tapping into conductors, feeders, and tampering service wires. Or customers can commit theft by damaging and manipulating electric equipment. Moreover there have been instances when the distribution utility's employers have made intentional billing errors in favor of fraudulent customers. However electricity theft is not the only component of non-technical loss faced by the distribution utility. In addition to electricity theft, energy diversion losses (and equivalently non-technical losses) can occur due to actions of a utility personnel(administrative losses), customer non-payment and theft by outsiders.

Non technical losses adversely affect the efficiency of distribution system. For example, the electricity consumption of non-paying parties is paid by other member of the society. Moreover nontechnical losses are usually covered via either higher electricity tariffs or higher taxes if the government decides to subsidize the distributor for these losses. Sometime these losses are not compensated for prolonged periods of time. Lastly, when the distribution utility ends up being the net bearer of losses for prolonged periods of time and no regulatory measures exist to recover these losses, his incentive to innovate and invest in the network and its maintenance are jeopardized.

In order to reduce non-technical losses, the distribution utility and the regulation bodies in developing countries should combat losses at each channel [2]. The distribution utility can advance hardware technology initiatives by installing IT-supported meters at distribution transformer and feeders, Automated Meter Reading (AMR), and Advanced Metering Infrastructures(AMI). Moreover the distribution utility can adopt management information systems equipped with data analytics for efficient detection of fraud and unmetered connections. Furthermore the regulatory body can impose strengthened enforcement mechanisms for efficient prosecution of theft and publicize theft for sharper public scrutiny. Lastly, the distribution utility can fix skewed tariff structures and ensure coordination and transparency in distribution efforts.

In this thesis, we want to build a business case for the distribution utility to leverage high resolution data to improve metering, billing and collection procedures, and fraud/theft identification complementing traditional detection tools of using balance meters and physical checks of tamper-evident seals by field personnel. As mentioned earlier, one of the most important technologies on the consumption side of smart grid is the Advanced Metering Infrastructure (AMI). AMI are crucial to the modernization of electricity metering system by replacing old meters by smart meters. Smart meters provide two-way communications between the utility and the customer. Moreover AMIs provide several capabilities including monitoring of network-wide and individual electricity consumption, faster remote diagnosis of outages, remote disconnect options and automated power restoration [1].

However installation of AMI alone is not the full answer to the widespread problem of electricity theft. Firstly, while AMI's are smart, the fraudulent customers are smarter. For example, many customers tap power lines in front of the smart meters thereby hindering the ability to identify unmetered consumption. In fact, according to [23],

the FBI has seen a surge in hacking of smart meters. Moreover the article mentions that one does not need sophisticated technological skills to compromise these meters and the fraudulent customers can achieve 50-75% reduction in power by simply placing a magnet to fake readings or hiring hackers to break into these meters. Secondly, although AMI's offer crucial technological advantages, their installation implies that investigation and utility personals are not visiting residential premises. As a result, the fraudulent customers are less deterred and more bold in committing theft. [38]
As discussed earlier, distribution utilities across the US are collecting fine-grained data from their networks, devices and consumers [25]. In fact, electricity customers generating as much data on the smart grid as they are on social media [21]. Due to this surge in data collection and rapid progress in AI and big data capabilities, there has been a proliferation in literature on data-driven anomaly detection systems. The focus of such works have been on detection of abnormal electricity traces that are highly correlated with electricity theft. Moreover, these works use a variety of machine learning techniques, including Support Vector Machines and Extreme Learning Machines to identify suspicious energy traces [33], [32], [16]. Furthermore there has been efforts to enable sensor fusion at the scale of electricity distribution to integrate consumption anomalies in the diagnostic system [30]. Lastly, there has been research in evaluating a class of theft detection schemes in the presence of strategic fraudulent customers who can evade the diagnostic system [27]. The survey article [20] provides a broader description of electricity theft problem.
However there has not been significant number of studies in the smart grid community about the value of these technologies. In this thesis, we evaluate the value of Intrusion Detection System (IDS), characterized by false alarm probability and detection probability, to the distribution utility for addressing non-technical losses.

1.1 Related Work

Though we are unaware of similar analysis in the energy diversion literature, researchers in other areas have investigated similar problems. Particularly, our work is closely related to Inspection Games. Inspection games are a class of attacker-defender games that account for illegal actions by a strategic inspectee who wants to evade detection by an inspector. The survey paper by Avenhaus et al. [5] provides an excellent summary of this topic. A fundamental feature of inspection games is that the inspector tries to prevent the inspectee from operating illegally in terms of violating an agreement or a set of rules. The problem is to design an optimal inspection scheme when the inspector is resource-constrained, and the detection can only be partial. This class of games is of particular interest to energy diversion problems, because it allows to extend the standard formulation of statistical detection tests to the settings when the inspectee is able to manipulate the observations collected by the inspector.

Similarly, there has been applications of Inspection games for security of IT architecture using Intrusion Detection Systems (IDS) in [12]. The work assess the value of IDS in firm's IT security distinguishing between its 'out-of-the-box' and 'optimal' configuration and finally shows that an IDS produces positive value only when the detection rate is higher than a critical value, which is determined by the attacker's cost parameters. Although the above framework addresses the optimal configuration of the IDS, it does not model the relationship between IDS detection probability and attacker's efforts. The above limitation is addressed by [13] which presents a systematic framework for comparing game-theoretic and decision-theoretic approaches to IT security based on investment levels, vulnerability, and equilibrium payoff. . The above model also compares the timing of the game showing that the sequential game results in the maximum payoff to the defender, followed by simultaneous game which is better than decision theoretic approach for most cases. However the model assumes a functional form for security breach probability based on attacker's effort without any implementation details. Furthermore both the models assume that all users/customers are attackers and does not reflect the property of incomplete information between the defender and users. Since majority of electricity customers are genuine, incomplete information about the type of customer is a crucial property of any game theoretical model for electricity theft.

To model the heterogeneity in customers, [26] proposes a game theoretic framework to analyze the interactions between pairs of attacking/defending nodes using a Bayesian formulation and evaluates Nash Equilibrium for both static and dynamic scenarios. However the model does not address (a) the optimal configuration of the IDS and (b) the interaction between IDS and attacker's efforts. Lastly, [1] investigates energy theft in smart utility networks by developing a leader-follower game-theoretic framework to model the interactions between distribution utility and population of strategic customers, a fraction of which are fraudulent customers. Furthermore the work evaluates pricing and investment decisions by a distribution utility in both unregulated monopoly and perfect competition environment. Although the model presents a systematic procedure to model electricity theft, it evaluates the equilibrium for only a certain type of ROC curves(namely, for exponential distribution) and assumes a functional relationship between level of investment by the distribution utility and false alarm probability of the IDS without any implementation details. Moreover the focus of this work has been on pricing of electricity and level of investment in AMI-based anomaly detection systems while we are more interested in configuration of these anomaly detection systems and evaluating the value of information in decision making for the distribution utility.

1.2 Focus / Our contribution

In this thesis, we present a *bayesian inspection game* framework to analyze the problem of electricity theft (and more broadly, to energy diversion attacks). In this model, the electricity distribution utility faces two types of consumers: genuine consumers and fraudulent ones. A genuine consumer's billed consumption of electricity is the same as her genuine consumption on average. However, the fraudulent consumers strategically manipulate their meter readings to under-report their actual consumption. In the absence of detection, the fraudulent consumers can continue to divert electricity, resulting in increased commercial losses to the distribution utility, and in some cases tariff increases for the genuine consumers.

In order to model the investment and operational decision-making process by the distribution utility, we consider a two stage formulation of its interaction with the customers. In stage 1, the distribution utility chooses the optimal configuration of the IDS and in stage 2, the distribution utility chooses the probability of using the IDS and the fraudulent customer chooses the probability of committing theft. Note that all the game parameters are known to both the players in stage 2. For both default configuration of the IDS and a tunable IDS, the distribution utility skips stage 1 and chooses the probability of using the IDS and the optimal configuration respectively in stage 2. Furthermore the distribution utility may not have complete information about the theft level committed by the fraudulent customer in stage 1 and consequently chooses the optimal configuration that maximizes his expected payoff. In order to evaluate the relevance of IDS technologies in security of smart grid, we seek to asses the value of IDS for the distribution utility in all the aforementioned configurations. The value of IDS is crucial to evaluate the difference between subsequent investment in smart grid security technologies and the recovered fine and tariff from fraudulent customers by deterring them from committing energy theft.

Similarly the distribution utilities often have an option to collect additional information about the customer and his consumption behavior. Through our game theoretic framework, we want to determine conditions when it is profitable for the distribution utility to gain further information. As a result, we define value of information to be the difference in defender's payoff between complete and incomplete information scenarios. Finally we argue that the total value that the distribution utility obtains from installing/operating an IDS is equal to the sum of value of IDS and value of information.

We summarize the significant findings of our analysis on equilibrium response of the distribution utility and fraudulent customer as follows:

i. In the default configuration of IDS, we show that, in equilibrium, the fraudulent customers do not commit theft and the distribution utility does not use the IDS if amount of theft is less than cost of attack. Moreover we show, provided amount of theft is greater than cost of attack, that there exists a critical fraction of fraudulent customers below which the distribution utility does not use the IDS and the fraudulent customer always commits theft. Finally, for fractions greater than critical fraction above, the fraudulent customer is indifferent between Theft and No Theft and the distribution utility chooses a constant false alarm probability.

ii. For a tunable IDS, like the default configuration, in equilibrium, the fraudulent customers do not commit theft and the distribution utility chooses zero false alarm probability of the IDS if amount of theft is less than cost of attack. Furthermore, there exists another critical fraction of fraudulent customers below which the fraudulent customer always commits theft and the distribution utility chooses a non-zero false alarm probability of the IDS. Finally, like the default configuration, for fractions greater than critical fraction above, the fraudulent customer is indifferent between Theft and No Theft and the distribution utility chooses a constant false alarm probability.

iii. In the optimal configuration of the IDS under perfect information, we show that, in equilibrium, the fraudulent customers do not commit theft and the distribution utility chooses zero false alarm probability in Stage 1 and does not use the IDS in Stage 2 if amount of theft is less than cost of attack. Furthermore, for amount of theft greater than cost of attack, the distribution utility uses the IDS in Stage 2 with probability 1. Similarly, the false alarm probability chosen by the distribution utility in Stage 1 is same as that chosen for a tunable IDS.

iv. In the optimal configuration of the IDS under imperfect information, we show that in equilibrium, the distribution utility chooses in Stage 1 a false alarm probability between the optimal false alarm probabilities corresponding to the two theft levels.

We summarize the significant findings of our analysis on value of IDS and value of information as follows:

i. The distribution utility always realizes non-negative value of IDS in all valid configurations (i.e., default, tunable and optimally configured under both perfect and imperfect information) with detection probability greater than equal to false alarm probability.

ii. In the default configuration of the IDS, the distribution utility realizes a positive value if amount of theft committed by the fraudulent customer is greater than cost of attack and the ratio of detection probability and false alarm probability is greater than a critical ratio defined in terms of cost parameters and theft level. Additionally, given the amount of theft is greater than cost of attack, the distribution utility obtains positive value for all fractions of fraudulent customers greater than a critical fraction determined by IDS configuration, cost parameters and theft level.

iii. For a tunable IDS, the distribution utility realizes a positive value if amount of theft committed by the fraudulent customer is greater than cost of attack. Moreover the value of IDS increases with increasing fraction of fraudulent customers.

iv. For complete information, the value of an optimally configured IDS is same as that of a tunable IDS. For incomplete information, the expected value of optimally configured IDS is always less than expected value of a tunable IDS.

v. For incomplete information, the distribution utility always realizes non-negative value of information. Furthermore, given any probability measure over theft level and under certain technical (yet realistic) conditions on the ROC curve, there always exists two theft levels such that the distribution utility obtains zero value of information.

Outline:

In chapter 2, we present the game-theoretic model and describe the environment in which the distribution utility and the customers interact followed by classification of Intrusion Detection System(IDS) available to the distribution utility. In chapter 3, we evaluate the equilibrium response of the distribution utility and customers for

Intrusion Detection System Parameters

α	False Alarm probability
$\rho(\cdot)$	Detection probability

Consumer Model Parameters

f	Fraudulent Consumer
g	Genuine Consumer
q_j^i	Average quantity for $j \in \{actual, billed\}$ $i \in \{fraudulent, genuine\}$
η	Level of theft (or threat)
$\delta^{\mathrm{f}} \in \{\mathbf{T}, \mathbf{NT}\}$	**T**: Theft, **NT**: No Theft
π	Fraction of fraudulent customers
$v(.)$	Customer valuation function
C_η	Cost of Theft
γ	Probability of committing Theft

Distribution Utility Parameters

D	Distribution Utility
$\delta^{\mathrm{D}} \in \{\mathbf{I}, \mathbf{NI}\}$	**I**: Investigate , **NI** : Not Investigate
T(.)	Tariff Scheme (For ex: $T(q) = A + pq$)
C_ρ	Cost of Investigation
C_α	Cost of False Alarm
F	Fine
$c(.)$	Cost of Electricity $c(q) = cq$

Table 1.1: Game-Theoretic Model Parameters

different configurations of the IDS - fixed, tunable and optimal configuration under imperfect information. In chapter 4, we utilize the results on equilibrium payoffs of the games in chapter 3 and determine the value of IDS(fixed, tunable/customizable), and the value of information on the theft level. Finally, in chapter 5, we discuss the manager implications of our work and discuss the current practices and future directions for distribution utilities.

Chapter 2

The Model

In this chapter, we present a game-theoretic framework to describe the interactions between distribution utility and a population of electricity customers. An exogenously known fraction of customers have vulnerable meter connections, and hence commit electricity diversion attacks. The remaining fraction of customers are genuine, i.e. they pay to the electric distribution utility for their actual consumption billed according to a predefined tariff rate. Next we describe the environment in which the distribution utility and the customers interact and define the class of Intrusion Detection System(IDS) available to the distribution utility. In 2.2 we define the two games that are both subject to a set of assumptions on the parameters that affect the players payoff and the IDS.

2.1 Environment

2.1.1 Distribution Utility

Our goal is to model scenarios for electricity theft and evaluate the optimal inspection strategies for the distribution utility facing a population of customers, a fraction of which can strategically conduct energy diversion attacks(fraud). In our setup, the Distribution Utility(Monopolist) is the *defender* who collects consumption data (e.g. AMI reading) to decide whether or not to investigate a customer for fraudulent electricity consumption. We assume that the defender employs an Intrusion Detection System which utilizes the classical statistical hypothesis testing to generate alert for the defender.

Following the classical hypothesis testing paradigm, we say that the distribution utility decides the underlying consumption distribution ($\mathcal{H}_0$ or $\mathcal{H}_1$) of the customer by observing realizations(meter readings) of electricity consumption which is random. The IDS's detection performance is governed by its false alarm probability and probability of detection. Thus in our model, the distribution utility acts like the statistician in classical hypothesis testing problem.

We view electricity theft as the monetary loss due to stolen electricity by fraudulent customers. A fraudulent customer (denoted by f) reports reduced electricity consumption to the distribution utility while the genuine customer (denoted by g) reports his/her actual consumption. The main feature of our setup is that the distribution utility does not know the "type" of the customers. We will model this information incompleteness as a Bayesian Inspection game; see sec. 2.2.

In the terminology of Bayesian games, we say that the fraction of fraudulent customers is common knowledge and each customer is privately informed about her type (type f or type g). The distribution utility chooses to conduct inspection or not based on alarm raised by the statistical test. Thus, the distribution utility's choice is to either *Investigate* (**I**) or *Not Investigate* (**NI**) a customer $\delta^{\mathrm{D}} = \{\mathbf{I}, \mathbf{NI}\}$. We allow randomized strategy, i.e. the distribution utility can choose to investigate with a certain probability. As we will describe subsequently, the distribution utility's inspection strategy is directly related to the detection probability ρ and false alarm probability α of the IDS.

We assume that the distribution utility buys electricity from the electricity market at a marginal cost C and sells it to customers according to a fixed tariff scheme $\mathrm{T}(\cdot)$. For practical purposes and for the sake of convenience, we will consider a two-part tariff scheme. To deter electricity theft, the customers are subject to a monetary fine (F) if fraud is confirmed upon investigation by the distribution utility.

However, the distribution utility incurs false alarm cost C_α for falsely investigating

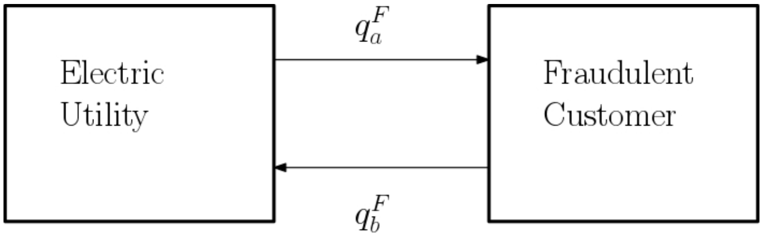


Figure 2.1: Interaction between a distribution utility and a fraudulent customer

a genuine customer for fraudulent behavior. Lastly, the distribution utility incurs a cost C_ρ to carry out an investigation for fraudulent behavior.

2.1.2 Customer

We now provide a (simple but non-exhaustive) classification of fraudulent behavior or energy diversion attacks that can be committed by strategic customers. Let q_a^{g} KWh denote the average quantity consumed by a genuine customer and q_b^{g} kWh the average quantity reported by a genuine customer. By definition, a genuine customer consumes and reports the same expected quantity, i.e $q_a^{\mathrm{g}} = q_b^{\mathrm{g}}$. Similarly let q_a^{f} kWh (resp. q_b^{f} kWh) is the average quantity consumed (resp. reported) by a fraudulent customer.

Naturally, $q_a^{\mathrm{f}} \neq q_b^{\mathrm{f}}$. The folllowing three cases arise:

i. Fraudulent customer reports lesser consumption but consumes same as the genuine customer, i.e. $q_b^{\mathrm{f}} < q_a^{\mathrm{f}} = q_a^{\mathrm{g}}$. In this case, the average stolen quantity of electricity, if undetected, is given by $(q_a^{\mathrm{f}} - q_b^{\mathrm{f}})$;

ii. Fraudulent customer consumes more than genuine customer, but reports same, i.e. $q_a^{\mathrm{f}} > q_a^{\mathrm{g}}$ and $q_b^{\mathrm{f}} = q_b^{\mathrm{g}}$;

iii. Fraudulent customer draws more electricity than the genuine one, but reports less than genuine customer, i.e. $q_a^{\mathrm{f}} > q_a^{\mathrm{g}}$ and $q_b^{\mathrm{f}} < q_b^{\mathrm{g}}$;

In our work, we focus on first case i.e., the fraudulent customer consumes the same expected quantity as a genuine customer but reports reduced consumption to the distribution utility.

We model the fraudulent customer's level of theft as a fraction $\eta \in [0,1]$, which is the ratio of stolen quantity and actual consumed quantity by the fraudulent customer, i.e. $\eta = (q_a^{\mathrm{f}} - q_b^{\mathrm{f}})/q_a^{\mathrm{f}} = 1 - q_b^{\mathrm{f}}/q_a^{\mathrm{f}}$. Equivalently, i.e. $q_b^{\mathrm{f}} = (1-\eta)q_a^{\mathrm{f}}$. The fraudulent customer chooses to either commit *Theft*(**T**) or *No Theft*(**NT**), i.e. $\delta^{\mathrm{f}} \in \{\mathbf{T},\ \mathbf{NT}\}$. For a mixed strategy $\gamma \in [0,1]$, the fraudulent customer randomizes between **T** and **NT**. Thus, γ can be viewed as the probability of committing theft **T**, $\mathrm{P}(\mathbf{T}) = \gamma$ and $\mathrm{P}(\mathbf{NT}) = 1-\gamma$. The fraudulent customer incurs cost of attack C_η if $\delta^{\mathrm{f}} = \mathbf{T}$ and pays a monetary fine F if the investigation by the distribution utility successfully detects theft.

A classical way to calculate q_a^{g} is to maximize genuine customer's utility as follows:

$$q_a^{\mathrm{g}*} = \arg\max_{q_a^{\mathrm{g}}} u^{\mathrm{g}}(q_a^{\mathrm{g}}, \mathrm{T}(.)) \tag{1}$$

where $u^{\mathrm{g}} := v(q_a^{\mathrm{g}}) - \mathrm{T}(q_b^{\mathrm{g}})$, and $v(\cdot)$ is the preference function for consuming electricity (assumed identical for all customers) and $\mathrm{T}(\cdot)$ is the tariff schedule. Similarly, for a given theft level η and $q_a^{\mathrm{g}} = q_a^{\mathrm{f}}$ as mentioned earlier, the fraudulent customer chooses γ to maximize his utility:

$$\begin{aligned} u^{\mathrm{f}*} :=& \max_{\gamma} v(q_a^{\mathrm{f}}) - (1-\gamma)\mathrm{T}(q_a^{\mathrm{f}}) - \gamma\mathrm{T}(q_b^{\mathrm{f}}) - \gamma\mathbb{E}_{\delta^{\mathrm{D}}}\mathcal{P}(q_a^{\mathrm{f}}, q_b^{\mathrm{f}}, \delta^{\mathrm{D}}) \\ =& \max_{\gamma} v(q_a^{\mathrm{g}*}) - (1-\gamma)\mathrm{T}(q_a^{\mathrm{g}*}) - \gamma\mathrm{T}(q_b^{\mathrm{f}}) - \gamma\mathbb{E}_{\delta^{\mathrm{D}}}[\mathcal{P}(q_a^{\mathrm{g}*}, q_b^{\mathrm{f}}, \delta^{\mathrm{D}})] \\ =& u^{\mathrm{g}*} + \max_{\gamma} \left(\gamma p(q_a^{\mathrm{g}*} - q_b^{\mathrm{f}}) - \gamma\mathbb{E}_{\delta^{\mathrm{D}}}[\mathcal{P}(q_a^{\mathrm{g}*}, q_b^{\mathrm{f}}, \delta^{\mathrm{D}})]\right) \end{aligned} \tag{2}$$

where $\mathcal{P}$ be the fine imposed by distribution utility given the fraudulent customer is committing theft **T**

$$\mathcal{P}(q_a^{\mathrm{g}*}, q_b^{\mathrm{f}}, \delta^{\mathrm{D}}) = \begin{cases} \mathrm{F} + \mathrm{C}_\eta + p q_a^{\mathrm{g}*}\eta & \text{if } \delta^{\mathrm{D}} = \mathbf{I} \\ \mathrm{C}_\eta & \text{if } \delta^{\mathrm{D}} = \mathbf{NI} \end{cases}$$

Hence, from (2), the fraudulent customer chooses γ to maximize,

$$\max_{\gamma} \gamma \left(p q_a^{\mathrm{g}*} \eta - \mathbb{E}_{\delta^{\mathrm{D}}}[\mathcal{P}(q_a^{\mathrm{g}*}, q_b^{\mathrm{f}}, \delta^{\mathrm{D}})] \right)$$

Note that since fraudulent customer always obtains $u^{\mathrm{g}*}$ irrespective of $\gamma, \delta^{\mathrm{D}}$ and genuine customers obtain $u^{\mathrm{g}*}$ (by definition) , we do not consider it in the payoff functions and further analysis. Lastly, since q_a^{g} is the only quantity in the payoff function, we will denote it by q for convenience.

2.1.3 Intrusion Detection System (IDS)

In this section, we introduce the model of an IDS and present its application for our game-theoretic model.

Broadly speaking, an intrusion detection system (IDS) is ICT product that monitors or a system to detect fraudulent or malicious activity. In our model, we will focus on IDS based on Neyman Pearson Decision Theory.

Theorem 1. ***Neyman Pearson lemma***

Let us define two hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ for the random variable $\boldsymbol{Y}$ with realization y and associated PDFs as $p_{y|\mathcal{H}_0}(y|\mathcal{H}_0)$ and $p_{y|\mathcal{H}_1}(y|\mathcal{H}_1)$ respectively. We define likelihood ratio $\mathcal{L}(y)$ and the Likelihood Ratio Test(LRT) as,

$$\mathcal{L}(y) \triangleq \frac{p_{y|\mathcal{H}}(\boldsymbol{Y} = y|\mathcal{H} = \mathcal{H}_1)}{p_{y|\mathcal{H}}(\boldsymbol{Y} = y|\mathcal{H} = \mathcal{H}_0)} \underset{\widehat{H}(y)=\mathcal{H}_0}{\overset{\widehat{H}(y)=\mathcal{H}_1}{\gtrless}} \mu$$

The decision maker chooses a decision rule $\widehat{H}$ that,

$$\max \rho \textit{ subject to } \alpha \leqslant \alpha_0,$$

where ρ is the probability of detection and α is the probability of false alarm (for continuous y):

$$\rho := \mathbb{P}(\widehat{H}(y) = \mathcal{H}_1|\mathcal{H} = \mathcal{H}_1) = \int_{y_1} p_{y|\mathcal{H}}(y|\mathcal{H}_1)dy$$

$$\alpha := \mathbb{P}(\widehat{H}(y) = \mathcal{H}_1 | \mathcal{H} = \mathcal{H}_0) = \int_{y_1} p_{y|\mathcal{H}}(y|\mathcal{H}_0) dy$$

where y_1 represents the region that corresponds $\widehat{H}(y) = \mathcal{H}_1$. Maximization of ρ subject to the constraint $\alpha < \alpha_0$ corresponds to using a decision rule with Likelihood Ratio above the threshold μ such that,

$$\alpha = \mathbb{P}(L(y) \geqslant \mu | \mathcal{H} = \mathcal{H}_0) = \alpha_0$$

Example 1. **Normal Distribution**

In our numerical illustrations, we assume that both fraudulent and genuine customer report consumption with underlying normal distribution. Let the average quantity of electricity reported by genuine customer and fraudulent customer be $\mathbf{Y}_b^{\mathrm{g}} \sim \mathcal{N}(q, \sigma^2)$ and $\mathbf{Y}_b^{\mathrm{f}} \sim \mathcal{N}(q_b^{\mathrm{f}}, \sigma^2)$ respectively. Here both genuine and fraudulent customer are assumed to have the same variance. The fraudulent customer can reduce the mean of reported consumption but not change the inherent noise. This assumption simplifies the analysis of our game-theoretic model which uses the IDS based on LRT.

The Receiver Operating Characteristic (ROC) curve for the normal distribution for theft level η committed by the fraudulent customer is

$$\rho(\alpha, \eta) = \phi\left(\phi^{-1}(\alpha) + \frac{q\eta}{\sigma}\right) \tag{3}$$

where ϕ is the standard normal CDF.

Proof. Let the PDF for genuine customers be $f_{\mathrm{g}}(x)$ and that of attacker by $f_{\mathrm{f}}(x)$

$$\rho = \int_{-\infty}^{t} f_{\mathrm{f}}(x) dx \quad \text{and} \quad \alpha = \int_{-\infty}^{t} f_{\mathrm{g}}(x) dx$$

$$\text{which implies that} \quad \rho = \phi(\frac{t - q_b^{\mathrm{f}}}{\sqrt{\sigma^2}}) \quad \text{and} \quad \alpha = \phi(\frac{t-q}{\sqrt{\sigma^2}})$$

where ϕ is the CDF function of standard normal. Equivalently, equating t in both α, ρ above,

$$\sigma\phi^{-1}(\rho) + q_b^{\mathrm{f}} = \sigma\phi^{-1}(\alpha) + q$$

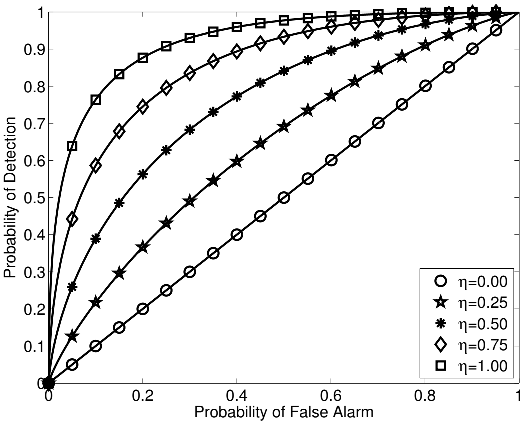


Figure 2.2: Receiver Operating Characteristic: Gaussian Distribution (different mean but same variance)

Furthermore, as shown earlier, $q_b^{\mathrm{f}} = (1-\eta)q$ and $\eta \in [0, 1]$. Hence, (3) follows. □

In Figure 2.2, we can visualize the variation of detection probability ρ probability with both false alarm probability α and theft level η. Note that for a given false alarm probability, the detection probability increases with theft level. Similarly for a given theft level, the detection probability increases with false alarm probability. In fact, we also observe that for a given theft level $\eta > 0$, probability of detection is a strictly concave function of false alarm probability. In the subsequent chapters and equilibrium results, we will use the ROC curve (3).

2.1.4 Interpretation of Cost/Payoff Parameters

i. **Cost of False Alarm** C_α: The cost of false alarm is the cost that the distribution utility incurs in investigating a genuine customer. The cost of false alarm can have multiple interpretations for the proposed security model. Precisely, the cost can represent the revenue lost in the form of reduced subscriptions to electric utility services. This can be attributed to inception of trust deficit between the distribution utility and genuine customer. Furthermore we can understand the importance of false alarm costs for outsourced security operations. For example, the distribution utilities across the US outsource its data operations to external analytics companies. One can argue that as the number of false alarm instance increases, the analytics company enjoys less faith from the distribution utility and the cost of false alarm, in this case, represents the cost of the lost subscription or early termination of security contracts.

ii. **Cost of Investigation** C_ρ : The cost of investigation is the cost that the distribution utility incurs by investigating the customer after the Intrusion Detection System indicates fraudulent activity. We interpret the cost of investigation as the dollar amount the distribution utility pays its investigation/monitoring crew for perusal of suspected AMI.

iii. **Cost of Theft** C_η : The cost of theft is the cost that the fraudulent customer incurs by choosing to commit electricity theft. One can argue that there exists two types of fraudulent customers: "sophisticated" and "amateur". The "sophisticated" fraudulent customers commit theft by cyber-phyiscal means and the cost of theft represents the investment in technology to hack AMI devices installed by the distribution utility. The "amateur" fraudulent customers commit electricity theft by physical measures for example meter tampering and the cost of theft represents the investment in installation of faulty meters or damaging the genuine meters. The cost of theft can also represent the social reward that the fraudulent customer faces, for example: suspicion by the society/neighbors or guilt of committing crime.

iv. **Fine** F: The fine is the cost that the fraudulent customer, if committing electricity theft, pays to the distribution utility on successful investigation. Equivalently, fine represents the revenue distribution utility collects from fraudulent customers after the IDS raises an alarm to investigate. Importantly, fine is imposed by the distribution utility to deter the fraudulent customer from committing theft. Fine can be executed by an increase in fixed fee, or an increased marginal price of electricity for the customer, or it can be termination of certain services by the distribution utility. The central idea is that the fraudulent customer incurs a cost and the distribution utility obtains a revenue on successful detection.

2.1.5 Assumptions

We will make the following assumptions throughout our analysis.

2.1.5.1 Fraudulent Customer

Assumption 1. i. The cost of attack C_{η} is less than average revenue from a genuine customer i.e., $\mathrm{C}_{\eta} < pq$.

This assumption implies that we restrict our attention to case in which the cost of attack is less than gain obtained from maximum theft (i.e. when $\eta = 1$)

ii. Fraudulent Consumer reports less but consumes the same average quantity as genuine customers i.e., $q_b^{\mathrm{f}} < q_a^{\mathrm{f}} = q_a^{\mathrm{g}}$.

This assumption is justified for both LOW type and HIGH type electricity customers.

(a) **Fraudulent "LOW" customers**: According to the World Bank Report on Non-Technical Losses(NTL) [3], fraudulent customers usually represent customers who do not pay tariff for electricity and consume effectively 50% more when they commit electricity theft. As a result, we argue that these LOW type customers, on an average, consume the same as HIGH type genuine customers.

(b) **Fraudulent "HIGH" customers**: We argue that although fraudulent "HIGH" customers save money by paying for part of the total electricity consumed and consequently face a reduced "effective" tariff rate. However, we argue that the fraudulent customer does not use the above savings to increase his electricity consumption. This is because the money saved is considered a "risky" income by the consumer and electricity consumption is influenced by permanent increase/decrease in the annual income. As a result, the fraudulent customer from "HIGH" category will not change his/her consumption. S/He will consume, on an average, same as a genuine customer of "HIGH" category but report less to the distribution utility.

2.1.5.2 Distribution Utility

Assumption 2. i. The cost of investigating a fraudulent customer C_ρ is less than the fine that can be imposed by the distribution utility on successful detection i.e., $\mathrm{C}_\rho < \mathrm{F}$.

This assumption implies that the maximum amount in the form of recoverable fine (i.e. for a perfect IDS) exceeds the cost incurred by the distribution utility in investigating a fraudulent customer upon successful detection.

ii. A two-part tariff schedule for the customer $\mathrm{T}(q) := pq + A$ where p is the marginal price and A is the fixed lump-sump fee.

This assumption is based on a standard practice in pricing implemented by distribution utilities.

iii. We adopt a linear production cost schedule for the distribution utility $C(q) := cq + B$. For convenience, we will assume $B = 0$

This assumption is made for simplicity and for sake of tractability. Also the assumption does not change the major insights/results of the game-theoretic model.

iv. We consider intrusion Detection Systems with ROC curves between detection probability ρ and (α, η), where α is the false alarm probability and η is the level of theft, satisfying the following:

$$\rho(\alpha, \eta = 0) = \alpha \quad , \quad \rho(\alpha = 0, \eta) = 0 \quad , \quad \rho(\alpha = 1, \eta) = 1$$

Furthermore $\rho(\alpha, \eta)$ is strictly concave and strictly increasing in α for all η with $\lim_{\alpha\to 0} \partial_\alpha \rho(\alpha) = \infty$ and $\lim_{\alpha\to 1} \partial_\alpha \rho(\alpha) = 0$

The above properties are satisfied for ROC curve representing the hypothesis testing between normal distributions. Additionally most, if not all, of the ML/anomaly detection algorithms have ROC curve that satisfy the aforementioned properties. Finally note that we do not make any additional assumptions on the dependence of $\rho(\alpha, \eta)$ on η.

2.2 Game

In this section, we will understand the extensive and normal form representation of the inspection game and model the decision making process using both simultaneous and sequential framework for the distribution utility. Next, we will discuss the payoffs for both the players.

2.2.1 Game Setting

In this section, we formulate the sequence of interactions between the distribution utility and fraudulent customer. Firstly, we present an extensive form representation of the game, secondly we present an abstract formulation of game timing and finally we will instantiate the abstract formulation to model various operational as well as investment aspects of IDS. The above formulations will help us develop deeper insights regarding the value added by the IDS, value derived from customizing it and finally value derived from gaining greater information about the customers.

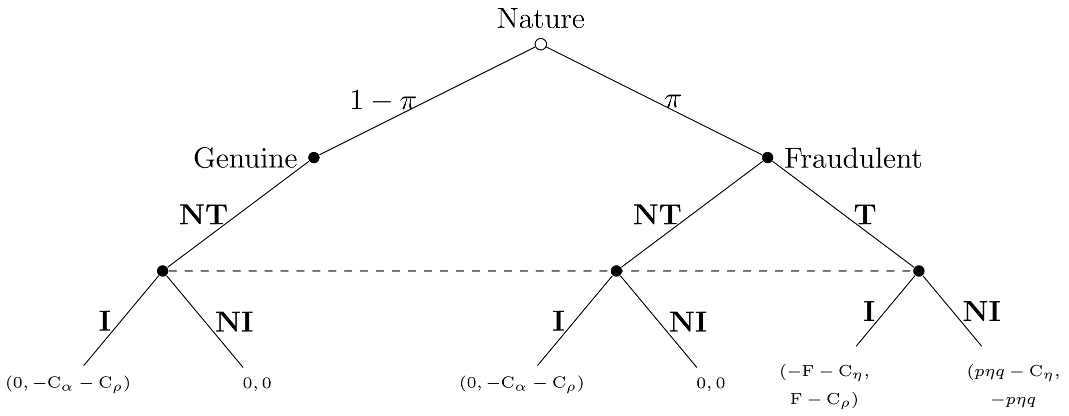


Figure 2.3: Extensive Form Representation of the Bayesian Inspection Game

2.2.1.1 Extensive Form Representation

Recall that we modeled the distribution utility - customer (both *Genuine* and *Fraudulent*) interaction as a Bayesian Inspection game. Figure 2.3 presents the extensive form representation. Before the distribution utility and customer interact, nature assigns fraudulent or genuine type to each customer with probability π and $1-\pi$ respectively. Subsequently, the fraudulent customer f chooses a pure strategy Theft **T** or No Theft **NT**, i.e $\delta^{\mathrm{f}} = \{\mathbf{T}, \mathbf{NT}\}$; the genuine customer g chooses the pure strategy No Theft **NT** i.e. $\delta^{\mathrm{g}} = \{\mathbf{NT}\}$; and the distribution utility D chooses a pure strategy Investigate **I** or No Investigate **NI**, i.e. $\delta^{\mathrm{D}} = \{\mathbf{I}, \mathbf{NI}\}$.

In the following sections, we will consider mixed-strategy equilibrium for the game. Following our earlier discussion, the fraudulent customer chooses probability of committing theft γ, i.e. $\mathrm{P}(\mathbf{T}) = \gamma$ and $\mathrm{P}(\mathbf{NT}) = 1-\gamma$. Furthermore, the mixing probability for the distribution utility is governed by the operating point of the IDS (α, ρ). Precisely consider the following pair of pure strategies for the distribution utility and fraudulent customer/genuine customer:

i. Distribution Utility vs Fraudulent Customer

   (a) $(\mathbf{I}, \mathbf{NT})$: $\mathrm{P}(\textbf{False Positive}) = \alpha$

   (b) $(\mathbf{I}, \mathbf{T})$: $\mathrm{P}(\textbf{True Positive}) = \rho$

   (c) $(\mathbf{NI}, \mathbf{T})$: $\mathrm{P}(\textbf{False Negative}) = 1-\rho$

   (d) $(\mathbf{NI}, \mathbf{NT})$: $\mathrm{P}(\textbf{True Negative}) = 1-\alpha$

ii. Distribution Utility vs Genuine Customer

   (a) $(\mathbf{I}, \mathbf{NT})$: $\mathrm{P}(\textbf{False Positive}) = \alpha$

   (b) $(\mathbf{NI}, \mathbf{NT})$: $\mathrm{P}(\textbf{True Negative}) = 1-\alpha$

Hence, we note that the distribution utility randomizes between **I** and **NI** using false alarm probability and detection probability of the IDS.

In subsequent sections we will consider scenarios where, given an IDS of default configuration (α_0, ρ_0) the distribution utility chooses to use the IDS or not use the IDS. Let the probability with which the distribution utility uses the IDS be β. Then, the mixing probability between **I** and **NI** becomes,

i. Distribution Utility vs Fraudulent Customer

(a) $(\mathbf{I}, \mathbf{NT})$: P(**False Positive**) $= \beta\alpha_0$

(b) $(\mathbf{I}, \mathbf{T})$: P(**True Positive**) $= \beta\rho_0$

(c) $(\mathbf{NI}, \mathbf{T})$: P(**False Negative**) $= 1 - \beta\rho_0$

(d) $(\mathbf{NI}, \mathbf{NT})$: P(**True Negative**) $= 1 - \beta\alpha_0$

ii. Distribution Utility vs Genuine Customer

(a) $(\mathbf{I}, \mathbf{NT})$: P(**False Positive**) $= \beta\alpha_0$

(b) $(\mathbf{NI}, \mathbf{NT})$: P(**True Negative**) $= 1 - \beta\alpha_0$

Hence, we conclude that the extensive form representation in Figure 2.3 models the pure strategies for both distribution utility and fraudulent customer in multiple scenarios. Since we are calculating the mixed strategy equilibrium of the inspection game, we consider the mixing probabilities, (α, ρ) and β for the distribution utility and γ for the fraudulent customer as pure strategies.
In the rest of thesis, we will consider (α, ρ, β) as the decision variable of the distribution utility and γ as the decision variable of the fraudulent customer.

2.2.1.2 Abstract Formulation

Consider a two stage game between the distribution utility and the electricity customer, who can be both genuine and fraudulent.

i. **Stage 1**: The distribution utility chooses the optimal operating point of the IDS: (α_0, ρ_0)

Discussion .

Equivalently, the distribution utility chooses the false alarm probability α_0. Since the level of theft is revealed in Stage 2, the distribution utility finds the detection probability ρ_0 using the corresponding ROC curve. The above model follows from the Neyman Pearson Theorem described in the earlier chapter. Although the IDS can use various machine learning/anomaly detection algorithms, we present our results for non-bayesian hypothesis testing whereby the distribution utility investigates a fraudulent customer if the reported consumption falls below certain threshold. From the definition of false alarm probability, the statistical test implemented is: $\{Investigate(\mathbf{I}) : \mathbf{Y} = \mathbf{y}|\mathbf{y} < (\phi^{-1}(\alpha_0) + \frac{q\eta}{\sigma})\}$ where the threshold is $\phi^{-1}(\alpha_0) + \frac{q\eta}{\sigma}$

We consider a two stage framework to model uncertainties in the decision making process of the distribution utility. The uncertainties can be attributed to both acquiring as well as customizing the Intrusion Detection System (IDS). For example, the distribution utility does not have complete information about the level of theft committed by the fraudulent customer and consequently chooses the optimal configuration by using a probability measure over the possible level of thefts. Furthermore one can also consider a case where the distribution utility does not have any information about the level of theft and consequently uses the IDS in its default configuration. Stage 1 analysis will help us answer questions related to value of information and value of customization for the distribution utility.

ii. **Stage 2**: The distribution utility chooses the probability of using the IDS. $\beta^{\dagger}$ and the fraudulent customer chooses the probability of committing theft $\gamma^{\dagger}$ in a simultaneous game

Discussion

Equivalently, the distribution utility chooses the "effective" false alarm probability *after* the fraudulent customer reveals the level of theft or the distribution utility learns the level of theft using machine learning or other cyber-physical procedures. Furthermore the fraudulent customer chooses the probability of Theft($\mathbf{T}$) using information about the current fraction of fraudulent customers and other cost parameters. Stage 2 analysis helps us gain deeper insights into the operational aspects of the IDS. For example, the distribution utility will vary the effective false alarm probability in accordance with the fraction of fraudulent customers and the level of theft. Additionally, we can answer questions related to value of IDS using Stage 2 subgame equilibrium.

2.2.1.3 Game Formulation

In this section, we will present the precise formulation of all the games we will consider in the subsequent sections using the abstract form presented earlier.

i. **Game 0** $\mathcal{G}^0$:

   In game $\mathcal{G}^0$, the distribution utility does not possess an IDS and consequently does not investigate, i.e Not Investigate **NI** is the only strategy. The game proceed in two stages:

   (a) **Stage 1**: The distribution utility chooses an IDS of zero false alarm probability $\alpha_0 = 0$.

   Since, $\alpha_0 = 0$, the probability of detection $\rho_0 = 0$. Note that this is equivalent to No IDS case or No Investigate **NT** strategy.

   (b) **Stage 2**: The fraudulent customer chooses the probability of theft **T**.

   Note that since the distribution utility chooses a zero false alarm probability IDS and it does not have any effect on the payoffs/equilibrium of the game, we do not analyze the probability of using the IDS β.

ii. **Game 1** $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$:

   In game $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$, the distribution utility uses an IDS of default configuration (α_0, ρ_0). Furthermore the distribution utility does not possess information about the ROC curve or theft level in Stage 1. The game proceeds in two stages:

   (a) **Stage 1**: The distribution utility chooses the default configuration of the IDS (α_0, ρ_0).

   Note that since the distribution utility has no information about the ROC curve, it does not configure the given IDS.

   (b) **Stage 2**: The distribution utility chooses the probability of using the IDS β. The fraudulent customer chooses the probability of committing theft

T.

Note that both fraction of fraudulent customers π and level of theft η becomes common knowledge before Stage 2.

iii. **Game 2** $\mathcal{G}_{\eta}^{\text{ROC}}$: In game $\mathcal{G}_{\eta}^{\text{ROC}}$, the distribution utility has complete information about the level of theft committed by the fraudulent customer.As a result, the distribution utility skips Stage 1 and find the optimal false alarm probability α^* in Stage 2. Equivalently, since the distribution utility has information about theft level, he also finds the optimal detection probability ρ^*. Furthermore the fraudulent customer finds the probability of committing theft γ.

iv. **Game 3a** $\widehat{\mathcal{G}}_{\eta}^{\text{ROC}}$: In game $\widehat{\mathcal{G}}_{\eta}^{\text{ROC}}$, the distribution utility has complete information about level of theft η committed by the fraudulent customer.

(a) **Stage 1**: The distribution utility chooses an optimal configuration of the IDS α_0^* using the ROC curve corresponding to η.

(b) **Stage 2**: The distribution utility chooses the probability of using the IDS β. The fraudulent customer chooses the probability of committing theft γ.

v. **Game 3b-** $\mathcal{G}_{\text{P}(\eta)}^{\text{ROC}}$: In game $\mathcal{G}_{\text{P}(\eta)}^{\text{ROC}}$, the distribution utility gets information about of a probability measure over level of theft. For simplicity, we assume that there exists two levels of theft $\eta^{\text{L}}, \eta^{\text{H}}$ with probability λ^{L} and $\lambda^{\text{H}} = 1 - \lambda^{\text{L}}$ respectively.

(a) **Stage 1**: The distribution utility chooses an optimal configuration of the IDS α_0^*.

Note that the distribution utility finds the above α_0 using the ROC curves corresponding to $\eta^{\text{L}}, \eta^{\text{H}}$. Furthermore, as described earlier, although the distribution utility chooses the false alarm probability in stage 1, the detection probability is determined in Stage 2 when the fraudulent customer reveals the true level of theft η^{L} or η^{H}

(b) **Stage 2**: The distribution utility chooses the probability of using the IDS β. The fraudulent customer chooses the probability of committing theft γ.

Game	Symbol	Stage 1	Stage 2	IDS Parameters	Theft Level
Game 0	$\mathcal{G}^0$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
Game 1	$\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$	$\varnothing$	(β,γ)	(α_0,ρ_0)	η
Game 2	$\mathcal{G}_\eta^{\text{ROC}}$	$\varnothing$	$((\alpha,\rho),\gamma)$	$\rho(\alpha)$	η
Game 3a	$\widehat{\mathcal{G}}_\eta^{\text{ROC}}$	$((\alpha_0,\rho_0),\varnothing)$	(β,γ)	$\rho(\alpha)$	η
Game 3b	$\mathcal{G}_{\text{P}(\eta)}^{\text{ROC}}$	$((\alpha_0,\rho_0),\varnothing)$	(β,γ)	$\rho(\alpha)$	$\text{P}(\eta)$

Table 2.1: Game Setting

The difference in game timing between $\widehat{\mathcal{G}}_\eta^{\text{ROC}}$ and $\mathcal{G}_\eta^{\text{ROC}}$ is subtle, but important. Note that, in both cases, the distribution utility finds the optimal configuration of the IDS. In game $\widehat{\mathcal{G}}_\eta^{\text{ROC}}$, the distribution utility is the **leader** and finds the optimal false alarm probability in Stage 1 using the equilibrium response of the fraudulent customer in Stage 2. In game $\mathcal{G}_\eta^{\text{ROC}}$, the distribution utility finds the optimal false alarm probability and the fraudulent customer chooses the probability of committing theft simultaneously.

2.2.2 Payoffs

In this section, we will first present the normal form representation of the game. Then, we will derive the payoff function for both distribution utility and fraudulent customer in all the scenarios/games defined earlier.

2.2.2.1 Normal Form Representation

Figure 2.2 presents the normal form representation of the game. Note that all the cost parameters of the game are common knowledge and known to both the players before start of the game.

Explanation:

i. For strategy profile (**I**, **T**): Due to successful detection, the distribution utility collects fine F from the fraudulent customer but incurs cost of monitoring C_ρ. The fraudulent customer incurs both fine F and cost of attack C_η.

ii. For strategy profile (**NI, T**): The distribution utility does not investigate and

		Distribution Utility	
		I	**NI**
Fraudulent Customer	**T**	$(-\mathrm{F}-\mathrm{C}_{\eta}, \mathrm{F}-\mathrm{C}_{\rho})$	$(pq\eta-\mathrm{C}_{\eta}, -pq\eta)$
	NT	$(0, -\mathrm{C}_{\alpha}-\mathrm{C}_{\rho})$	$(0, 0)$

		Distribution Utility	
		I	**NI**
Genuine Customer	**NT**	$(0, -\mathrm{C}_{\alpha}-\mathrm{C}_{\rho})$	$(0, 0)$

Table 2.2: Normal Form Representation for Distribution Utility vs Fraudulent Customer (top) and Genuine Customer (bottom)

loses $pq\eta$. The fraudulent customer obtains revenue $pq\eta$ and incurs cost of attack C_{η}.

iii. For strategy profile (**I, NT**): The distribution utility commits False Alarm and loses both false alarm cost C_{α} and cost of investigation C_{ρ}. The fraudulent customer does not commit theft and obtains zero payoff.

iv. For strategy profile (**NI, NT**): The distribution utility does not investigate **I** and obtains zero payoff. The fraudulent customer does not commit theft **NT** and obtains zero payoff.

Note that the payoff for genuine customer is equal to fraudulent customer's payoff when it commits No Theft **NT**.

2.2.2.2 Distribution Utility

From the normal form representation and mixing probabilities discussed earlier, we can write the profit term and the payoff term for the distribution utility as follows,

$$\Pi^{\mathrm{D}}(\alpha_0, \rho_0) = A + pq + u^{\mathrm{D}}((\alpha_0, \rho_0, \beta), \gamma) - cq \tag{4}$$

where A is the fixed fee and pq is the variable tariff that the distribution utility collects from the customer. Additionally, the distribution utility pays cq to supply electricity to the customer. Furthermore, the distribution utility payoff from the inspection game for a strategy profile $((\alpha_0, \rho_0, \beta), \gamma)$ is:

$$u^{\mathrm{D}}((\alpha_0, \rho_0, \beta), \gamma) = \pi\gamma\left(\beta\rho_0\left(\mathrm{F} - \mathrm{C}_\rho\right) - \left(1 - \beta\rho_0\right) pq\eta\right) - \left(1 - \pi\gamma\right)\beta\alpha_0\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) \tag{5}$$

2.2.2.3 Fraudulent Customers

From the normal form representation 2.2, the payoff for fraudulent customer can be written as,

$$u^{\mathrm{f}}((\alpha_0, \rho_0, \beta), \gamma) = \gamma\left(\left(1 - \beta\rho_0\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right) \tag{6}$$

Remark 1. The utility of a fraudulent customer u^{f} at equilibrium is always greater than equal to 0

Proof. From (6), the fraudulent customer will choose $\gamma^* > 0$ if $(1 - \beta\rho_0)(\mathrm{F} + pq\eta) > (\mathrm{F} + \mathrm{C}_\eta)$ and choose $\gamma^* = 0$ if $(1 - \beta\rho_0)(\mathrm{F} + pq\eta) \leqslant (\mathrm{F} + \mathrm{C}_\eta)$. Hence, $u^{\mathrm{f}} \geqslant 0$ □

Chapter 3

Configuration of Intrusion Detection System

3.1 Game $\mathcal{G}^0$: No IDS

Recall that in game $\mathcal{G}^0$, the distribution utility does not have access to an IDS, i.e. $\alpha_0 = 0$ and $\rho_0 = 0$. From (5), we obtain $u^{\mathrm{D}}((0,0),\gamma) = -\pi\gamma pq\eta$. Similarly from (6), $u^{\mathrm{f}}((0,0),\gamma) = \gamma(pq\eta - \mathrm{C}_\eta)$. Since the payoff of fraudulent customer does not depend on distribution utility's strategy, it cannot deter the fraudulent customer from committing theft $\mathbf{T}$. Hence, $\gamma^* = 1$, and in equilibrium, the player's payoffs are $u^{\mathrm{D}*} = -\pi pq\eta$ and $u^{\mathrm{f}*} = pq\eta - \mathrm{C}_\eta$

3.2 Game $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$: Default IDS

We now consider the case where the distribution utility uses an IDS with "default" configuration. Here we assume that the default IDS may not be configured in an optimal manner for the given theft level. This case models a situation when the distribution utility is not able to configure the underlying anomaly detection algorithm and its IDS only operates at a particular point on the ROC curve. In the following, we first find the equilibrium of the game $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$ and analyze its behavior with respect to the level of theft and the default configuration of IDS.

Under a given default IDS (α_0, ρ_0), the player payoffs for the Game $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$ for a strategy profile (β, γ) can be written as follows:

$$u^{\mathrm{D}}(\beta,\gamma) = \pi\gamma\left(\beta\rho_0\left(\mathrm{F}-\mathrm{C}_\rho\right) - \left(1-\beta\rho_0\right)pq\eta\right) - \left(1-\pi\gamma\right)\beta\alpha_0\left(\mathrm{C}_\alpha+\mathrm{C}_\rho\right) \tag{7}$$

$$u^{\mathrm{f}}(\beta,\gamma) = \gamma\left(\left(1-\beta\rho_0\right)\left(\mathrm{F}+pq\eta\right) - \left(\mathrm{F}+\mathrm{C}_\eta\right)\right) \tag{8}$$

We know that a strategy profile $\left(\beta^\dagger, \gamma^\dagger\right)$ is an equilibrium of $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$ if and only if,

$$u^{\mathrm{D}}(\beta^\dagger,\gamma^\dagger) \geqslant u^{\mathrm{D}}(\beta,\gamma^\dagger) \quad \forall\beta \tag{9}$$

$$u^{\mathrm{f}}(\beta^\dagger,\gamma^\dagger) \geqslant u^{\mathrm{f}}(\beta^\dagger,\gamma) \quad \forall\gamma \tag{10}$$

We define three quantities of interest: Critical level of theft η_c^I, critical fraction of fraudulent customers π_c and critical detection probability ρ_c.

$$\eta_c^I := \frac{\mathrm{C}_\eta}{pq} \tag{11}$$

$$\pi_c := \frac{(\mathrm{C}_\alpha+\mathrm{C}_\rho)\alpha_0}{(\mathrm{C}_\alpha+\mathrm{C}_\rho)\alpha_0 + \rho_0(pq\eta+\mathrm{F}-\mathrm{C}_\rho)} \tag{12}$$

$$\rho_c := \frac{pq\eta-\mathrm{C}_\eta}{pq\eta+\mathrm{F}} \tag{13}$$

We claim that in equilibrium, η_c is the critical fraction of theft level η below which the fraudulent customers do not attack with probability 1. That is, the fraudulent customers commit theft with non-zero probability only when $\eta > \eta_c^I$. Also, π_c is the critical fraction of fraudulent customers such that if $\pi \leqslant \pi_c$ and $\eta > \eta_c^I$, the fraudulent customer always commit theft. However, when $\pi > \pi_c^I$ and $\eta > \eta_c^I$, the distribution utility (resp. fraudulent customer) choose to investigate (resp. commit Theft) with probability 1 if ρ_0 the default IDS detection probability $\rho_0 \leqslant \rho_c$; however the players chose randomized (mixed strategy) if $\rho_0 > \rho_c$. These claims are formalized in the following proposition.

Proposition 1. ***Game with Default IDS*** $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$ ***Equilibrium***

Consider the game $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$ *with a given default IDS specified by* (α_0, ρ_0) *and for a* *given two-part tariff schedule* $\mathrm{T}(q) := A + pq$*. The equilibrium of* $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$ *is unique and is given by:*

$$(\beta^\dagger, \gamma^\dagger) = \begin{cases} (0,0) & \text{if } \eta \leqslant \eta_c^I, \\ (0,1) & \text{if } \eta > \eta_c^I, \pi \leqslant \pi_c, \\ (1,1) & \text{if } \eta > \eta_c^I, \pi > \pi_c, \text{ and } \rho_0 < \rho_c, \\ \left(\frac{\rho_c}{\rho_0}, \frac{\pi_c}{\pi}\right) & \text{if } \eta > \eta_c^I, \pi > \pi_c, \text{ and } \rho_0 \geqslant \rho_c. \end{cases} \tag{14}$$

Proof. Let us re-write (8) as follows:

$$u^{\mathrm{f}}(\beta,\gamma) = \gamma\left(\Phi(\beta) - (\mathrm{F} + \mathrm{C}_\eta)\right), \tag{15}$$

where $\Phi(\beta) := (1 - \beta\rho_0)(p\eta q + \mathrm{F})$. Using (7), we can write the derivative of u^{D} with respect to β:

$$\partial_\beta u^{\mathrm{D}} = \pi[(1-\gamma)(-(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_0)] + (1-\pi)[-(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_0] + \pi[\gamma(p\eta q + \mathrm{F} - \mathrm{C}_\rho)\rho_0] \tag{16}$$

Now consider the following cases:

Case (i) $[\eta \leqslant \eta_c^I]$ We see from (11) that $pq\eta \leqslant \mathrm{C}_\eta$, which implies that $\Phi(\beta) \leqslant (1 - \beta\rho_0)(\mathrm{C}_\eta + \mathrm{F}) \leqslant (\mathrm{C}_\eta + \mathrm{F})$. From (15), $u^{\mathrm{f}}(\beta,\gamma) \leqslant 0$ for any strategy profile (β,γ). Thus,we conclude that, in equilibrium, $\gamma^\dagger = 0$.

Now, for a strategy profile $(\beta, 0)$, we can express (7) as $u^{\mathrm{D}}(\beta, 0) = -\beta\alpha_0(\mathrm{C}_\alpha + \mathrm{C}_\rho)$, which implies that in equilibrium, $\beta^\dagger = 0$.

Case (ii) $\left[\eta > \eta_c^I, \quad \pi \leqslant \pi_c\right]$

From (15), we see $\gamma = 1$ is a best response to β if and only if,

$$\Phi(\beta) > \mathrm{F} + \mathrm{C}_\eta \tag{17}$$

Consider a strategy profile $(\beta, 1)$, using (16) we obtain that:

$$\begin{aligned}\partial_\beta u^{\mathrm{D}}\big|_{(\beta,1)} &= -\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right)\alpha_0 + \pi\left(\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right)\alpha_0 + \left(pq\eta + \mathrm{F} - \mathrm{C}_\rho\right)\rho_0\right) \\ &< 0\end{aligned}$$

This implies that $\beta = 0$ is a BR to $\gamma = 1$.

Now, substituting $\beta = 0$ in (17), we check that $\mathrm{F} + p\eta q > \mathrm{F} + \mathrm{C}_\eta$; or equivalently, $\eta > \eta_c^I$, which indeed holds for this case. Hence, $\gamma = 1$ is a BR to $\beta = 0$. Thus, $\left(\beta^\dagger, \gamma^\dagger\right) = (0, 1)$ is an equilibrium.

To argue uniqueness, assume the contrary. Then, for any $\gamma \in [0, 1)$, from (16) we can write $\partial_\beta u^{\mathrm{D}} = \gamma\pi((\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_0 + (pq\eta + \mathrm{F} - \mathrm{C}_\rho)\rho_0) - (\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_0 < 0$, which would imply that $\beta = 0$ is a BR to any $\gamma \neq 1$. However, from (17) we know that $\gamma = 1$ is the unique BR to $\beta = 0$, which is a contradiction.

Case (iii) $\left[\eta > \eta_c^I, \quad \pi > \pi_c, \quad \rho_0 < \rho_c\right]$

For $\rho_0 \leqslant \rho_c$, we can verify that $\Phi(\beta) > \mathrm{F} + \mathrm{C}_\eta$ for all $\beta \in [0, 1]$. From (15), we obtain that $\gamma = 1$ is the BR to any β. Furthermore, for a strategy profile $(\beta, 1)$, using (16) and the fact that $\pi > \pi_c$, we can write $\partial_\beta u^{\mathrm{D}}(\beta, 1)\big|_{\gamma * = 1} = -\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right)\alpha_0 + \pi\left(\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right)\alpha_0 + \left(pq\eta + \mathrm{F} - \mathrm{C}_\rho\right)\rho_0\right) > 0$. This implies that $\beta = 1$ is a unique BR to $\gamma = 1$. Thus, $(\beta^\dagger, \gamma^\dagger) = (1, 1)$ is the unique equilibrium.

Case (iv) $\left[\eta > \eta_c^I, \quad \pi > \pi_c, \quad \rho_0 \geqslant \rho_c\right]$

Consider a strategy profile $(\beta, \gamma) \in (0, 1)^2$. Then, for (β, γ) to be an equilibrium, we necessarily need $\Phi\left(\gamma\right) = (\mathrm{F} + \mathrm{C}_\eta)$. Solving for β we obtain $\beta^\dagger = \rho_c/\rho_0 \in (0, 1)$. From (16), we put $\partial_\beta u^{\mathrm{D}} = 0$. Solving for γ, we obtain $\gamma^\dagger = \pi_c/\pi$. Thus, $(\beta^\dagger, \gamma^\dagger) = (1, 1)$ is the unique equilibrium.

To argue uniqueness for $\gamma \in (0, 1)$: Consider the following two cases: (a) If in equilibrium, $\frac{\partial u^{\mathrm{D}}}{\partial \beta} < 0$, then $\beta^\dagger = 0$ and $\gamma(\beta^\dagger) = 1$ which is a contradiction (b) If in equilibrium, $\frac{\partial u^{\mathrm{D}}}{\partial \beta} > 0$ then $\beta^\dagger = 1$ and $\gamma(\beta^\dagger) = 0$ which is a contradiction. Lastly, for $\frac{\partial u^{\mathrm{D}}}{\partial \beta} = 0$, $\gamma^\dagger$ is the unique solution. □

Remark 2. We can obtain the equilibrium payoffs for game $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$ by plugging $(\beta^\dagger, \gamma^\dagger)$ from Proposition 1 into (7) and (8)

$$u^{\mathrm{D}}(\beta^\dagger, \gamma^\dagger) = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ -\pi p q \eta & \text{if } \eta > \eta_c^I, \pi \leqslant \pi_c \\ -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_0 + \pi\Psi & \text{if } \eta > \eta_c^I, \pi > \pi_c \text{ and } \rho_0 < \rho_c \\ -\pi_c p q \eta & \text{if } \eta > \eta_c^I, \pi > \pi_c \text{ and } \rho_0 \geqslant \rho_c \end{cases} \tag{18}$$

and (19)

$$u^{\mathrm{f}}(\beta^\dagger, \gamma^\dagger) = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ p q \eta - \mathrm{C}_\eta & \text{if } \eta > \eta_c^I, \pi \leqslant \pi_c \\ (\mathrm{F} + p q \eta)(1-\rho_0) - (\mathrm{F} + \mathrm{C}_\eta) & \text{if } \eta > \eta_c^I, \pi > \pi_c \text{ and } \rho_0 < \rho_c \\ 0 & \text{if } \eta > \eta_c^I, \pi > \pi_c \text{ and } \rho_0 \lambda \rho_c \end{cases} \tag{20}$$

where we define $\Psi := ((-\mathrm{C}_\rho + \mathrm{F})\rho_0 - (pq\eta)(1-\rho_0))$
Furthermore using (5) and (4) the total profit collected by the distribution utility is $\Pi^{\mathrm{D}} = u^{\mathrm{D}}(\beta^\dagger, \gamma^\dagger) + A + pq - cq$. Throughout the thesis, we will use Π^{D} in simulations to show the equilibrium payoff of the distribution utility.

Remark 3. Under a default IDS (α_0, ρ_0), Proposition 1 implies that the "effective" operating point in equilibrium is $\left(\beta^\dagger\alpha_0, \beta^\dagger\rho_0\right)$. That is, the distribution utility finds the optimal operating point on the straight line joining $(0,0)$ and (α_0, ρ_0) as shown in Figure 3.1. Moreover, since under the assumptions, the ROC curve is concave function, the distribution utility has lower detection probability in this case, for a given false alarm probability; relative to the case when it has a complete knowledge of ROC curve.

In the following sections, we show the variation of equilibrium strategies and payoff for both distribution utility (or *defender*) and fraudulent customer (or *attacker*) with theft level and default false alarm probability.

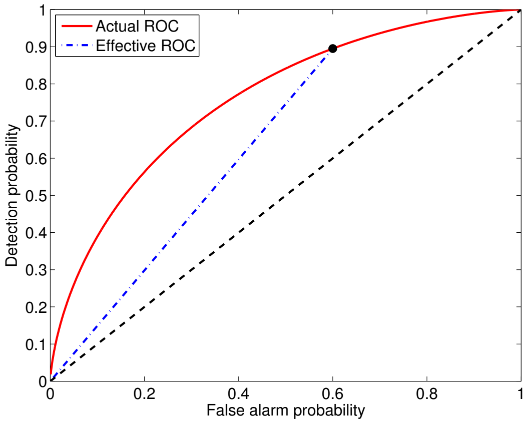


Figure 3.1: For $\eta = 0.5$, the "effective" ROC curve in blue with operating point $(\beta\alpha_0, \beta\rho_0)$

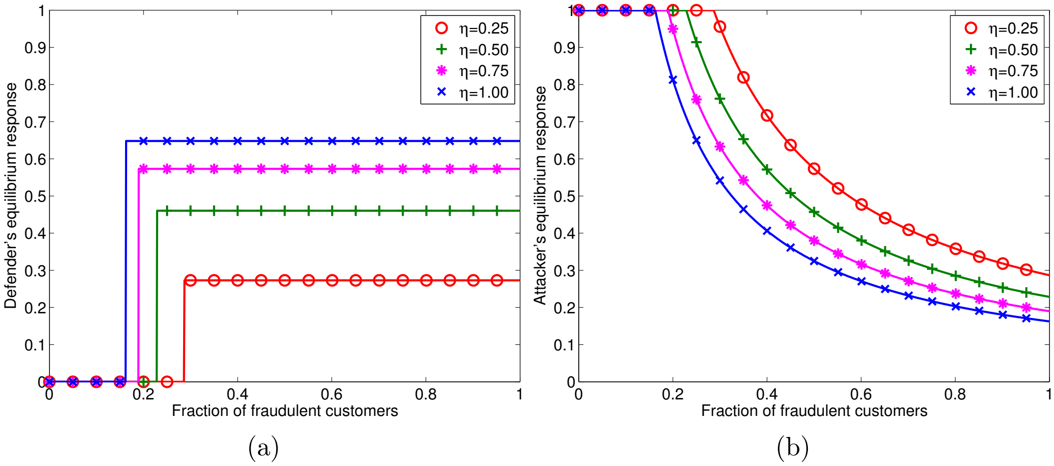


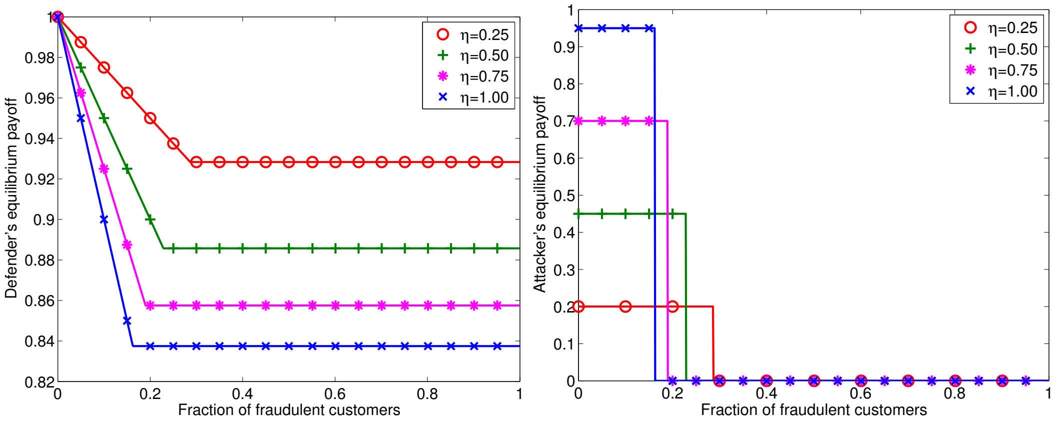
Figure 3.2: Equilibrium Strategies for $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$: Variation with fraction of fraudulent customers π for different theft levels η

3.2.1 Equilibrium Analysis - Level of theft

For game $\mathcal{G}_\eta^{(\alpha_0,\rho_0)}$, Figure 3.2a and Figure 3.2b (resp.3.3a and 3.3b) illustrate the equilibrium strategies (resp. equilibrium payoffs) of a fraudulent customer and distribution utility respectively when $\eta > \eta_c^I$.

Fraudulent Customer's Equilibrium Response

As shown in Fig 3.2b, a fraudulent customer commits theft with probability 1 when the fraction of fraudulent customers is below the critical threshold π_c. A fraudulent



(a) (b)

Figure 3.3: Equilibrium Payoffs for $\mathcal{G}_{\eta}^{(\alpha_0,\rho_0)}$: Variation with fraction of fraudulent customersπ for different theft levels η

customer's equilibrium payoff is constant in this range. However, the equilibrium payoff is non-decreasing in theft level η. One interpretation of this result is that when $\pi \leqslant \pi_c$, the fraudulent customer does not anticipate investigation **I** by the distribution utility.

For π beyond the critical fraction π_c, we observe from Fig 3.2b that the probability the fraudulent customer commits theft **T** decreases in the fraction of fraudulent customers, and increasing theft level. Intuitively, as the fraction of fraudulent customers increases, the fraudulent customer is deterred from committing theft due to an increased probability of investigation. A fraudulent customer is also deterred from committing theft as level of theft increases, since the distribution utility is more likely to investigate the fraudulent customer. Finally, from Fig, 3.3b, the fraudulent customer obtains zero payoff for $\pi > \pi_c$ and $\rho_0 > \rho_c$ because in equilibrium, the distribution utility chooses β such that the fraudulent customer becomes indifferent between **T** and **NT**.

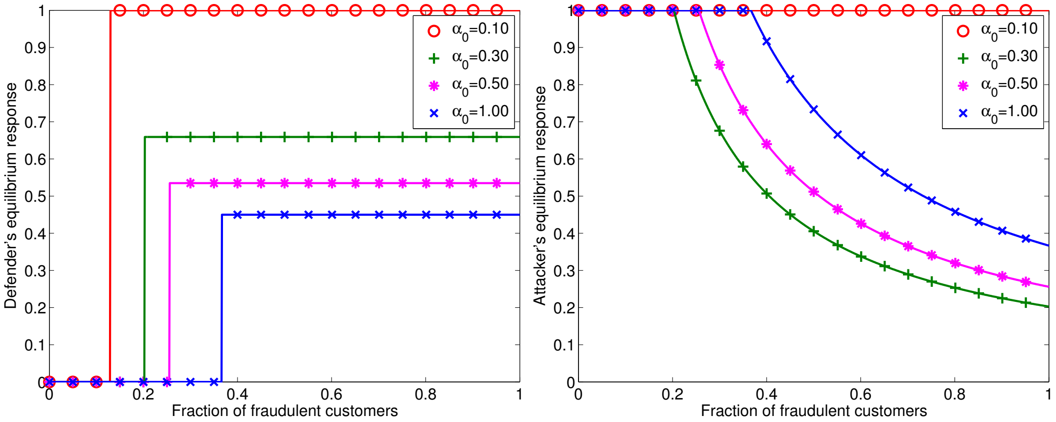
Distribution Utility's Equilibrium Response

From Fig 3.2a, we observe that that the distribution utility does not investigate **NI** with probability 1 when $\pi < \pi_c$. Additionally from Fig 3.3a, we observe in the above range, a decreasing payoff with increasing fraction of fraudulent customers η. One of the interpretations of critical fraction is that for $\pi < \pi_c$ the distribution utility incurs greater false alarm costs than recovered tariff and fine with the given IDS configuration. Consequently the distribution utility, for a given η and $\pi < \pi_c$, does not investigate and incurs a linearly decreasing payoff with π. Moreover, following the above reasoning, we can argue that for a given π, the distribution utility incurs a higher loss with greater theft η committed by the fraudulent customer.

However, for π beyond critical fraction π_c and a given IDS configuration (α_0, ρ_0), we observe in Fig 3.2a that the probability the distribution utility uses the IDS, or equivalently investigates **I**, remains constant with increasing fraction of fraudulent customers π. Furthermore the above constant probability, for a given π, increases with increasing theft level η. As we have seen earlier, the constant probability follows from the equilibrium concept that the distribution utility makes the fraudulent customer indifferent between Theft **T** and No theft **NT**. Similarly, we can argue from Fig, 3.3a that the distribution utility obtains a constant payoff for $\pi > \pi_c$. Moreover, following the equilibrium concept, as the level of theft η increases the distribution utility increases its probability to investigate **I** thereby reducing the expected theft of the fraudulent customer. Lastly, from Fig. 3.3a, the distribution utility incurs greater losses with increasing theft level η as the increased detection rate is coupled with increased false alarm costs. it is worth noting that for a given configuration of the IDS, the distribution utility does not leverage the concavity of the ROC curve but rather follows a linear ROC curve. Additionally, the distribution utility, for a given false alarm probability, does not obtain higher detection probability with increasing theft level for $\mathcal{G}_{\eta}^{(\alpha_0,\rho_0)}$.

3.2.2 Equilibrium Analysis - Default false alarm probability

Fraudulent Customer's Equilibrium Response From Fig 3.4b, as we have seen earlier, the fraudulent customer commits theft with probability 1 below a critical fraction π_c of fraudulent customers. Note that the definition and interpretation of critical fraction is same as in Fig 3.2b. Moreover, we observe that for very small default false alarm probability $\alpha_0 = 0.10$, the fraudulent customer always commits theft. Intu-

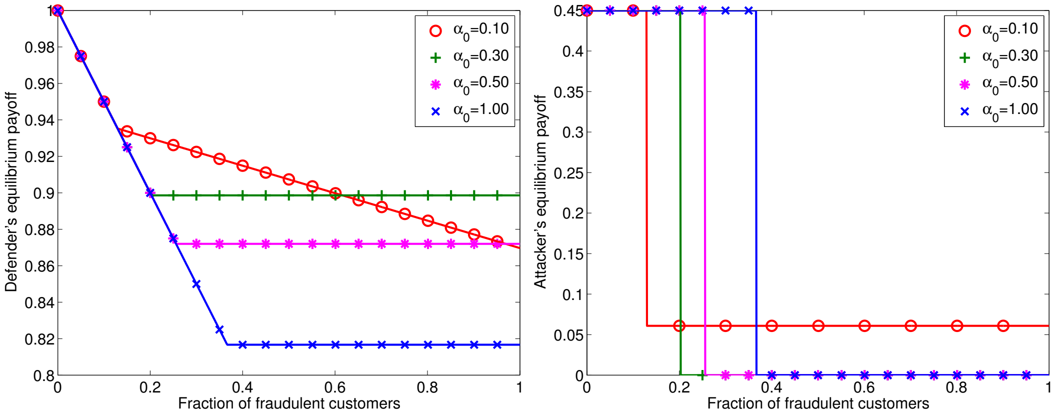


(a)

(b)

Figure 3.4: Equilibrium Strategies for $\mathcal{G}_{\eta}^{(\alpha_0,\rho_0)}$: Variation with fraction of fraudulent customers π for different default false alarm probability α_0

itively, since α_0 is very small, the distribution utility cannot procure enough detection probability to deter the fraudulent customer. Precisely, following from the ROC curve ($\rho(.)$) defined earlier, the detection probability is bounded above by $\rho(\alpha_0) = \rho(0.1)$. Furthermore, we also observe that for large default false alarm probability, for example $\alpha_0 = 1$, the fraudulent customer commits theft with larger probability. Note that for $\alpha_0 = 1$, from Figure 3.1, we can conclude that the distribution utility is using an inefficient IDS (similar to a biased coin toss) and consequently becomes an inefficient deterrent for the fraudulent customer. Additionally we observe, from Fig 3.5b, that the fraudulent customer obtains a constant payoff below the critical fraction. Although the payoff is constant across the different false alarm probabilities, the critical fraction is maximum for very small ($\alpha_0 = 0.1$) and very large ($\alpha_0 = 1$) default false alarm probabilities. In fact, for $\alpha_0 = 0.1$, the fraudulent customer obtains non-zero payoff for all fractions π. Lastly, as we have seen earlier, for π beyond critical fraction π_c, the fraudulent customer commits theft with decreasing probability with increasing fraction of fraudulent customers. Furthermore, we also observe in Figure 3.5b that the fraudulent customer obtains zero payoff for $\pi > \pi_c$. Again this follows from equilibrium concept that the distribution utility makes the fraudulent customer indifferent between Theft **T** and No theft **NT**.



(a) (b)

Figure 3.5: Equilibrium Payoff for $\mathcal{G}_{\eta}^{(\alpha_0,\rho_0)}$: Variation with fraction of fraudulent customers π for different default false alarm probability α_0

Distribution Utility's Equilibrium Response

From Fig 3.4a, we observe that the distribution utility does not investigate **NI** with probability 1 below a critical fraction π_c of fraudulent customers. Additionally we observe, from Fig 3.5a, that the distribution utility obtains, in the above range, a decreasing payoff with increasing fraction of fraudulent customers π. We observed a similar effect in Fig 3.2a, 3.3a and can similarly argue that since the distribution utility incurs greater false alarm costs than recovered tariff and fine for $\pi < \pi_c$, s/he does not investigate and incurs a linearly decreasing payoff with π.

However, for π beyond critical fraction π_c, we observe in Fig 3.4a that probability with which the distribution utility uses the IDS, or equivalently investigates **I**, remains constant with increasing fraction of fraudulent customers π. Furthermore the above constant probability, for a given π, decreases with increasing default false alarm probability η. As we have seen earlier, the constant probability follows from the equilibrium concept such that the distribution utility makes the fraudulent customer indifferent between Theft **T** and No theft **NT**. Moreover as the default false alarm probability increases, the distribution utility uses the IDS with smaller probability to save false alarm costs. Since the distribution utility uses the IDS with constant probability, we can argue from Fig, 3.5a that s/he obtains a constant payoff for $\pi > \pi_c$.

However for very small default false alarm probability(i.e. $\alpha_0 = 0.1$), as we have seen earlier, the distribution utility cannot procure sufficient detection probability such that the fraudulent customer commits theft with probability 1. As a result, the distribution utility always uses the inefficient IDS and incurs a linearly decreasing payoff with π.

Discussion:
From Figure 3.5a, it is evident that the distribution utility obtains maximum payoff from different default false alarm probability over different regimes of fraction of fraudulent customer. For instance, for $\pi < 0.2$ the distribution utility obtains same payoff from all default false alarm probabilities, for $0.2 < \pi < 0.7$ the distribution utility obtains maximum payoff from $\alpha_0 = 0.1$ and lastly for $\pi > 0.7$, the distribution utility obtains maximum payoff from $\alpha_0 = 0.3$. In fact, the difference in *best* and *worst* distribution utility payoff is $\sim 10 - 15\%$. Consequently, we expect that if the distribution utility can configure the IDS, the "optimal" false alarm rate(which also determines the detection rate) should be determined while accounting for both the fraction of fraudulent customers and the level of theft. This idea forms the basis of game with tunable IDS which we consider next.

3.3 Game $\mathcal{G}_{\eta}^{\textbf{ROC}}$: Tunable IDS

In this section, we consider the case when the distribution utility has the capability to "tune" or configure the IDS, i.e. choose an optimal false alarm probability α. As described earlier in section 2.2.1.3, in game $\mathcal{G}_{\eta}^{\text{ROC}}$, the distribution utility configures the IDS i.e. choose an optimal false alarm probability with the knowledge of theft level obtained in stage 2. Furthermore, for game $\mathcal{G}_{\eta}^{\text{ROC}}$, both the distribution utility and fraudulent customer have complete information about the ROC curve, i.e. $\rho(\alpha)$. In the following, we will present the equilibrium of game $\mathcal{G}_{\eta}^{\text{ROC}}$ and study its behavior for different levels of theft.

For ρ_c as defined in (13), we define α_0^c as :

$$\alpha_0^c := \rho^{-1}(\rho_c) \tag{21}$$

Recall that α_0^c can be viewed as critical false alarm probability, such that the resulting detection probability $\rho(\alpha_0^c) = \rho_c$ makes the fraudulent customer indifferent between **T**(Theft) and **NT**(No Theft).

Analogous to π_c (see (12)), we define another threshold,

$$\widetilde{\pi}_c := \frac{\mathrm{C}_\alpha + \mathrm{C}_\rho}{\mathrm{C}_\alpha + \mathrm{C}_\rho + \partial_\alpha \rho(\alpha_0^c)(\mathrm{F} + pq\eta - \mathrm{C}_\rho)} \tag{22}$$

Under ROC curve $\rho(\cdot)$, the player payoffs for the Game $\mathcal{G}_\eta^{\mathrm{ROC}}$ for a strategy profile (α, γ) can be written as follows (using (5) and (6)):

$$u^{\mathrm{D}}(\alpha,\gamma) = \pi\gamma\left(\rho(\alpha)\left(\mathrm{F} - \mathrm{C}_\rho\right) - \left(1 - \rho(\alpha)\right) pq\eta\right) - \left(1 - \pi\gamma\right)\alpha\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) \tag{23}$$

$$u^{\mathrm{f}}(\alpha,\gamma) = \gamma\left(\left(1 - \rho(\alpha)\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right) \tag{24}$$

Proposition 2. ***Game with Tunable IDS $\mathcal{G}_\eta^{ROC}$ Equilibrium***

Consider the game $\mathcal{G}_\eta^{ROC}$ with a IDS specified by ROC curve $\rho(\alpha)$ and for a given tariff schedule $\mathrm{T}(q) := A + pq$. The equilibrium of $\mathcal{G}_\eta^{ROC}$ is unique and given by:

$$(\alpha^*, \gamma^*) = \begin{cases} (0,0) & \text{if } \eta \leqslant \eta_c^I \\ \left(\chi^{-1}\left\{\frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{pq\eta + \mathrm{F} - \mathrm{C}_\rho}\right\}, 1\right) & \text{if } \eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c \\ \left(\alpha_0^c, \frac{\widetilde{\pi}_c}{\pi}\right) & \text{if } \eta > \eta_c^I, \pi > \widetilde{\pi}_c \end{cases} \tag{25}$$

where $\chi(\alpha) = \partial\rho/\partial\alpha$. Furthermore, for the reported consumption represented by random variable $\boldsymbol{Y}$ with realization $\boldsymbol{y}$, the statistical test implemented by the defender in equilibrium is given by $\{Investigate\ (\boldsymbol{I}) : \boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{y} < (\phi^{-1}(\alpha^*) + \frac{q\eta}{\sigma})\}$

Proof. We can re-write (24) as follows:

$$u^{\mathrm{f}}(\alpha,\gamma) = \gamma\left[\Phi(\alpha) - (\mathrm{F} + \mathrm{C}_\eta)\right] \tag{26}$$

where $\Phi(\alpha) := (1 - \rho(\alpha))\,(p\eta q + \mathrm{F})$

Case (i) $\left[\eta \leqslant \eta_c^I\right]$

For this case, we see from (11) that $pq\eta \leqslant \mathrm{C}_\eta$, which implies $\Phi(\alpha) \leqslant (1-\rho(\alpha))(\mathrm{C}_\eta + \mathrm{F}) \leqslant (\mathrm{C}_\eta + \mathrm{F})$. Thus, for any strategy profile (α, γ), $u^{\mathrm{f}}(\alpha, \gamma) \leqslant 0$. From (26), we conclude that, in equilibrium, $\gamma^* = 0$.

Now, for a strategy profile $(\alpha, 0)$, we can express (37) as $u^{\mathrm{D}}(\alpha, 0) = -\alpha(\mathrm{C}_\alpha + \mathrm{C}_\rho)$. Thus, in equilibrium, $\alpha^* = 0$.

Case (ii) $\left[\eta > \eta_c^I, \quad \pi \leqslant \tilde{\pi}_c\right]$

The proof consists of two parts:

- For $\gamma = 1$, α^* is the unique BR; where $\alpha^* \in [0, \alpha_c]$
- $\gamma = 1$ is the unique BR of α^*

Lemma 1. *For $\gamma = 1$, α^* is the unique BR; where $\alpha^* \in [0, \alpha_c]$ is given by,*

$$\frac{\partial \rho}{\partial \alpha}(\alpha^*) = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{pq\eta + \mathrm{F} - \mathrm{C}_\rho}$$

Proof. For a strategy profile $(\alpha, 1)$, from (23);

$$u^{\mathrm{D}}(\alpha, 1) = \pi\left(\rho(\alpha)\left(\mathrm{F} - \mathrm{C}_\rho\right) - (1-\rho(\alpha))\,pq\eta\right) - (1-\pi)\,\alpha\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) \tag{27}$$

Solving for FOC $u^{\mathrm{D}}_\alpha(\alpha, 1) = 0$,

$$\rho_\alpha(\alpha^*) = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{pq\eta + \mathrm{F} - \mathrm{C}_\rho} \tag{28}$$

Recall that, by assumption on the ROC curve, $\rho(\alpha)$ is a strictly increasing, concave function. Furthermore $\lim_{\alpha_0 \to 0} \rho_\alpha(\alpha_0) = \infty$ and $\lim_{\alpha_0 \to 1} \rho_\alpha(\alpha_0) = 0$. As a result, by Intermediate Value Theorem (IVP), (28) **always** has a unique solution α^*

Next, we argue that α^* obtained by solving 28 is less than α_0^c for $\pi < \tilde{\pi}_c$. For convenience of notation, let us represent the LHS of 28 by $f(\alpha)$ and RHS by $g(\pi)$. Note

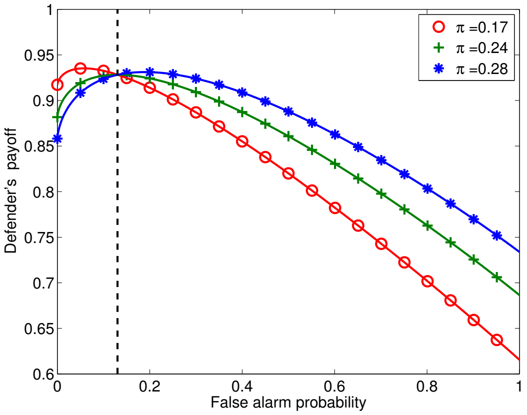


Figure 3.6: For $\gamma^* = 1$, characterization of distribution utility's payoff $u^{\mathrm{D}}(\alpha, 1)$ for $\pi < \widetilde{\pi}_c, \pi = \widetilde{\pi}_c, \pi > \widetilde{\pi}_c$

that f is decreasing in α and $g(\pi)$ is decreasing in π. Consider $\alpha_1, \alpha_2, \pi_1, \pi_2$ such that, $f(\alpha_1) = g(\pi_1), f(\alpha_2) = g(\pi_2)$ Furthermore consider $\pi_1 < \pi_2$ which means $g(\pi_1) > g(\pi_2)$ and $f(\alpha_1) > f(\alpha_2)$. Since f is a decreasing function, $\alpha_1 < \alpha_2$. Substituting $\alpha_2 = \alpha_0^c$ and $\pi_2 = \widetilde{\pi}_c$, we get, $\pi_1 < \widetilde{\pi}_c$ implies $\alpha_1 < \alpha_0^c$. Please note that $\widetilde{\pi}_c$ is the fraction of fraudulent customers for $\alpha^* = \alpha_0^c$. □

From (26), we see that for this case, $\gamma^* = 1$ is and only if,

$$\Phi(\alpha) > \mathrm{F} + \mathrm{C}_\eta \tag{29}$$

Equivalently, (29) implies $\alpha < \alpha_0^c$. From Lemma 1, $\alpha^* < \alpha_0^c$. Hence, $\gamma^* = 1$ is unique BR to $(\alpha^*, 1)$. To visualize the result, let us look at distribution utility's utility for three cases $\pi < \widetilde{\pi}_c, \pi = \widetilde{\pi}_c, \pi > \widetilde{\pi}_c$ with $\gamma^* = 1$ in Figure 3.6.

Case (iii) $\left[\eta > \eta_c^I, \quad \pi > \widetilde{\pi}_c\right]$

Consider a strategy profile (α, γ) such that $\alpha, \gamma \in (0, 1)^2$. Then for (α, γ) to be an equilibrium, we necessarily need $\Phi(\alpha) = \mathrm{F} + \mathrm{C}_\eta$. Solving above, we obtain $\rho(\alpha) = \rho_c$, equivalently, $\alpha = \alpha_0^c$. Next we show that α_0^c is a unique equilibrium response of the distribution utility. Again,

$$\partial_\alpha u^{\mathrm{D}}(\alpha_0^c, \gamma) = \pi\gamma\left(\partial_\alpha \rho(\alpha_0^c)\left(\mathrm{F} + pq\eta - \mathrm{C}_\rho\right)\right) - (1 - \pi\gamma)\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) = 0 \tag{30}$$

Solving for γ, $\gamma^* = \widetilde{\pi}_c/\pi$.

Note that $\gamma^* \in [0,1)$ cannot be an equilibrium for $\pi \in (0, \widetilde{\pi}_c)$. Assume the contrary and let there exist a $\gamma^* \in [0,1)$. But we know, for $\pi < \widetilde{\pi}_c$, $\partial_\alpha u^{\mathrm{D}} = \pi\gamma((\mathrm{C}_\alpha + \mathrm{C}_\rho) + (pq\eta + \mathrm{F} - \mathrm{C}_\rho)\rho'(\alpha)) - (\mathrm{C}_\alpha + \mathrm{C}_\rho) < 0 \;\; \forall\, \gamma$. Then, $\alpha^* = 0$. But we know that $\gamma^* = 1$ is a BR for $\alpha^* = 0$ which is a contradiction. □

Remark 4. : We obtain the equilibrium payoffs for game $\mathcal{G}_\eta^{\mathrm{ROC}}$ by plugging (α^*, γ^*) from Proposition 2 into (23) and (24),

$$u^{\mathrm{D}}\left(\alpha^*, \gamma^*\right) = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ -\left(1-\pi\right)\kappa_1\alpha^* + \pi\left(\kappa_2\rho\left(\alpha^*\right) - pq\eta\right) & \text{if } \eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c \\ -\left(1-\widetilde{\pi}_c\right)\kappa_1\alpha_0^c + \widetilde{\pi}_c\left(\kappa_2\rho_c - pq\eta\right) & \text{if } \eta > \eta_c^I, \pi > \widetilde{\pi}_c \end{cases} \tag{31}$$

and

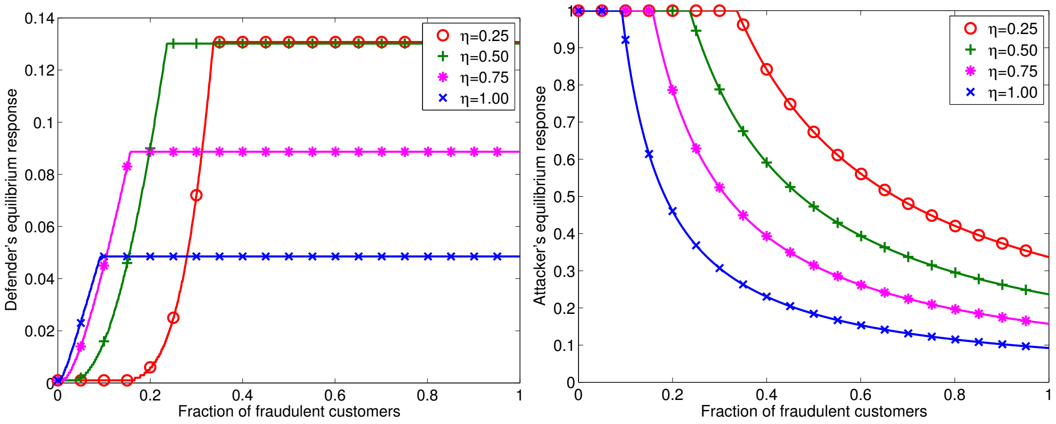
$$u^{\mathrm{f}}\left(\alpha^*, \gamma^*\right) = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ \left(1-\rho\left(\alpha^*\right)\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right) & \text{if } \eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c \\ 0 & \text{if } \eta > \eta_c^I, \pi > \widetilde{\pi}_c \end{cases} \tag{32}$$

where $\kappa_1 := (\mathrm{C}_\alpha + \mathrm{C}_\rho)$ and $\kappa_2 := (\mathrm{F} + pq\eta - \mathrm{C}_\rho)$ Furthermore, like before, using (5) and (4) the total profit collected by the distribution utility is $\Pi^{\mathrm{D}} = u^{\mathrm{D}}(\beta^*, \gamma^*) + A + pq - cq$. We will use Π^{D} in simulations to show the equilibrium payoff of the distribution utility.

In the following sections, we show the variation of equilibrium strategies and payoff for both distribution utility (or *defender*) and fraudulent customer (or *attacker*) with theft level η.

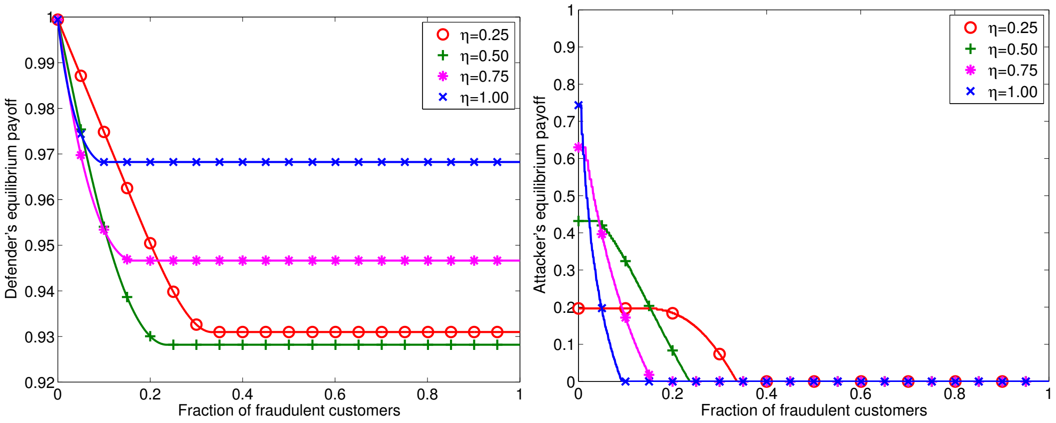
3.3.1 Equilibrium Analysis - Level of Theft

Figures 3.7a, 3.7b and 3.8a , 3.8b illustrate the equilibrium strategy and player payoffs for game $\mathcal{G}_\eta^{\mathrm{ROC}}$. In Figure 3.9a, we show the variation of effective detection probabil-



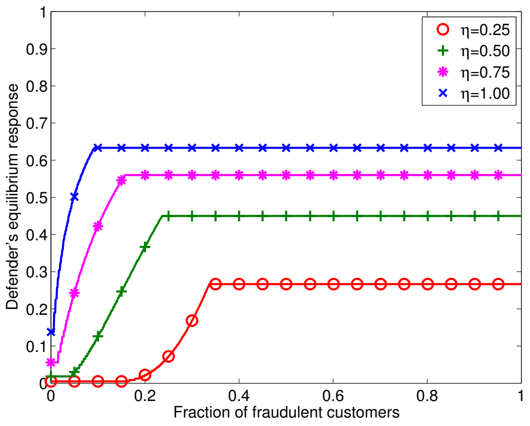
(a) (b)

Figure 3.7: Equilibrium Strategies for the game $\mathcal{G}_{\eta}^{\mathrm{ROC}}$: Variation with fraction of fraudulent customers π for different theft levels η



(a) (b)

Figure 3.8: Equilibrium Payoffs for the game $\mathcal{G}_{\eta}^{\mathrm{ROC}}$: Variation with fraction of fraudulent customers π for different theft levels η



(a)

Figure 3.9: Equilibrium detection probability for the game $\mathcal{G}_{\eta}^{\text{ROC}}$

ity with π for different levels of theft.

Fraudulent Customer's Equilibrium Response

As shown in Figure 3.7b, when $\pi > \widetilde{\pi}_c$, the fraudulent customer's probability of committing theft decreases in π. Indeed in equilibrium as π increases, the distribution utility becomes more effective in deterring the fraudulent customers. Moreover as theft level η increases, (a) the range of π for which $\gamma^* = 1$ decreases (i.e. π_c is smaller) (b) For a given fraction of fraudulent customers π, the probability of committing theft decreases with increasing theft. These results follow from the fact that with increasing theft level, the fraudulent customer derives the same "effective" or expected theft with a lesser probability of attack, and the defender's effective probability of detection improves for the chosen optimal false alarm rate α^*.

From Figure 3.8b, we observe that the equilibrium payoff for the fraudulent customer decreases with π. This can be understood by the fact that as fraction of fraudulent customers increases, the distribution utility improves its detection/investigation probability becomes more aggressive and the equilibrium payoff for the individual customer decreases. Furthermore, as seen in Figure 3.8b, the fraudulent customer derives zero payoff when $\pi > \widetilde{\pi}_c$. Indeed from Proposition 2, after $\widetilde{\pi}_c$, the distribution utility makes the fraudulent customer, indifferent between Theft **T** and No Theft **NT** and consequently the fraudulent customer derives zero payoff.

We also make the following interesting observation: As π increases, a smaller theft level produces larger equilibrium payoff for the fraudulent customer. Equivalently, if the fraudulent customer has the leverage to choose the optimal theft level, then she should choose smaller theft level as the fraction of fraudulent customer increases. Moreover, as we have seen earlier, the equilibrium payoff of the fraudulent customer decreases with increasing π. We will formally state and prove the above observations in Section 3.3.2, Proposition 3.

Distribution Utility's Equilibrium Response

In contrast to default configuration, from Figure 3.8a we observe that the distribution utility never chooses zero false alarm probability. In fact, from Proposition 2, it is evident that the distribution utility chooses $\alpha^* \to 0$ as the fraction of fraudulent customer approaches zero, i.e. $\pi \to 0$. Furthermore we observe that for $\pi < \widetilde{\pi}_c$ (22), the distribution utility chooses α^* less than the critical false alarm probability α_0^c (21) because for small π, the increase in false alarm costs more than offset the revenue recovered in investigation. Furthermore as π increases, the distribution utility recovers larger revenue and consequently chooses higher false alarm probability. Finally, since $\alpha^* < \alpha_0^c$ for $\pi < \widetilde{\pi}_c$, from Theorem 2, the fraudulent customer always commits Theft **T**. Hence, we observe in Figure 3.8b, the distribution utility's payoff decreases with increasing fraction of fraudulent customers (31).

Like the default configuration, we observe in Figure 3.8a that the distribution utility chooses a constant false alarm probability $\alpha^* = \alpha_0^c$ for $\pi > \widetilde{\pi}_c$ in order to make the fraudulent customer indifferent between Theft **T** and No Theft **NT**. Furthermore since the fraudulent customer randomizes between **T** and **NT**, he obtains zero payoff at equilibrium. Consequently we observe that, from Figure 3.8b, the distribution utility obtains a constant payoff for $\pi > \widetilde{\pi}_c$.

Lastly, we also make the following interesting observation: Although the detection probability increases with increase in theft level from Figure 3.9a, the optimal false alarm probability α^* chosen does not have a monotonic relationship with η. We will state the intuitive reason here and rigorously analyze this apparent paradox in section

4.3.1, Proposition 10. Note that from the definition of ROC curve (more specifically for normal distribution) that, for a given level of false alarm probability, the probability of detection increases with increase in theft level η. We also know that the probability of detection increases with increase in false alarm probability α. As a result as theft level increases, the distribution utility gets higher detection probability. However it may be sufficient or insufficient at the current theft level to make the fraudulent customer indifferent between Theft **T** and No Theft **NT**. As a result, the distribution utility will choose either a higher or lower false alarm probability at equilibrium.

3.3.2 Hipster effect and Fraudulent Customers

In this section, we want to study the effect of fraction of fraudulent customers on equilibrium payoff of the fraudulent customer who has leverage to find the optimal level of theft. Since the level of theft chosen will be the best case for the fraudulent customer, we can use the following analysis and results to argue for variation of his payoff for any arbitrary level of theft.

We will use the game-theoretic framework introduced earlier and propose the following two-stage formulation. In stage 1, the fraudulent customer chooses the optimal level of theft η^*. In stage 2, the distribution utility and fraudulent customer play a simultaneous game where the distribution utility chooses the optimal false alarm probability α^* for a given ROC curve $\rho(\alpha)$ and the fraudulent customer chooses the optimal probability of attack γ^*. Define $\Lambda(\eta)$,

$$\Lambda(\eta) := (\mathrm{T}\eta q + \mathrm{F} - \mathrm{C}_\rho)\rho_\alpha(\alpha_c(\eta), \eta) \tag{33}$$

where $\alpha_c(\eta) = \phi\{\phi^{-1}(\frac{\mathrm{T}q\eta - \mathrm{C}_\eta}{\mathrm{T}q\eta + \mathrm{F}}) - \frac{q\eta}{\sigma}\}$ from Proposition 2.

Proposition 3. *For a given ROC curve $\rho(\alpha)$, the equilibrium payoff u^{f^*} decreases with increasing fraction of fraudulent customers π, i.e. for $\pi_{\mathrm{L}} \leqslant \pi_{\mathrm{H}}$, we have, $u^{\mathrm{f}^*}_{\mathrm{L}} \geqslant u^{\mathrm{f}^*}_{\mathrm{H}}$*

Proof. From Proposition 2, we know that $u^{\mathrm{f}} > 0$ for $\pi < \widetilde{\pi}_c$ and $u^{\mathrm{f}} = 0$ for $\pi \geqslant \widetilde{\pi}_c$. Furthermore, we observed that $\widetilde{\pi}_c$ is a function of η. Consequently, we claim that to obtain a nonzero payoff, the fraudulent customer will choose η such that $\pi < \widetilde{\pi}_c$. Equivalently, following from (33) and from Proposition 2, optimal η must satisfy,

$$\Lambda(\eta) < \frac{1-\pi}{\pi}\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) \tag{34}$$

Following (34), we define $\mathcal{N} := \left\{\eta : \eta \in [0,1] \text{ and } \Lambda(\eta) < \frac{1-\pi}{\pi}\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right)\right\}$. Similarly we define $\mathcal{N}_{\mathrm{L}}, \mathcal{N}_{\mathrm{H}}$ for fraction $\pi_{\mathrm{L}}, \pi_{\mathrm{H}}$ respectively, Consequently, the attacker will obtain for $\mathcal{N} \neq \varnothing$,

$$\begin{aligned} u^{\mathrm{f}*}(\alpha^*, (\gamma^*, \eta^*)) :&= \max_{\eta} \gamma^*(\eta) \left(\left(1-\rho(\alpha^*(\eta))\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right) \\ &= \max_{\eta \in \mathcal{N}} \left(\left(1-\rho(\alpha^*(\eta))\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right) \end{aligned}$$

Furthermore, from (34), $\mathcal{N}_{\mathrm{H}} \subset \mathcal{N}_{\mathrm{L}}$ because $\frac{1-\pi_{\mathrm{L}}}{\pi_{\mathrm{L}}}\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) > \frac{1-\pi_{\mathrm{H}}}{\pi_{\mathrm{H}}}\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right)$. As a result,

$$\begin{aligned} &\max_{\eta \in \mathcal{N}_{\mathrm{L}}} \left(\left(1-\rho(\alpha^*(\eta))\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right) \geqslant \\ &\max_{\eta \in \mathcal{N}_{\mathrm{H}}} \left(\left(1-\rho(\alpha^*(\eta))\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right) \end{aligned}$$

This completes the proof of the proposition. □

Remark 5. The optimal level of theft is given by,

$$\eta^* = \arg \max_{\eta \in \mathcal{N} \neq \varnothing} \left(\left(1-\rho(\alpha^*(\eta))\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right)$$

where $\alpha^*(\eta)$ is the unique solution of $\rho_\alpha(\alpha^*, \eta) = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{\mathrm{T}q\eta + \mathrm{F} - \mathrm{C}_\rho}$

3.4 Optimal Configuration of the IDS

In this section, we consider the extensive form game $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$ where the distribution utility chooses the optimal false alarm probability α^{*} for a given ROC curve $\rho(\alpha)$ in Stage 1. Subsequently in Stage 2, both distribution utility and fraudulent/genuine customer play a simultaneous game $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$. As we are considering a two-stage game, let us revisit the definition of subgame perfect equilibrium to argue for the equilibrium strategy by the distribution utility and fraudulent customers.

Subgame Perfect Equilibrium: A strategy profile s^{*} is a Subgame Perfect Nash equilibrium (SPE) in game $\mathcal{G}$ if for any subgame $\mathcal{G}'$ of $\mathcal{G}$, $\mathrm{s}^{*}|_{\mathcal{G}'}$ is a Nash equilibrium of $\mathcal{G}'$.

For finite-horizon games, we can find the subgame perfect equilibrium using a technique called *Backward Induction.* Loosely speaking, backward induction refers to starting from the last subgames of a finite game, then finding the best response strategy profiles or the Nash equilibria in the subgames, then assigning these strategies profiles and the associated payoffs to be subgames, and moving successively towards the beginning of the game. Another classical result is that, backward induction gives the entire set of SPE. Furthermore, every finite extensive form game G has a Subgame Perfect Equilibrium.

In accordance with the backward induction principle, we will first solve the subgame $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$ to find $\left(\beta^{\dagger}, \gamma^{\dagger}\right)$ for a given default configuration of the IDS (α_0, ρ_0). Subsequently, in stage 1, we will find the optimal α_0^{*}, ρ_0^{*} for given ROC curve $\rho(\alpha)$ and Stage 2 equilibrium.

3.4.1 Game $\widehat{\mathcal{G}}_{\eta}^{\mathbf{ROC}}$: Perfect Information

In Stage 1, the distribution utility obtains the optimal default false alarm probability from the following optimization formulation:

$$\max_{\alpha_0,\rho_0} \ \Pi^{\mathrm{D}}(\alpha_0, \rho_0) = A + pq + u^{\mathrm{D}}((\alpha_0, \rho_0, \beta), \gamma) - cq$$

$$s.t. \quad \rho_0 \leqslant \rho(\alpha_0), \alpha_0 \in [0, 1]$$

Since only u^{D} depends on (α_0, ρ_0),

$$\Leftrightarrow \quad \max_{\alpha_0,\rho_0} \; u^{\mathrm{D}}((\alpha_0, \rho_0, \beta), \gamma) \\ s.t. \; \rho_0 \leqslant \rho(\alpha_0), \alpha_0 \in [0, 1] \tag{35}$$

Remark 6. Under the ROC curve specified by $\rho(\alpha)$, the constraint $\rho_0 \leqslant \rho(\alpha)$ is equivalent to $\rho_0 = \rho(\alpha)$ for (35). From (5), we can argue that any (ρ', α_0) such that $\rho' < \rho_0$ is dominated by (ρ_0, α_0).

Optimization Formulation:

$$\Leftrightarrow \quad \max_{\alpha_0} \; u^{\mathrm{D}}((\alpha_0, \beta), \gamma) \\ s.t. \; \alpha_0 \in [0, 1] \tag{36}$$

where under default IDS specified by ROC curve $\rho(\alpha)$, player payoffs in the game $\widehat{\mathcal{G}}_\eta^{\mathrm{ROC}}$ for a strategy profile $((\alpha_0, \beta), \gamma)$ can be written as follows:

$$u^{\mathrm{D}}((\alpha_0, \beta), \gamma) = \pi\gamma\left(\beta\rho(\alpha_0)\left(\mathrm{F} - \mathrm{C}_\rho\right) - \left(1 - \beta\rho(\alpha_0)\right)pq\eta\right) - (1 - \pi\gamma)\,\beta\alpha_0\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) \tag{37}$$

$$u^{\mathrm{f}}((\alpha_0, \beta), \gamma) = \gamma\left(\left(1 - \beta\rho(\alpha_0)\right)\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right)\right) \tag{38}$$

In the following result, we show that, in equilibrium, η_c^I is the critical fraction of theft level η below which the fraudulent customers do not attack with probability 1. The results also shows that $\widetilde{\pi}_c$ is the critical fraction of fraudulent customers below which attackers always attacks provided $\eta \geqslant \eta_c^I$. Finally, when $\pi > \widetilde{\pi}_c$, the attacker is indifferent between **T**(Theft) and **NT**(No Theft) actions for the detection probability $\rho = \rho_c$.

Theorem 2. ***Game with Optimal default IDS $\widehat{\mathcal{G}}_\eta^{ROC}$ Equilibrium***

Consider the game $\widehat{\mathcal{G}}_\eta^{ROC}$ with a default IDS specified by ROC curve $\rho(\alpha)$ and for a *given tariff schedule* $\mathrm{T}(q) := A + pq$. *The equilibrium of* $\widehat{\mathcal{G}}_\eta^{ROC}$ *is unique and given by:*

$$\left((\alpha_0^*, \beta^\dagger), \gamma^\dagger\right) = \begin{cases} (0,0;0) & \text{if } \eta \leqslant \eta_c^I \\ \left(\chi^{-1}\left\{\frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha+\mathrm{C}_\rho)}{pq\eta+\mathrm{F}-\mathrm{C}_\rho}\right\}, 1; 1\right) & \text{if } \eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c \\ \left(\alpha_0^c, 1; \frac{\widetilde{\pi}_c}{\pi}\right) & \text{if } \eta > \eta_c^I, \pi > \widetilde{\pi}_c \end{cases} \tag{39}$$

where $\chi(\alpha) = \partial\rho/\partial\alpha$. *Furthermore, for the reported consumption represented by random variable* $\boldsymbol{Y}$ *with realization* $\boldsymbol{y}$*, the statistical test implemented by the distribution utility in equilibrium is given by* $\left\{Investigate(\boldsymbol{I}) : \boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{y} < (\phi^{-1}(\alpha_0^*) + \frac{q\eta}{\sigma})\right\}$

Remark 7. We obtain the equilibrium payoffs for game $\widehat{\mathcal{G}}_\eta^{\mathrm{ROC}}$ by plugging $((\alpha_0, \rho_0, \beta), \gamma)$ from Theorem 2 into (37) and (38),

$$u^{\mathrm{D}}\left((\alpha_0^*, \beta^\dagger), \gamma^\dagger\right) = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ -(1-\pi)\kappa_1\alpha_0^* + \pi\left(\kappa_2\rho\left(\alpha_0^*\right) - pq\eta\right) & \text{if } \eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c \\ -(1-\widetilde{\pi}_c)\kappa_1\alpha_0^c + \widetilde{\pi}_c\left(\kappa_2\rho_c - pq\eta\right) & \text{if } \eta > \eta_c^I, \pi > \widetilde{\pi}_c \end{cases} \tag{40}$$

and

$$u^{\mathrm{f}}\left((\alpha_0^*, \beta^\dagger), \gamma^\dagger\right) = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ (1-\rho\left(\alpha_0^*\right))\left(\mathrm{F} + pq\eta\right) - \left(\mathrm{F} + \mathrm{C}_\eta\right) & \text{if } \eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c \\ 0 & \text{if } \eta > \eta_c^I, \pi > \widetilde{\pi}_c \end{cases} \tag{41}$$

where $\kappa_1 = (\mathrm{C}_\alpha + \mathrm{C}_\rho)$ and $\kappa_2 = (\mathrm{F} + pq\eta - \mathrm{C}_\rho)$

Lemma 2. *For an increasing, concave ROC curve* $\rho(.)$ *with* $\rho(0) = 0,\ \rho(1) = 1$ *then* $\rho'(\alpha) \leqslant \rho(\alpha)/\alpha$ *for all* $\alpha \in (0, 1]$.

Proof. Let us first geometrically argue that no tangent to the ROC curve with the given properties can have a negative intercept. Precisely, let us represent the equation of tangent at $(\alpha, \rho(\alpha))$ as: $y = \frac{\partial\rho}{\partial\alpha}x + C$. We claim that $C \geqslant 0$ for any tangent to the ROC curve $\rho(.)$ above.

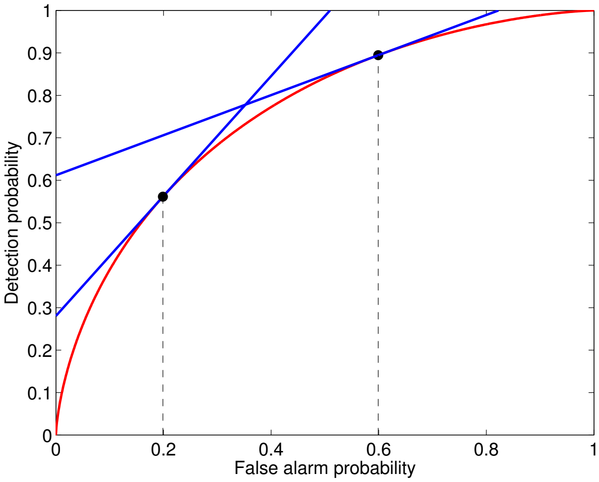


Figure 3.10: Relationship between $\rho(\alpha_0)/\alpha_0$ and $\rho'(\alpha_0)$

Assume the contrary. Let there exist a tangent that has a negative intercept term. We know this is possible only if the tangent has its X-intercept $\geqslant 0$. But this violates the concavity of the ROC curve.

Lastly substituting $y = \rho(\alpha)$ and $x = \alpha$, we get the equation as $\rho(\alpha) = \frac{\partial \rho}{\partial \alpha}\alpha + C$. Since $C \geqslant 0$, we get $\rho(\alpha) - \frac{\partial \rho}{\partial \alpha}\alpha \geqslant 0$ □

Explanation of Theorem 2:

i. For theft level $\eta \leqslant \eta_c^I$, the fraudulent customer does not commit theft and the defender chooses 0 false alarm probability. Note that since the distribution utility will not use the IDS in Stage 2, i.e. $\beta^\dagger = 0$, he can choose an arbitrary default false alarm probability. For convenience, we have taken $\alpha_0 = 0$ which can be interpreted as lack of IDS. We can also posit that the distribution utility can choose $\alpha_0 = 0$ in Stage 1 and then can choose an arbitrary probability of using the IDS β. Hence, $\beta\alpha_0 = 0$ at equilibrium.

ii. The effective false alarm probability is same for both sequential game with complete information $\widehat{\mathcal{G}}_\eta^{\mathrm{ROC}}$ and simultaneous game with tunable IDS $\mathcal{G}_\eta^{\mathrm{ROC}}$, i.e. $\beta^\dagger\alpha_0^* = \alpha^*$.

iii. The distribution utility always uses the IDS, i.e. $\beta^\dagger = 1$ for an optimally configured IDS with complete information ($\widehat{\mathcal{G}}_\eta^{\mathrm{ROC}}$). Intuitively, this follows from concavity of ROC curve such that for a given effective false alarm probability, the detection probability of operating on the ROC curve is strictly better than any linear combination of other operating points. We will rigorously show this in the following proof of the theorem.

Proof. We can re-write (38) as follows:

$$u^{\mathrm{f}}(\alpha_0, \beta; \gamma) = \gamma \left[\Phi(\alpha_0, \beta) - (\mathrm{F} + \mathrm{C}_\eta)\right] \tag{42}$$

where $\Phi(\alpha_0, \beta) := (1 - \beta\rho(\alpha_0))\,(p\eta q + \mathrm{F})$ and $\rho_0 = \rho(\alpha_0)$

Case (i) $\left[\eta \leqslant \eta_c^I\right]$

For this case, we see from (11) that $pq\eta \leqslant \mathrm{C}_\eta$, which implies that $\Phi(\alpha_0, \beta) \leqslant (1 - \beta\rho_0)\,(\mathrm{C}_\eta + \mathrm{F}) \leqslant (\mathrm{C}_\eta + \mathrm{F})$. Thus, for any strategy profile $(\alpha_0, \beta; \gamma)$, $u^{\mathrm{f}}(\alpha_0, \beta; \gamma) \leqslant 0$. From (42), we conclude that, in equilibrium, $\gamma^* = 0$.

Now, for a strategy profile $(\alpha_0, \beta, 0)$, we can express (37) as follows $u^{\mathrm{D}}(\alpha_0, \beta; 0) = -\beta\alpha_0\,(\mathrm{C}_\alpha + \mathrm{C}_\rho)$. Thus, in equilibrium, $\alpha_0^*\beta^* = \alpha_I^* = 0$.

Case (ii) $\left[\eta > \eta_c^I, \quad \pi \leqslant \tilde{\pi}_c\right]$

Let us define region $\mathcal{R}(\pi)$ of false alarm probability α, under ROC curve $\rho(\alpha)$ using Proposition 1 and (12),

$$\mathcal{R}(\pi) := \left\{\alpha : \alpha \in [0,1],\, \pi < \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha}{(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha + \rho(\alpha)(pq\eta + \mathrm{F} - \mathrm{C}_\rho)}\right\} \tag{43}$$

Plan of the proof

- For $\gamma = 1$ and for any (α', β'), $\alpha' \leqslant 1$ and $\beta' < 1$, there exists an $(\tilde{\alpha}, 1)$, $\tilde{\alpha} \leqslant 1$, which dominates (α', β').

- For $\gamma = 1$, $(\alpha_0^*, 1)$ is the unique BR; where $\alpha_0^* \in \mathcal{R}$ is given by,

$$\rho_\alpha(\alpha_0^*) = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{\mathrm{T}q\eta + \mathrm{F} - \mathrm{C}_\rho}$$

- $\gamma = 1$ is a unique BR of $(\alpha_0^*, 1)$

Lemma 3. *For $\gamma = 1$ and for any (α', β'), $\alpha' \leqslant 1$ and $\beta' < 1$, there exists an $(\widetilde{\alpha}, 1)$, $\widetilde{\alpha} \leqslant 1$, which dominates (α', β').*

Proof. We define $\widetilde{\alpha} := \beta' {\alpha_0}'$

Assume the contrary, i.e. (β', α') such that $\beta' < 1$ and $\beta' {\alpha_0}' < 1$ gives a better payoff to the distribution utility than $(\widetilde{\alpha}, 1)$. Recall that $\rho(\alpha)$, by Assumption 2, is strictly concave and strictly increasing.

Then,

$$\rho(\widetilde{\alpha}) = \rho(\beta' {\alpha_0}') = \rho(\beta' {\alpha_0}' + (1 - \beta')\, 0) > \beta' \rho({\alpha_0}') + (1 - \beta')\, 0 \tag{44}$$

Consider two strategy profiles: $(\widetilde{\alpha}, 1, 1)$ and $({\alpha_0}', \beta', 1)$. From (37),

$$\begin{aligned} u^{\mathrm{D}}(\widetilde{\alpha}, 1; 1) - u^{\mathrm{D}}({\alpha_0}', \beta'; 1) &= \pi \left(\rho(\widetilde{\alpha}) \left(\mathrm{F} - \mathrm{C}_\rho\right) - (1 - \rho(\widetilde{\alpha}))\, pq\eta\right) - (1 - \pi)\, \widetilde{\alpha} \left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) \\ + (1 - \pi)\, \beta' {\alpha_0}' \left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) &- \pi \left(\beta' \rho({\alpha_0}') \left(\mathrm{F} - \mathrm{C}_\rho\right) - (1 - \beta' \rho({\alpha_0}'))\, pq\eta\right) \\ &= \pi \left(\rho(\widetilde{\alpha}) - \beta' \rho({\alpha_0}')\right) \left(\mathrm{F} + pq\eta - \mathrm{C}_\eta\right) > 0 \end{aligned} \tag{45}$$

From(44) and (45), we can see $u^{\mathrm{D}}({\alpha_0}', \beta'; 1) < u^{\mathrm{D}}(\alpha_0^*, 1; 1)$. **A contradiction.** □

Proposition 4. *For $\gamma = 1$, $(\alpha_0^*, 1)$ is the unique BR; where $\alpha_0^* \in \mathcal{R}$ is given by,*

$$\rho_\alpha(\alpha_0^*) = \frac{1 - \pi}{\pi} \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{\mathrm{T} q\eta + \mathrm{F} - \mathrm{C}_\rho}$$

Proof. Consider $\beta^* = 1$. For a strategy profile $(\alpha_0, 1, 1)$, from (37);

$$u^{\mathrm{D}}(\alpha_0, 1; 1) = \pi \left(\rho(\alpha_0) \left(\mathrm{F} - \mathrm{C}_\rho\right) - (1 - \rho(\alpha_0))\, pq\eta\right) - (1 - \pi)\, \alpha_0 \left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) \tag{46}$$

Solving for FOC $u^{\mathrm{D}}_{\alpha_0}(\alpha_0, 1; 1) = 0$,

$$\rho_\alpha(\alpha_0^*) = \frac{1 - \pi}{\pi} \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{\mathrm{T} q\eta + \mathrm{F} - \mathrm{C}_\rho} \tag{47}$$

By Assumption 2, $\rho(\alpha)$ is a strictly increasing, concave function. Furthermore $\lim_{\alpha_0 \to 0}$ $\rho_\alpha(\alpha_0) = \infty$ and $\lim_{\alpha_0 \to 1} \rho_\alpha(\alpha_0) = 0$. As a result, by Intermediate Value Theorem, (47) **always** has a unique solution α_0

Moreover, from Lemma 2, we know that

$$\rho(\alpha_0^*)/\alpha_0^* > \rho_\alpha(\alpha_0^*) = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{pq\eta + \mathrm{F} - \mathrm{C}_\rho} \tag{48}$$

which implies $\alpha_0 \in \mathcal{R}$.

Next, we argue that α_0^* obtained by solving 47 is less than α_0^c for $\pi < \tilde{\pi}_c$. For convenience of notation, let us represent the LHS of 47 by $f(\alpha_0)$ and RHS by $g(\pi)$. Note that f is decreasing in α_0 and $g(\pi)$ is decreasing in π. Consider $\alpha_1, \alpha_2, \pi_1, \pi_2$ such that, $f(\alpha_1) = g(\pi_1), f(\alpha_2) = g(\pi_2)$ Furthermore consider $\pi_1 < \pi_2$ which means $g(\pi_1) > g(\pi_2)$ and $f(\alpha_1) > f(\alpha_2)$. Since f is a decreasing function, $\alpha_1 < \alpha_2$. Substituting $\alpha_2 = \alpha_0^c$ and $\pi_2 = \tilde{\pi}_c$, we get, $\pi_1 < \tilde{\pi}_c$ implies $\alpha_1 < \alpha_0^c$. Please note that $\tilde{\pi}_c$ is the fraction of fraudulent customers for $\alpha_0^* = \alpha_0^c$.

To sum up, we have shown that (a) α_0^* always exists and is unique (b) $\alpha_0^* \in \mathcal{R}$ and (c) $\alpha_0^* < \alpha_0^c$, equivalently(from (21)) $\rho(\alpha_0^*) < \rho_c$. Hence, from Proposition 1, $\beta^\dagger(\alpha_0^*) = 1$ under subgame $\widehat{\mathcal{G}}_\eta^{\mathrm{ROC}}$. Therefore $(\alpha_0^*, 1)$ is a valid equilibrium.

Lastly, we show that there cannot be any other equilibrium that has a higher payoff. Assume the contrary and consider the strategy (β', α') for the following cases:

i. $\alpha' \in \mathcal{R}$ such that $\alpha' \neq \alpha_0^*$. Consider $\alpha' \leqslant \alpha_0^c$, then from Proposition 1, we know that $\beta^\dagger(\alpha') = 1$. However we have shown earlier that α_0^* is the unique BR. A contradiction. Consider $\alpha' > \alpha_0^c$, then from Proposition 1, we know that $\beta^\dagger(\alpha') = \rho_c/\rho(\alpha') < 1$. Furthermore, from Lemma 3, there exists $(\tilde{\alpha}, 1)$ such that it dominates (α', β'), or equivalently $u^{\mathrm{D}}((\alpha', \beta'), \gamma) < u^{\mathrm{D}}((\tilde{\alpha}, 1), \gamma)$. However, we have shown above that $u^{\mathrm{D}}((\tilde{\alpha}, 1), \gamma) < u^{\mathrm{D}}((\alpha_0^*, 1), \gamma)$ for all $\tilde{\alpha}$. Hence, $u^{\mathrm{D}}((\alpha', \beta'), \gamma) < u^{\mathrm{D}}((\alpha_0^*, 1), \gamma)$. A contradiction.

ii. $\alpha' \notin \mathcal{R}$, then from Proposition 1, $\beta^\dagger(\alpha') = 0$. Furthermore, from (48),

$$u^{\mathrm{D}}((\alpha_0^*, 1), 1) - u^{\mathrm{D}}((\alpha', 0), 1) =$$
$$\pi\left(\rho(\alpha_0^*)\left(\mathrm{F} + pq\eta - \mathrm{C}_\rho\right)\right) - (1-\pi)\,\alpha_0^*\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) > 0$$

A contradiction.

Hence for $\gamma = 1$, $(\alpha_0^*, 1)$ is the unique BR.

From (42), we see that for this case, $\gamma^* = 1$ is and only if,

$$\Phi(\alpha_0, \beta) > \mathrm{F} + \mathrm{C}_\eta \tag{49}$$

Equivalently, (49) implies $\alpha_0 < \alpha_0^c$. From Proposition 4, $\alpha_0^* < \alpha_0^c$. Hence, $\gamma^* = 1$ is unique BR to $(\alpha_0^*, 1)$.

□

Case (iii) $\left[\eta > \eta_c^I, \quad \pi > \tilde{\pi}_c\right]$

Consider a strategy profile $((\alpha_0, \beta), \gamma)$ such that $\alpha_0, \gamma \in (0, 1)$ and $\beta \in [0, 1]$. Then for $((\alpha_0, \beta), \gamma)$ to be an equilibrium, we necessarily need $\Phi(\alpha_0, \beta) = \mathrm{F} + \mathrm{C}_\eta$. Solving above, we obtain $\beta\rho(\alpha_0) = \rho_c$.

Lemma 4. *For $\gamma^* \in (0, 1)$ and for any $({\alpha_0}', \beta')$, $\beta' < 1$ and $\beta'\rho({\alpha_0}') = \rho_c$, there exists an $(\tilde{\alpha}, 1)$ which dominates $({\alpha_0}', \beta')$.*

Proof. We define $\tilde{\alpha}$, $\rho(\tilde{\alpha}) = \beta'\rho({\alpha_0}')$

Assume the contrary, i.e. (α', β') such that $\beta' < 1$ and $\beta'\rho({\alpha_0}') = \rho_c$ gives a better payoff to the distribution utility than $(\tilde{\alpha}, 1)$. Recall that $\rho(\alpha)$, by Assumption 2, is strictly concave and strictly increasing.

Then,

$$\rho(\beta'{\alpha_0}') = \rho(\beta'{\alpha_0}' + (1 - \beta')\, 0) > \beta'\rho({\alpha_0}') + (1 - \beta')\, 0 = \rho(\tilde{\alpha}) \tag{50}$$

Since $\rho(.)$ is a strictly increasing function, from (50), $\tilde{\alpha} < \beta'{\alpha_0}'$.

Consider two strategy profiles: $((\alpha_0^*, 1), \gamma^*)$ and $(({\alpha_0}', \beta'), \gamma^*)$. From (37),

$$\begin{aligned}
&u^{\mathrm{D}}((\tilde{\alpha}, 1), \gamma^*) - u^{\mathrm{D}}(({\alpha_0}', \beta'), \gamma^*) \\
&= \pi\gamma^* \left(\rho(\tilde{\alpha})\,(\mathrm{F} - \mathrm{C}_\rho) - (1 - \rho(\tilde{\alpha}))\, pq\eta\right) - (1 - \pi\gamma^*)\, \tilde{\alpha}\, (\mathrm{C}_\alpha + \mathrm{C}_\rho) \\
&- \pi\gamma^* \left(\beta'\rho({\alpha_0}')\,(\mathrm{F} - \mathrm{C}_\rho) - (1 - \beta'\rho({\alpha_0}'))\, pq\eta\right) + (1 - \pi\gamma^*)\, \beta'{\alpha_0}'\, (\mathrm{C}_\alpha + \mathrm{C}_\rho) \\
&= (1 - \pi\gamma^*)\, (\mathrm{C}_\alpha + \mathrm{C}_\rho)\, (\beta'{\alpha_0}' - \tilde{\alpha}) > 0
\end{aligned} \tag{51}$$

From(50) and (51), we get $u^{\mathrm{D}}((\alpha_0{}', \beta'), \gamma) < u^{\mathrm{D}}((\widetilde{\alpha}, 1), \gamma)$. **A contradiction.**. Hence $\widetilde{\alpha} = \alpha_0^c$ and $(\alpha_0^c, 1)$ dominates $(\alpha_0{}', \beta')$. Next we show that $(\alpha_0^c, 1)$ is a valid strategy profile. Firstly, from (21), $\rho(\alpha_0^c) = \rho_c$. Secondly, from (22),

$$\pi > \widetilde{\pi}_c = \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \rho'(\alpha_0^c)(pq\eta + \mathrm{F} - \mathrm{C}_\rho)}$$

Or equivalently,

$$\frac{\rho(\alpha_0^c)}{\alpha_0^c} > \rho'(\alpha_0^c) > \frac{1-\pi}{\pi} \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(pq\eta + \mathrm{F} - \mathrm{C}_\rho)}$$

where the first inequality follows from Lemma 2. Finally from Proposition 1, equivalently (14), $(\alpha_0^c, \rho(\alpha_0^c))$ falls in regime 3 or $\beta^\dagger(\alpha_0^c) = 1$. □

We now show that $(\alpha_0^c, 1)$ is a unique equilibrium response of the distribution utility. Again,

$$u_\alpha^{\mathrm{D}}((\alpha_0^c, 1), \gamma) = \pi\gamma\left(\rho_\alpha(\alpha_0^c)\left(\mathrm{F} + pq\eta - \mathrm{C}_\rho\right)\right) - (1 - \pi\gamma)\left(\mathrm{C}_\alpha + \mathrm{C}_\rho\right) = 0 \qquad (52)$$

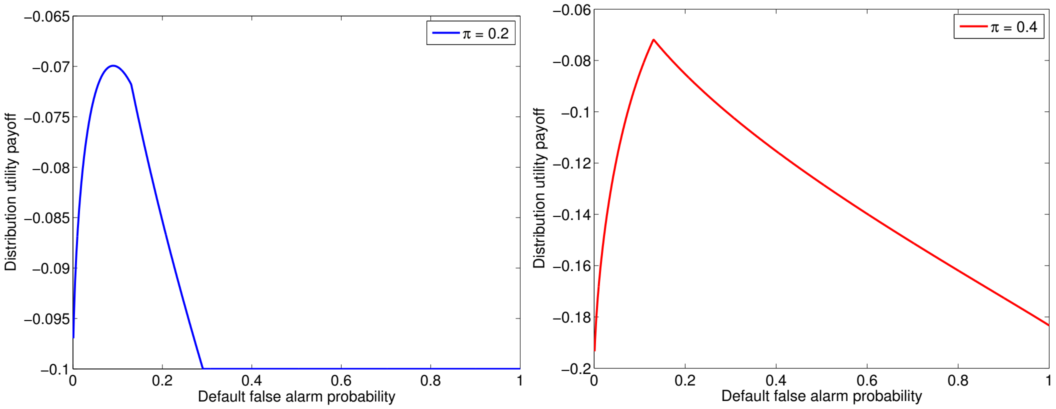
Solving for γ, $\gamma^* = \widetilde{\pi}_c/\pi$.

Note that $\gamma^* \in (0,1)$ for $\pi \in (0, \widetilde{\pi}_c))$ cannot be an equilibrium. Let $\exists \gamma^* \in [0,1)$, $\partial_\alpha u^{\mathrm{D}} = \pi(\mathrm{C}_\alpha + \mathrm{C}_\rho) + (pq\eta + \mathrm{F} - \mathrm{C}_\rho)\rho'(\alpha_0)\gamma^* - (\mathrm{C}_\alpha + \mathrm{C}_\rho) < 0 \;\; \forall\, \gamma$. Then, $\alpha_0^* = 0$. But we know that $\gamma^* = 1$ is a BR for $\alpha_0^* = 0$ which is a contradiction. □

In Figure 3.11, the optimal default false alarm probability for small prior $\pi < \pi_c$ and large prior $\pi > \pi_c$.

3.4.2 Game $\mathcal{G}^{\mathbf{ROC}}_{\mathrm{P}(\eta)}$: Imperfect Information

As described earlier, the distribution utility in game $\mathcal{G}^{\mathrm{ROC}}_{\mathrm{P}(\eta)}$, does not have perfect information about the theft level η committed by the fraudulent customer. Rather, the distribution utility has a prior (or knows the probability distribution) on theft level, $\mathrm{P}(\eta)$. Precisely, there are two levels of theft $\eta^{\mathrm{L}}, \eta^{\mathrm{H}}$ with probabilities λ^{L} and $\lambda^{\mathrm{H}}(= 1 - \lambda^{\mathrm{L}})$ respectively.



(a) (b)

Figure 3.11: Variation of u^{D} with α_0 for (a) $\pi = 0.2$ (b) $\pi = 0.4$

Like in the perfect information game $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$ earlier, for a given ROC $\rho(\alpha_0)$ the distribution utility obtains the optimal default false alarm probability in Stage 1 from the following optimization formulation:

$$\max_{\alpha_0} \; \Pi^{\mathrm{D}}(\alpha_0, \rho(\alpha_0)) = A + pq + \mathbb{E}_{\eta}\left[u^{\mathrm{D}}((\alpha_0, \rho(\alpha_0), \beta^{\dagger}(\alpha_0)), \gamma^{\dagger}(\alpha_0))\right] - cq$$
$$s.t. \quad \alpha_0 \in [0,1], \eta \sim \mathrm{P}(\eta)$$

Optimization Formulation:

$$\Leftrightarrow \quad \max_{\alpha_0} \; \mathbb{E}_{\eta}\left[u^{\mathrm{D}}((\alpha_0, \rho(\alpha_0), \beta^{\dagger}(\alpha_0)), \gamma^{\dagger}(\alpha_0))\right]$$
$$s.t. \quad \alpha_0 \in [0,1], \eta \sim \mathrm{P}(\eta) \tag{53}$$

Note that,

$$\mathbb{E}_{\eta}\left[u^{\mathrm{D}}((\alpha_0, \rho(\alpha_0), \beta^{\dagger}(\alpha_0)), \gamma^{\dagger}(\alpha_0))\right] = \lambda^{\mathrm{L}} u_{\mathrm{L}}^{\mathrm{D}}(\left(\alpha_0, \beta^{\mathrm{L}}\right), \gamma^{\mathrm{L}}) + \lambda^{\mathrm{H}} u_{\mathrm{H}}^{\mathrm{D}}(\left(\alpha_0, \beta^{\mathrm{H}}\right), \gamma^{\mathrm{H}}) \tag{54}$$

equivalently,

$$\alpha_0^* := \arg\max_{\alpha_0} \lambda^{\mathrm{L}} u_{\mathrm{L}}^{\mathrm{D}}(\left(\alpha_0, \beta^{\mathrm{L}}\right), \gamma^{\mathrm{L}}) + \lambda^{\mathrm{H}} u_{\mathrm{H}}^{\mathrm{D}}(\left(\alpha_0, \beta^{\mathrm{H}}\right), \gamma^{\mathrm{H}}) \tag{55}$$

where,

i. α_0^* is the Stage 1 optimal false alarm probability chosen by distribution utility.

ii. η^{L} is low theft level

iii. η^{H} is high theft level

iv. β^{H} is the equilibrium (subgame) probability of using the IDS for theft level η^{H}

v. β^{L} is the equilibrium (subgame) probability of using the IDS for theft level η^{L}

vi. $u_{\mathrm{H}}^{\mathrm{D}}$ is the equilibrium payoff of the distribution utility for theft level η^{H}

vii. $u_{\mathrm{L}}^{\mathrm{D}}$ is the equilibrium payoff of the distribution utility for theft level η^{L}

viii. γ^{H} is the equilibrium probability of committing η^{H} theft level by the fraudulent customer

ix. γ^{L} is the equilibrium probability of committing η^{L} theft level by the fraudulent customer

Furthermore, under default IDS specified by ROC curve $\rho(\alpha)$, $(\beta^{\dagger}, \gamma^{\dagger})$ is given by Proposition 1 and player payoffs $u^{\mathrm{D}}, u^{\mathrm{f}}$ for a strategy profile $((\alpha_0, \beta), \gamma)$ are given by (37) and (38).

Lemma 5. *For $\pi < \widehat{\pi}_c$, where $\widehat{\pi}_c$ is the critical fraction of fraudulent customers*

$$\widehat{\pi}_c = \frac{\mathrm{C}_\alpha + \mathrm{C}_\rho}{\mathrm{C}_\alpha + \mathrm{C}_\rho + pq\eta + \mathrm{F} - \mathrm{C}_\rho} \tag{56}$$

there exists a unique $\widehat{\alpha}_c \in (0,1)$ such that,

$$\frac{\rho(\widehat{\alpha}_c)}{\widehat{\alpha}_c} = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{\mathrm{T}q\eta + \mathrm{F} - \mathrm{C}_\rho} \tag{57}$$

Proof. We first show that for a given ROC curve $\rho(\alpha)$ and for any α_1 and α_2 such that $\alpha_1 < \alpha_2$, we have $\rho(\alpha_1)/\alpha_1 > \rho(\alpha_2)/\alpha_2$. Furthermore, $\rho(\alpha)/\alpha$ achieves its minimum value 1 at $\alpha = 1$ and is unbounded at $\alpha = 0$.

Consider for notational convenience, $\Psi(\alpha) = \rho(\alpha)/\alpha$. Then,

$$\Psi'(\alpha) = \frac{\alpha\rho'(\alpha) - \rho(\alpha)}{\alpha^2} < 0$$

where the inequality follows from Lemma 2. Furthermore following from the monotonicity of $\Psi(\alpha)$, it achieves its minimum at $\alpha = 1$. Therefore $\Psi^{min}(1) = 1$. Additionally, using L'Hospital rule,

$$\lim_{\alpha\to 0} \frac{\rho(\alpha)}{\alpha} = \lim_{\alpha\to 0} \rho'(\alpha) = \infty$$

proving it to be unbounded at $\alpha = 0$.

From (56), we can see that for $\pi > \widehat{\pi}_c$, $\frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha+\mathrm{C}_\rho)}{\mathrm{T}q\eta+\mathrm{F}-\mathrm{C}_\rho} < 1$, or equivalently, $\frac{\rho(\widehat{\alpha}_c)}{\widehat{\alpha}_c} < 1$. But we have shown earlier that $\rho(\alpha)/\alpha$ is a decreasing function of α and is equal to 1 at $\alpha = 1$. Hence, (57) has no solution for $\pi < \widehat{\pi}_c$. Lastly, for $\pi < \widehat{\pi}_c$, there exists a unique $\widehat{\alpha}_c \in (0, 1)$. The existence follows from Intermediate Value Theorem(IVP) and uniqueness follows from monotonicity of $\rho(\alpha)/\alpha$. □

Discussion: *Significance of* $\widehat{\alpha}_c$

Lemma 5 characterizes the existence and uniqueness of $\widehat{\alpha}_c$ with π. From Proposition 1, and by monotonicity of $\rho(\alpha)/\alpha$, we argue, using (12), that $\alpha_0 < \widehat{\alpha}_c$ implies $\pi > \pi_c$. Hence we can study the effect of α_0 on the subgame equilibrium $(\beta^\dagger, \gamma^\dagger)$(Theorem 2) using $\widehat{\pi}_c, \widehat{\alpha}_c$.

Before we outline the lower and upper bounds for α_0^* (defined in (55)), we will characterize the payoff of the distribution utility for game $\widehat{\mathcal{G}}_\eta^{\mathrm{ROC}}$ with single level of theft η. Intuitively, in the forthcoming propositions 5, 6, we show that distribution utility obtains a higher payoff as α_0 gets closer to α_0^*. Precisely, we argue for the monotonicity of defender's payoff for $\alpha_0 < \alpha_0^*$ and $\alpha_0 > \alpha_0^*$ with both $\pi < \widetilde{\pi}_c$ (Proposition 5) and $\pi > \widetilde{\pi}_c$ (Proposition 6).

Proposition 5. *For all* $\pi < \widetilde{\pi}_c$ *and for any* α^1 *and* α^2, $\alpha^1 < \alpha^2 < \alpha^*$ *the distri-*

bution utility obtains a greater payoff for α^2 than α^1 such that, $u^{\mathrm{D}}((\alpha^1,\beta^1),\gamma^1) \leqslant u^{\mathrm{D}}((\alpha^2,\beta^2),\gamma^2)$. Similarly, for $\alpha^2 > \alpha^1 > \alpha^*$ the distribution utility obtains a greater payoff for α^1 than α^2 i.e., $u^{\mathrm{D}}((\alpha^1,\beta^1),\gamma^1) \geqslant u^{\mathrm{D}}((\alpha^2,\beta^2),\gamma^2)$.

Proof. As we have seen in Theorem 2, the BR of the defender is $(\alpha^*, 1)$ where

$$\rho_\alpha(\alpha^*) = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{\mathrm{T}q\eta + \mathrm{F} - \mathrm{C}_\rho}$$

And, from Lemma 2, $\frac{\rho(\alpha^*)}{\alpha^*} > \rho_\alpha(\alpha^*) = \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha+\mathrm{C}_\rho)}{\mathrm{T}q\eta+\mathrm{F}-\mathrm{C}_\rho}$

Case (i) $\pi < \widetilde{\pi}_c \leqslant \widehat{\pi}_c$:

Consider $\alpha_1 < \alpha_2 < \alpha^*$. From Lemma 5, we have $\rho(\alpha_1)/\alpha_1 > \rho(\alpha_2)/\alpha_2 > \rho(\alpha^*)/\alpha^* > \rho_\alpha(\alpha^*)$. Recall from Proposition 1, this implies that α_1, α_2 lie in regime 3, or equivalently, $\beta^\dagger(\alpha_1) = 1$ and $\beta^\dagger(\alpha_2) = 1$. But for $\beta = 1$, since u^{D} is a concave function, we know that $u^{\mathrm{D}}((\alpha_1, 1), 1) < u^{\mathrm{D}}((\alpha_2, 1), 1)$.

Next, consider $\alpha_2 > \alpha_1 > \alpha^*$. From Proposition 1, we know that for $\alpha > \widehat{\alpha}_c$, α is in regime 2, or equivalently, $\beta^\dagger(\alpha) = 0$ and the payoff $u^{\mathrm{D}} = -\pi pq\eta$. Let us consider the following two cases:

i. For $\widehat{\alpha}_c \leqslant \alpha_0^c$, let us consider the following two cases:

   (a) $\alpha_1 < \widehat{\alpha}_c$, $\alpha_2 > \widehat{\alpha}_c$: For $\alpha_1 < \widehat{\alpha}_c < \alpha_0^c$, we know that α_1 lies in regime 3, or equivalently, $\beta^\dagger(\alpha_1) = 1$. Furthermore, $u^{\mathrm{D}}((\alpha_1, 1), 1) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_1 + \pi((-\mathrm{C}_\rho + \mathrm{F})\rho(\alpha_1) - (pq\eta)(1-\rho(\alpha_1)))$. Lastly, from (57), we can see that $u^{\mathrm{D}}((\alpha_1, 1), 1) > u^{\mathrm{D}}((\alpha_2, 0), 1)$

   (b) $\alpha_1 > \widehat{\alpha}_c$, $\alpha_2 > \widehat{\alpha}_c$: For $\alpha_1, \alpha_2 > \widehat{\alpha}_c$, we have $u^{\mathrm{D}}((\alpha_1, 0), 1) = u^{\mathrm{D}}((\alpha_2, 0), 1)$ as shown above.

ii. For $\widehat{\alpha}_c > \alpha_0^c$, let us consider the following cases:

   (a) $\alpha_1 < \alpha_0^c$, $\alpha_2 < \alpha_0^c$: In this case, we know that for both α, $\rho(\alpha) < \rho_c$ and $\frac{\rho(\alpha)}{\alpha} > \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha+\mathrm{C}_\rho)}{\mathrm{T}q\eta+\mathrm{F}-\mathrm{C}_\rho}$. Consequently, both α lie in regime 3 and $\beta^\dagger(\alpha) =$

1. Furthermore by concavity of $u^{\mathrm{D}}((\alpha,1),1)$, we have $u^{\mathrm{D}}((\alpha_1,1),1) > u^{\mathrm{D}}((\alpha_2,1),1)$.

(b) $\alpha_1 < \alpha_0^c$, $\widehat{\alpha}_c > \alpha_2 > \alpha_0^c$: In this case, we know from above that on one hand, α_1 lies in regime 3 and $\beta^\dagger(\alpha_1) = 1$. On the other hand, α_2 lies in regime 4 and $\beta^\dagger(\alpha_2) = \rho_c/\rho(\alpha)$. Furthermore, we know from Lemma 4, $(\alpha_2, \beta^\dagger)$ is dominated, or equivalently, gives lesser payoff to the distribution utility than $(\alpha_0^c, 1)$. And $(\alpha_0^c, 1)$ lies in regime 3 with $\gamma^\dagger(\alpha_0^c) = 1$. But we know that for $\pi < \widetilde{\pi}_c$, $u^{\mathrm{D}}((\alpha,1),1)$ is a concave function with maximum at α^*. Consequently, we infer that $u^{\mathrm{D}}((\alpha^*,1),1) > u^{\mathrm{D}}((\alpha_1,1),1) > u^{\mathrm{D}}((\alpha_2,\beta^\dagger),\gamma^\dagger)$.

(c) $\alpha_1 < \alpha_0^c$, $\alpha_2 > \widehat{\alpha}_c$: Again in this case, α_1 lies in regime 3 and $\beta^\dagger(\alpha_1) = 1$ and $\gamma^\dagger(\alpha_1) = 1$. However α_2 lies in regime 2 and $\beta^\dagger(\alpha_2) = 0$ with $\gamma^\dagger(\alpha_2) = 1$. From discussion 2, we have $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_1 + \pi((-\mathrm{C}_\rho + \mathrm{F})\rho(\alpha_1) - (pq\eta)(1-\rho(\alpha_1)))$ and $u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger) = -\pi pq\eta$. However, from (57), we can see that $-(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_1 + \pi((-\mathrm{C}_\rho + \mathrm{F} + pq\eta)\rho(\alpha_1)) > 0$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) > u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

(d) $\widehat{\alpha}_c > \alpha_1 > \alpha_0^c$, $\alpha_2 < \widehat{\alpha}_c$: In this case, both α_1 an α_2 lie in regime 4 and $\beta^\dagger(\alpha) = \rho_c/\rho(\alpha)$. Therefore, we know $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -\pi_c(\alpha)qp\eta$. Lastly, we know that from Lemma 5 that $\rho(\alpha_1)/\alpha_1 > \rho(\alpha_2)/\alpha_2$ and from (12), $\pi_c(\alpha_1) < \pi_c(\alpha_2)$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) > u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

(e) $\widehat{\alpha}_c > \alpha_1 > \alpha_0^c$, $\alpha_2 > \widehat{\alpha}_c$: In this case, α_1 lies in regime 4 and $\beta^\dagger(\alpha) = \rho_c/\rho(\alpha)$. Furthermore, α_2 lies in regime 2 with $\beta^\dagger(\alpha_2) = 0$. The payoffs are given by, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) = -\pi_c(\alpha_1)qp\eta$ and $u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger) = -\pi qp\eta$. Lastly, from (57), $\pi_c(\alpha_1) < \pi$. Hence, $u^{\mathrm{D}}(\alpha_1, 0, 1) > u^{\mathrm{D}}(\alpha_2, 0, 1)$

(f) $\alpha_1 > \widehat{\alpha}_c$, $\alpha_2 > \widehat{\alpha}_c$: In this case, the payoff $u^{\mathrm{D}} = -\pi pq\eta$ for both α_1, α_2. Hence, $u^{\mathrm{D}}(\alpha_1, 0, 1) = u^{\mathrm{D}}(\alpha_2, 0, 1)$

Case (ii) $\pi < \widehat{\pi}_c < \widetilde{\pi}_c$

The proof will be same as that for Case (i)

Case (iii) $\widehat{\pi}_c < \pi < \widetilde{\pi}_c$

In this case, as per our argument earlier, there does not exist $\widehat{\alpha}_c \in (0,1)$. Let us consider all the cases:

i. $\alpha_1 < \alpha_2 < \alpha^*$

   From Lemma 5, since $\alpha_1 < \alpha_2 < \alpha^*$, we have $\rho(\alpha_1)/\alpha_1 > \rho(\alpha_2)/\alpha_2 > \rho(\alpha^*)/\alpha^* > \rho_\alpha(\alpha^*)$. Recall from Proposition 1, this implies that α_1, α_2 lie in regime 3, or equivalently, $\beta^\dagger(\alpha_1) = 1$ and $\beta^\dagger(\alpha_2) = 1$. But for $\beta = 1$, since u^{D} is a concave function, we know that $u^{\mathrm{D}}((\alpha_1, 1), 1) < u^{\mathrm{D}}((\alpha_2, 1), 1)$.

ii. $\alpha_1 < \alpha_0^c$, $\alpha_2 < \alpha_0^c$: In this case, we know that for both α, $\rho(\alpha) < \rho_c$ and $\frac{\rho(\alpha)}{\alpha} > \frac{1-\pi}{\pi}\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{\mathrm{T}q\eta + \mathrm{F} - \mathrm{C}_\rho}$. Consequently, both α lie in regime 3 and $\beta^\dagger(\alpha) = 1$. Furthermore by concavity of $u^{\mathrm{D}}((\alpha, 1), 1)$, we have $u^{\mathrm{D}}((\alpha_1, 1), 1) > u^{\mathrm{D}}((\alpha_2, 1), 1)$.

iii. $\alpha_1 < \alpha_0^c$, $\alpha_2 > \alpha_0^c$: In this case, we know from above that on one hand, α_1 lies in regime 3 and $\beta^\dagger(\alpha_1) = 1$. On the other hand, α_2 lies in regime 4 and $\beta^\dagger(\alpha_2) = \rho_c/\rho(\alpha)$. Furthermore, we know from Lemma 4, $(\alpha_2, \beta^\dagger)$ is dominated, or equivalently, gives lesser payoff to the distribution utility than $(\alpha_0^c, 1)$. And $(\alpha_0^c, 1)$ lies in regime 3 with $\gamma^\dagger(\alpha_0^c) = 1$. But we know that for $\pi < \widetilde{\pi}_c$, $u^{\mathrm{D}}((\alpha, 1), 1)$ is a concave function with maximum at α^*. Consequently, we infer that $u^{\mathrm{D}}((\alpha^*, 1), 1) > u^{\mathrm{D}}((\alpha_1, 1), 1) > u^{\mathrm{D}}((\alpha_2, \beta^\dagger), \gamma^\dagger)$.

iv. $\alpha_1 > \alpha_0^c, \alpha_2 > \alpha_0^c$: In this case, both α_1 an α_2 lie in regime 4 and $\beta^\dagger(\alpha) = \rho_c/\rho(\alpha)$. Therefore, we know $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -\pi_c(\alpha)qp\eta$. Lastly, we know that from Lemma 5 that $\rho(\alpha_1)/\alpha_1 > \rho(\alpha_2)/\alpha_2$ and from (12), $\pi_c(\alpha_1) < \pi_c(\alpha_2)$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) > u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

□

Proposition 6. *For all $\pi > \widetilde{\pi}_c$ and for any α^1 and α^2, $\alpha^1 < \alpha^2 < \alpha^*$ the distribution utility obtains a greater payoff for α^2 than α^1 i.e., $u^{\mathrm{D}}((\alpha^1, \beta^1), \gamma^1) \leqslant u^{\mathrm{D}}((\alpha^2, \beta^2), \gamma^2)$. Similarly, for $\alpha^2 > \alpha^1 > \alpha^*$ the distribution utility obtains a greater payoff for α^1 than α^2 i.e., $u^{\mathrm{D}}((\alpha^1, \beta^1), \gamma^1) \geqslant u^{\mathrm{D}}((\alpha^2, \beta^2), \gamma^2)$.*

Proof. We will consider the following cases for $\pi, \widetilde{\pi}_c, \widehat{\pi}_c$.

Case (i) $\pi > \widetilde{\pi}_c > \widehat{\pi}_c$

From Theorem 2, we know that for $\pi > \widetilde{\pi}_c$, $\alpha_0^* = \alpha_0^c$. Furthermore since $\pi > \widehat{\pi}_c$, we know that there does not exist a $\widehat{\alpha}_c$ as defined earlier. Consider the following cases:

i. $\alpha_1 < \alpha_2 < \alpha_0^c$: In this case, we know that both α_1 and α_2 lie in regime 3, i.e. $\beta^\dagger(\alpha) = 1$ and $\gamma^\dagger(\alpha) = 1$. Furthermore the payoff is given by $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha + \pi((-\mathrm{C}_\rho + \mathrm{F})\rho(\alpha) - (pq\eta)(1-\rho(\alpha)))$ and $\partial_\alpha u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \pi((-\mathrm{C}_\rho + \mathrm{F} + pq\eta)\rho'(\alpha)) > 0$ from $\pi > \widetilde{\pi}_c$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) < u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

ii. $\alpha_0^c < \alpha_1 < \alpha_2$: In this case, we know that both α_1 and α_2 lie in regime 4, i.e. $\beta^\dagger(\alpha) = \rho_c/\rho(\alpha)$ and $\gamma^\dagger(\alpha) = \pi_c(\alpha)/\pi$. Furthermore $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -\pi_c(\alpha)pq\eta$. Lastly, for $\alpha_1 < \alpha_2$, $\pi_c(\alpha_1) < \pi_c(\alpha_2)$ and hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) > u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$.

Case (ii) $\pi > \widehat{\pi}_c > \widetilde{\pi}_c$

The proof for this case follows from **Case (i)** as both situations are equivalent for the distribution utility.

Case (iii) $\widehat{\pi}_c > \pi > \widetilde{\pi}_c$

In this case, we know that there exist $\widehat{\alpha}_c$ given by (57). Like we have seen in Proposition 5, we will define regimes for α_0^c and $\widehat{\alpha}_c$. Consider the following cases:

i. $\alpha_0^c < \widehat{\alpha}_c$: Consider α_1, α_2, such that $\alpha_1 < \alpha_2$

(a) $\alpha_1 < \alpha_0^c$ and $\alpha_2 < \alpha_0^c$: In this case, we know that both α_1 and α_2 lie in regime 3, i.e. $\beta^\dagger(\alpha) = 1$ and $\gamma^\dagger(\alpha) = 1$. Furthermore the payoff is given by $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha + \pi((-\mathrm{C}_\rho + \mathrm{F})\rho(\alpha) - (pq\eta)(1-\rho(\alpha)))$ and $\partial_\alpha u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \pi((-\mathrm{C}_\rho + \mathrm{F} + pq\eta)\rho'(\alpha)) > 0$ from $\pi > \widetilde{\pi}_c$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) < u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

(b) $\alpha_0^c < \alpha_1 < \widehat{\alpha}_c$ and $\alpha_0^c < \alpha_2 < \widehat{\alpha}_c$: In this case, we know that both α_1 and α_2 lie in regime 4, i.e. $\beta^\dagger(\alpha) = \rho_c/\rho(\alpha)$ and $\gamma^\dagger(\alpha) = \pi_c(\alpha)/\pi$. Furthermore

$u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -\pi_c(\alpha)pq\eta$. Lastly, for $\alpha_1 < \alpha_2$, $\pi_c(\alpha_1) < \pi_c(\alpha_2)$ and hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) > u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$.

(c) $\alpha_0^c < \alpha_1 < \widehat{\alpha}_c$ and $\alpha_2 > \widehat{\alpha}_c$: In this case, we know that α_1 lies in regime 4, i.e. $\beta^\dagger(\alpha_1) = \rho_c/\rho(\alpha_1)$ and $\gamma^\dagger(\alpha_1) = \pi_c(\alpha)/\pi$. Also $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) = -\pi_c(\alpha_1)pq\eta$. Furthermore, we know that α_2 lies in regime 2 i.e. $\beta^\dagger(\alpha_2) = 0$ and $\gamma^\dagger(\alpha_2) = 1$. Also $u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger) = -\pi pq\eta$. Since $\pi > \pi_c(\alpha_1)$, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) > u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

(d) $\alpha_1 > \widehat{\alpha}_c$ and $\alpha_2 > \widehat{\alpha}_c$: In this case, both α_1 and α_2 lie in regime 2 i.e. $\beta^\dagger(\alpha) = 0$ and $\gamma^\dagger(\alpha) = 1$. Also $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -\pi pq\eta$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) = u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

ii. $\widehat{\alpha}_c < \alpha_0^c$: Consider α_1, α_2 such that $\alpha_1 < \alpha_2$

(a) $\alpha_1 < \widehat{\alpha}_c$ and $\alpha_2 < \widehat{\alpha}_c$: In this case, we know that both α_1 and α_2 lie in regime 3, i.e. $\beta^\dagger(\alpha) = 1$ and $\gamma^\dagger(\alpha) = 1$. Furthermore the payoff is given by $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha + \pi((-\mathrm{C}_\rho + \mathrm{F})\rho(\alpha) - (pq\eta)(1-\rho(\alpha)))$ and $\partial_\alpha u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \pi((-\mathrm{C}_\rho + \mathrm{F} + pq\eta)\rho'(\alpha)) > 0$ from $\pi > \widetilde{\pi}_c$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) < u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

(b) $\alpha_1 < \widehat{\alpha}_c$ and $\widehat{\alpha}_c < \alpha_2 < \alpha_0^c$: In this case, we know that α_1 lies in regime 3, i.e. $\beta^\dagger(\alpha_1) = 1$ and $\gamma^\dagger(\alpha_1) = 1$. Also the payoff is given by $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) = -(1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_1 + \pi((-\mathrm{C}_\rho + \mathrm{F})\rho(\alpha_1) - (pq\eta)(1-\rho(\alpha_1)))$. Furthermore, α_2 lies in regime 2, i.e $\beta^\dagger(\alpha_2) = 0$ and $\gamma^\dagger(\alpha_2) = 1$. Also $u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger) = -\pi pq\eta$. Lastly, since $\alpha_1 < \widehat{\alpha}_c$, $\pi(\mathrm{F} - \mathrm{C}_\rho + pq\eta)\rho(\alpha_1) - (1-\pi)(\mathrm{C}_\alpha + \mathrm{C}_\rho)\alpha_1 > 0$ and hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) < u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$.

(c) $\widehat{\alpha}_c < \alpha_1 < \alpha_0^c$ and $\widehat{\alpha}_c < \alpha_2 < \alpha_0^c$: In this case, both α_1 and α_2 lie in regime 2 i.e. $\beta^\dagger(\alpha) = 0$ and $\gamma^\dagger(\alpha) = 1$. Also $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -\pi pq\eta$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) = u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

(d) $\alpha_1 > \alpha_0^c$ and $\alpha_2 > \alpha_0^c$: In this case, both α_1 and α_2 lie in regime 2 i.e. $\beta^\dagger(\alpha) = 0$ and $\gamma^\dagger(\alpha) = 1$. Also $u^{\mathrm{D}}(\alpha, \beta^\dagger, \gamma^\dagger) = -\pi pq\eta$. Hence, $u^{\mathrm{D}}(\alpha_1, \beta^\dagger, \gamma^\dagger) = u^{\mathrm{D}}(\alpha_2, \beta^\dagger, \gamma^\dagger)$

□

Let us define two quantities of interest for game $\mathcal{G}^{\mathrm{ROC}}_{\mathrm{P}(\eta)}$ with default IDS,

$$\begin{aligned}{\alpha_0}^{\mathrm{H}} &:= \arg\max_{\alpha_0} \ u^{\mathrm{D}}_{\mathrm{H}}(\left(\alpha_0, \beta^{\mathrm{H}}\right), \gamma^{\mathrm{H}}) \\ {\alpha_0}^{\mathrm{L}} &:= \arg\max_{\alpha_0} \ u^{\mathrm{D}}_{\mathrm{L}}(\left(\alpha_0, \beta^{\mathrm{L}}\right), \gamma^{\mathrm{L}}) \end{aligned} \tag{58}$$

Theorem 3. ***Optimal default false alarm probability for two level of thefts*** *For given two levels of theft $\eta^{\mathrm{H}}, \eta^{\mathrm{L}}$ with probability $\lambda^{\mathrm{H}}, \lambda^{\mathrm{L}}(= 1 - \lambda^{\mathrm{H}})$ respectively and a given tariff scheme $\mathrm{T}(q) = A + pq$. Furthermore, let ${\alpha_0}^{\mathrm{H}}$ and ${\alpha_0}^{\mathrm{L}}$ be the optimal default false alarm probability for low η^{L} and high η^{H} theft level respectively. Then, $\alpha_0^* \in [{\alpha_0}^{\mathrm{L}}, {\alpha_0}^{\mathrm{H}}]$ if ${\alpha_0}^{\mathrm{L}} < {\alpha_0}^{\mathrm{H}}$ and $\alpha_0^* \in [{\alpha_0}^{\mathrm{H}}, {\alpha_0}^{\mathrm{L}}]$ if ${\alpha_0}^{\mathrm{H}} < {\alpha_0}^{\mathrm{L}}$ for all π.*

Proof. For simplicity, let ${\alpha_0}^{\mathrm{L}} < {\alpha_0}^{\mathrm{H}}$, then we show that $\alpha_0^* \in [{\alpha_0}^{\mathrm{L}}, {\alpha_0}^{\mathrm{H}}]$. Assume the contrary and consider the following cases:

i. $\alpha_0^* < {\alpha_0}^{\mathrm{L}}$: From Proposition 5 and 6, we know that $u^{\mathrm{D}}_{\mathrm{H}}((\alpha_0^*, \beta^{\mathrm{H}}(\alpha_0^*)), \gamma^{\mathrm{H}}(\alpha_0^*)) \leqslant u^{\mathrm{D}}_{\mathrm{H}}({\alpha_0}^{\mathrm{L}}, \beta^{\mathrm{H}}({\alpha_0}^{\mathrm{L}}), \gamma^{\mathrm{H}}({\alpha_0}^{\mathrm{L}}))$. Similarly, from Theorem 2, equivalently (58), $u^{\mathrm{D}}_{\mathrm{L}}((\alpha_0^*, \beta^{\mathrm{L}}(\alpha_0^*)), \gamma^{\mathrm{L}}(\alpha_0^*)) \leqslant u^{\mathrm{D}}_{\mathrm{L}}(({\alpha_0}^{\mathrm{L}}, \beta^{\mathrm{L}}({\alpha_0}^{\mathrm{L}})), \gamma^{\mathrm{L}}({\alpha_0}^{\mathrm{L}}))$. Taking the weighted average of inequalities above,

$$\begin{aligned}&\lambda^{\mathrm{L}} u^{\mathrm{D}}_{\mathrm{L}}(\left(\alpha_0^*, \beta^{\mathrm{L}}(\alpha_0^*)\right), \gamma^{\mathrm{L}}(\alpha_0^*)) + \lambda^{\mathrm{H}} u^{\mathrm{D}}_{\mathrm{H}}\left(\left(\alpha_0^*, \beta^{\mathrm{H}}(\alpha_0^*)\right), \gamma^{\mathrm{H}}(\alpha_0^*)\right) \\ &\leqslant \lambda^{\mathrm{L}} u^{\mathrm{D}}_{\mathrm{L}}(\left({\alpha_0}^{\mathrm{L}}, \beta^{\mathrm{L}}({\alpha_0}^{\mathrm{L}})\right), \gamma^{\mathrm{L}}({\alpha_0}^{\mathrm{L}})) + \lambda^{\mathrm{H}} u^{\mathrm{D}}_{\mathrm{H}}(\left({\alpha_0}^{\mathrm{L}}, \beta^{\mathrm{H}}({\alpha_0}^{\mathrm{L}})\right), \gamma^{\mathrm{H}}({\alpha_0}^{\mathrm{L}}))\end{aligned}$$

which is a contradiction, from (55).

ii. $\alpha_0^* > {\alpha_0}^{\mathrm{H}}$: From Proposition 5 and 6, we know that $u^{\mathrm{D}}_{\mathrm{L}}((\alpha_0^*, \beta^{\mathrm{L}}(\alpha_0^*)), \gamma^{\mathrm{L}}(\alpha_0^*)) \leqslant u^{\mathrm{D}}_{\mathrm{L}}({\alpha_0}^{\mathrm{H}}, \beta^{\mathrm{L}}({\alpha_0}^{\mathrm{H}}), \gamma^{\mathrm{L}}({\alpha_0}^{\mathrm{H}}))$. Similarly, from Theorem 2, equivalently (58), $u^{\mathrm{D}}_{\mathrm{H}}((\alpha_0^*, \beta^{\mathrm{H}}(\alpha_0^*)), \gamma^{\mathrm{H}}(\alpha_0^*)) \leqslant u^{\mathrm{D}}_{\mathrm{H}}(({\alpha_0}^{\mathrm{H}}, \beta^{\mathrm{H}}({\alpha_0}^{\mathrm{H}})), \gamma^{\mathrm{H}}({\alpha_0}^{\mathrm{H}}))$. Taking the weighted average of inequalities above,

$$\begin{aligned}&\lambda^{\mathrm{L}} u^{\mathrm{D}}_{\mathrm{L}}(\left(\alpha_0^*, \beta^{\mathrm{L}}(\alpha_0^*)\right), \gamma^{\mathrm{L}}(\alpha_0^*)) + \lambda^{\mathrm{H}} u^{\mathrm{D}}_{\mathrm{H}}\left(\left(\alpha_0^*, \beta^{\mathrm{H}}(\alpha_0^*)\right), \gamma^{\mathrm{H}}(\alpha_0^*)\right) \\ &\leqslant \lambda^{\mathrm{L}} u^{\mathrm{D}}_{\mathrm{L}}(\left({\alpha_0}^{\mathrm{H}}, \beta^{\mathrm{L}}({\alpha_0}^{\mathrm{H}})\right), \gamma^{\mathrm{L}}({\alpha_0}^{\mathrm{H}})) + \lambda^{\mathrm{H}} u^{\mathrm{D}}_{\mathrm{H}}(\left({\alpha_0}^{\mathrm{H}}, \beta^{\mathrm{H}}({\alpha_0}^{\mathrm{H}})\right), \gamma^{\mathrm{H}}({\alpha_0}^{\mathrm{H}}))\end{aligned}$$

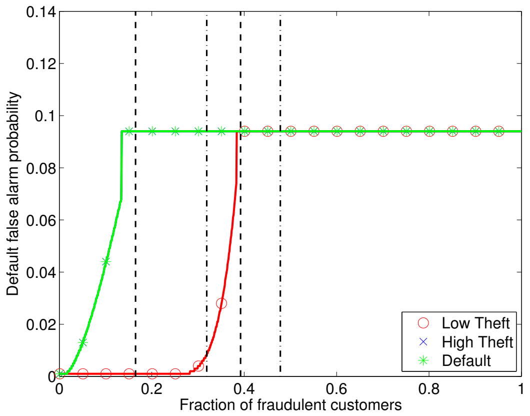


Figure 3.12: Game $\mathcal{G}^{\mathrm{ROC}}_{\mathrm{P}(\eta)}$ Equilibrium Response of Distribution Utility for $\eta^{\mathrm{L}} = 0.15, \eta^{\mathrm{H}} = 0.72$

which is a contradiction, from (55).

We can similarly argue for ${\alpha_0}^{\mathrm{L}} > {\alpha_0}^{\mathrm{H}}$ case. □

Discussion

In the plots below, we show the defender's equilibrium response α_0^* for two levels of theft with $\lambda^{\mathrm{L}} = \lambda^{\mathrm{H}} = 0.5$. The red curve corresponds to low theft η^{L}, the blue curve corresponds to η^{H} and the green curve corresponds to the equilibrium response. It can be verified that the green curve always lies inside the blue and red curve as given in Theorem 3.

i. *The critical false alarm probability is same for both level of thefts i.e.* $\eta^{\mathrm{L}} = 0.15, \eta^{\mathrm{H}} = 0.724$*:*

This case presents the IDS with same optimal default configuration for both theft levels. It is evident that both the low and high theft level have the same critical false alarm probability after a certain threshold for fraction of fraudulent customers. Furthermore, we observe in Figure 3.12 that the distribution utility, below the critical fraction, chooses a false alarm probability that is optimally configured for high theft level fraudulent customers η^{H}. In the later section 4.3.1, we will further discuss the underlying conditions for same critical false alarm probability and understand its relationship to value of information.

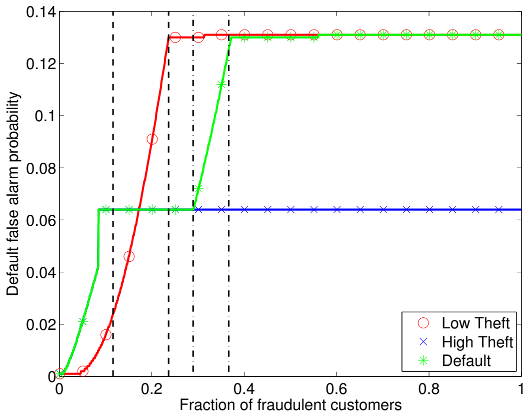


Figure 3.13: Game $\mathcal{G}^{\mathrm{ROC}}_{\mathrm{P}(\eta)}$ Equilibrium Response of Distribution Utility for $\eta^{\mathrm{L}} = 0.5, \eta^{\mathrm{H}} = 0.9$

ii. *The critical false alarm probability for low theft is more than critical false alarm probability for hight theft i.e.* $\eta^{\mathrm{L}} = 0.5, \eta^{\mathrm{H}} = 0.9$*:*

This case presents the non-monotonic behavior of α_0^c with level of theft η. In Figure 3.13, it is evident that the critical false alarm probability α_0^c is higher for low theft level fraudulent customer. Furthermore, we observe that for small fraction of fraudulent customers, the distribution utility "addresses" the high theft level customers while for large fraction of fraudulent customers, the distribution utility "addresses" the low theft level customers.

iii. *The critical false alarm probability for low theft is less than critical false alarm probability for high theft i.e.* $\eta^{\mathrm{L}} = 0.2, \eta^{\mathrm{H}} = 0.35$*:*

For the given mixing probabilities (or equivalently lack of information on level of theft $\lambda^{\mathrm{L}} = \lambda^{\mathrm{H}} = 0.5$), we observe in Figure 3.14 that the distribution utility always addresses the high theft fraudulent customers. One of the interpretations is that, for the same false alarm probability, the increased theft level is not sufficient for the distribution utility to detect the fraudulent customers. As a result, the distribution utility increases its false alarm probability to address the high theft committed by half of the total population of fraudulent customers.

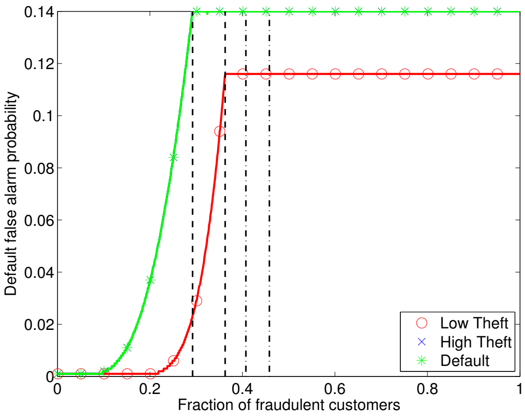


Figure 3.14: Game $\mathcal{G}^{\mathrm{ROC}}_{\mathrm{P}(\eta)}$ Equilibrium Response of Distribution Utility for $\eta^{\mathrm{L}} = 0.2, \eta^{\mathrm{H}} = 0.35$

Chapter 4

Value of Information

In this chapter, we utilize the results on equilibrium payoffs of the game $\mathcal{G}_{\eta}^{(\alpha_0,\rho_0)}$, $\mathcal{G}_{\eta}^{\mathrm{ROC}}$, and $\mathcal{G}_{\mathrm{P}(\eta)}^{\mathrm{ROC}}$ to determine the value of IDS(fixed, tunable/customizable), and the value of information on the theft level. We choose the cost incurred by the distribution utility under no IDS as the base cost to compute the value of IDS. This corresponds to the case when the fraudulent customer is able to divert energy at the level η without facing any investigation from the distribution utility. We use the following notation:

$\mathcal{V}_{\eta}^{(\alpha_0,\rho_0)}$	Value of default IDS with parameters (α_0, ρ_0) and η(avg. theft level)
$\mathcal{V}_{\eta}^{\mathrm{ROC}}$	Value of tunable IDS with ROC curve $(\alpha, \rho(\alpha))$ and η(actual theft level)
$\mathcal{V}_{\eta}^{\mathrm{cust}}$	Value of customizable IDS rel. for a fixed configuration and known η
$\mathcal{V}_{\mathrm{info}}^{\mathrm{ROC}}$	Value of information on η with tunable IDS

Table 4.1: Metrics of Value of Information

4.1 Value of Intrusion Detection Systems

Value of default IDS $\mathcal{V}(\alpha_0, \rho_0)$: For a given fraction of customers π and average level of theft η, the value of fixed IDS with parameters (α_0, ρ_0), denoted by $\mathcal{V}_{\eta}^{(\alpha_0,\rho_0)}$

is defined as follows:

$$\mathcal{V}_\eta^{(\alpha_0,\rho_0)} = u^{\mathrm{D}}(\mathcal{G}_\eta^{(\alpha_0,\rho_0)}) - u^{\mathrm{D}}(\mathcal{G}^0)$$

From 3.1 and Proposition 1

$$\mathcal{V}_\eta^{(\alpha_0,\rho_0)} = \pi pq\eta + u^{\mathrm{D}}(\beta^\dagger, \gamma^\dagger) \tag{59}$$

4.1.1 Default IDS

The following result characterizes $\mathcal{V}_\eta^{(\alpha_0,\rho_0)}$:

Proposition 7. *For a given π and η, the equilibrium value of fixed IDS (α_0, ρ_0) is given by:*

$$\mathcal{V}_\eta^{(\alpha_0,\rho_0)} = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ 0 & \text{if } \eta > \eta_c^I, \pi \leqslant \pi_c \\ -(1-\pi)\kappa_1\alpha_0 + \pi\kappa_2\rho_0 & \text{if } \eta > \eta_c^I, \pi > \pi_c \text{ and } \rho_0 \leqslant \rho_c \\ (\pi - \pi_c)pq\eta & \text{if } \eta > \eta_c^I, \pi > \pi_c \text{ and } \rho_0 > \rho_c \end{cases} \tag{60}$$

where $\kappa_1 := \mathrm{C}_\alpha + \mathrm{C}_\rho$ and $\kappa_2 := -\mathrm{C}_\rho + \mathrm{F} + pq\eta$, ρ_c given by 13 and π_c given by 12. Furthermore for any $\pi \in [0,1]$ and $\eta \in [0,1]$, $\mathcal{V}_\eta^{(\alpha_0,\rho_0)}$ is non negative at equilibrium.

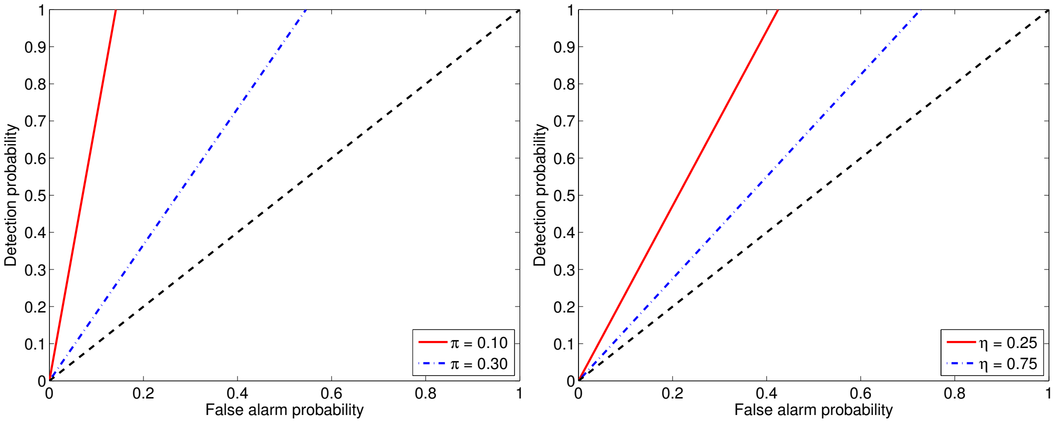
i. $\mathcal{V}_\eta^{(\alpha_0,\rho_0)} > 0$, if and only if the IDS configuration satisfies

$$\frac{\rho_0}{\alpha_0} > \frac{(1-\pi)}{\pi}\frac{\mathrm{C}_\alpha + \mathrm{C}_\rho}{\mathrm{F} + pq\eta - \mathrm{C}_\rho} \tag{61}$$

ii. For given η, the IDS is always valuable, i.e. $\mathcal{V}_\eta^{(\alpha_0,\rho_0)} > 0$ for any $\rho_0/\alpha_0 \geqslant 1$, if

$$\pi > \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{C}_\alpha + \mathrm{C}_\rho) + (pq\eta + \mathrm{F} - \mathrm{C}_\rho)} \tag{62}$$

Proof. By substituting u^{D} from (18) in (59), we obtain (60). Furthermore for $\eta \leqslant \eta_c^I$ and $\eta > \eta_c^I, \pi \leqslant \pi_c$, we obtain that value of IDS is zero. For $\eta > \eta_c^I, \pi > \pi_c$ and $\rho_0 \leqslant$



(a)

(b)

Figure 4.1: Variation of minimum IDS configuration ρ_0/α_0 with (a) Fraction of fraudulent customer π (b) Theft level η

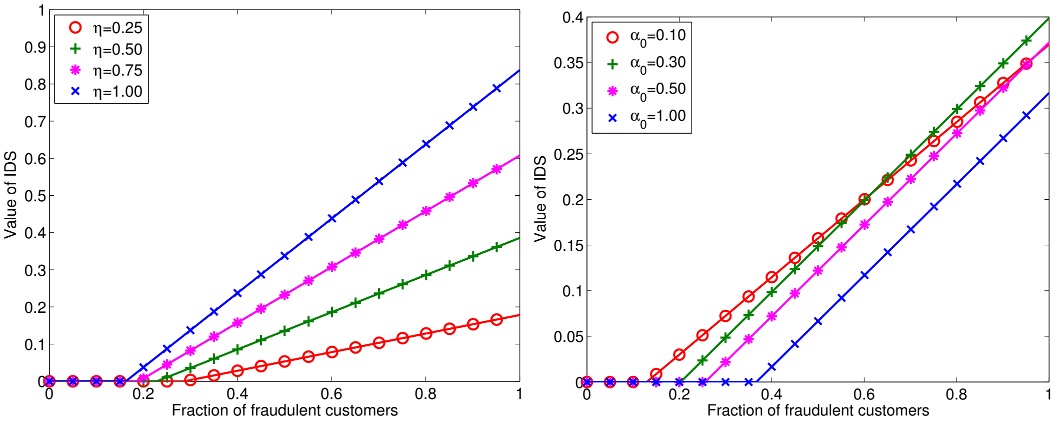
ρ_c, from definition of π_c (see (12)), we obtain $\pi\kappa_2\rho_0 > (1-\pi)\kappa_1\alpha_0$ if $\pi > \pi_c$ and consequently the value of IDS is strictly positive. Analogously, for $\eta > \eta_c^I, \pi > \pi_c$ and $\rho_0 > \rho_c$, since $\pi > \pi_c$ the value is strictly positive. Hence, $\mathcal{V}_\eta^{(\alpha_0,\rho_0)}$ is non negative at equilibrium.

i. From (60), we can observe that, in equilibrium, $\mathcal{V} > 0$ if and only if $\pi > \pi_c$ and $\eta > \eta_c^I$. Hence using $\pi > \pi_c$ from (12), we get (61)

ii. Any valid IDS will have $\rho_0 \geqslant \alpha_0$. By substituting the lower bound 1 on ρ_0/α_0 into (61) and rearranging, we obtain the conclusion in (62)

□

Discussion The above result (part (i)) gives a useful lower bound on the ratio ρ_0/α_0 for the IDS to have non-zero value to the distribution utility. Thus, the ratio ρ_0/α_0 can be viewed as an indicator of "quality" of the IDS. For a given average level of theft η, from (61), we obtain that with decreasing π, the distribution utility needs a better quality IDS(i.e. higher ρ_0/α_0) to derive a positive value form it. At a first glance, this result may seem counterintuitive. However, note that as π reduces, the false alarm cost increases and this requires a better IDS. Also, from (61), we obtain that with increasing theft level η, the distribution utility can derive a positive value



(a) (b)

Figure 4.2: Value of default IDS $\mathcal{V}_{\eta}^{(\alpha_0,\rho_0)}$ with fraction of fraudulent customer π for different (a) Theft level η (b) Default IDS configuration (α_0, ρ_0)

with lower ρ_0/α_0. Indeed as η increases, the distribution utility's payoff increased due to successful investigation. Consequently, a positive expected payoff even from a smaller ρ_0. These points are illustrated in Fig 4.1a and 4.1b respectively. For example, for $\pi = 0.1$ in Fig 4.1a, all IDS configurations above the corresponding lower bound provides a positive value to the distribution utility.

4.1.2 Tunable IDS

Next, the following result characterizes the value of tunable IDS ($\mathcal{V}_{\eta}^{\mathrm{ROC}}$) where the assumption is that the η is known to the distribution utility (i.e. Game $\mathcal{V}_{\eta}^{\mathrm{ROC}}$).

Proposition 8. *For a given ROC curve $\rho(\cdot)$, the equilibrium value of IDS under knowledge of η is given by,*

$$\mathcal{V}_{\eta}^{ROC} = \begin{cases} 0 & \text{if } \eta \leqslant \eta_c^I \\ -(1-\pi)\,\kappa_1\alpha^* + \pi\kappa_2\rho\,(\alpha^*) & \text{if } \eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c \\ -(1-\widetilde{\pi}_c)\,\kappa_1\alpha_0^c + \widetilde{\pi}_c\kappa_2\rho_c + pq\eta\,(\pi - \widetilde{\pi}_c) & \text{if } \eta > \eta_c^I, \pi > \widetilde{\pi}_c \end{cases} \tag{63}$$

where $\kappa_1 = (\mathrm{C}_\alpha + \mathrm{C}_\rho)$ and $\kappa_2 = (\mathrm{F} + pq\eta - \mathrm{C}_\rho)$ Furthermore, at equilibrium, the value of IDS is non-negative for all η, π, and is strictly positive for $\pi \in (0,1]$ and $\eta \in (\eta_c^I, 1]$

where η_c^I is given by 11. Lastly, the value of IDS $\mathcal{V}$ increases with increasing fraction of fraudulent customers π if and only if $\eta > \eta_c^I$.

Proof. By substituting u^{D} from (40) in (59), we obtain (63). We first show that $\mathcal{V}_\eta^{\mathrm{ROC}}$ is always non-negative and is strictly positive for $\eta > \eta_c^I$ and positive π

- For $\eta \leqslant \eta_c^I$, we obtain the value of IDS as zero.
- For $\eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c$, the value of IDS is greater than zero if $-(1-\pi)\,\kappa_1\alpha^* + \pi\kappa_2\rho\,(\alpha^*) > 0$, or equivalently,

  $$\pi\,(\mathrm{F} + pq\eta - \mathrm{C}_\rho)\,\rho\,(\alpha^*) > (1-\pi)\,(\mathrm{C}_\alpha + \mathrm{C}_\rho)\,\alpha^*$$
  $$\mathit{or}\ \frac{\rho\,(\alpha^*)}{\alpha^*} > \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{F} + pq\eta - \mathrm{C}_\rho)}$$

  From Proposition 2, the above inequality is equivalent to,

  $$\frac{\rho\,(\alpha^*)}{\alpha^*} > \frac{\partial\rho}{\partial\alpha}(\alpha^*)$$

  Finally from Lemma 2, $\rho(\alpha)/\alpha > \rho'(\alpha)$ for all α. This proves that value of IDS is positive for $\eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c$

- For $\eta > \eta_c^I, \pi > \widetilde{\pi}_c$, $\mathcal{V}_\eta^{\mathrm{ROC}} > 0$ if $\widetilde{\pi}_c\,(\mathrm{F} + pq\eta - \mathrm{C}_\rho)\,\rho_c > (1-\widetilde{\pi}_c)\,(\mathrm{C}_\alpha + \mathrm{C}_\rho)\,\alpha_0^c$. Recall the definition of $\widetilde{\pi}_c$ from (22),

  $$\widetilde{\pi}_c = \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \rho'(\alpha_0^c)(pq\eta + \mathrm{F} - \mathrm{C}_\rho)} > \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \frac{\rho_c}{\alpha_0^c}(pq\eta + \mathrm{F} - \mathrm{C}_\rho)} \tag{64}$$

  where the inequality follows from $\rho_c/\alpha_0^c > \rho'(\alpha_0^c)$ (using Lemma 2). Simplifying (64), we obtain $\widetilde{\pi}_c\,(\mathrm{F} + pq\eta - \mathrm{C}_\rho)\,\rho_c > (1-\widetilde{\pi}_c)\,(\mathrm{C}_\alpha + \mathrm{C}_\rho)\,\alpha_0^c$. Hence, $\mathcal{V}_\eta^{\mathrm{ROC}} > 0$ for $\eta > \eta_c^I, \pi > \widetilde{\pi}_c$

Next, we show that $\mathcal{V}_\eta^{\mathrm{ROC}}$ is increasing in fraction of fraudulent customer π.

- For $\eta \leqslant \eta_c^I$, the value of IDS is zero for all π

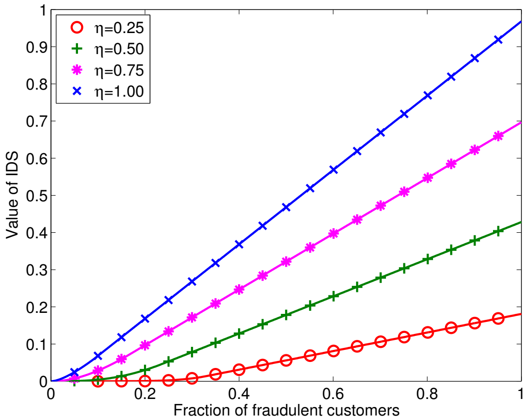


Figure 4.3: Value of tunable IDS $\mathcal{V}_{\eta}^{\mathrm{ROC}}$ with fraction of fraudulent customers π for different levels of theft η

- For $\eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c$, the value of tunable IDS $\mathcal{V}_{\eta}^{\mathrm{ROC}}$ can be rewritten as follows:

$$\begin{aligned}\mathcal{V}_{\eta}^{\mathrm{ROC}} &= \pi\left(\mathrm{F} + pq\eta - \mathrm{C}_{\rho}\right)\rho\left(\alpha^{*}\right)\left[1 - \frac{\left(1-\pi\right)\left(\mathrm{C}_{\alpha} + \mathrm{C}_{\rho}\right)\alpha^{*}}{\pi\left(\mathrm{F} + pq\eta - \mathrm{C}_{\rho}\right)\rho\left(\alpha^{*}\right)}\right] && (65)\\ &= \pi\left(\mathrm{F} + pq\eta - \mathrm{C}_{\rho}\right)\left[\rho\left(\alpha^{*}\right) - \rho'\left(\alpha^{*}\right)\alpha^{*}\right] && (66)\end{aligned}$$

  From Proposition 2, for $\eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c$, α^* increases with π. Let us define $\Gamma(x) = \rho(x) - \rho'(x)x$. Then, $\Gamma'(x) = -\rho''(x)$. Recall that $\rho(\alpha)$ is strictly concave in α. Hence, $\Gamma'(x) > 0$, or equivalently, $\Gamma(x)$ is increasing in x. To sum up, we obtain that $\left[\rho\left(\alpha^{*}\right) - \rho'\left(\alpha^{*}\right)\alpha^{*}\right]$ is increasing in π. Thus, $\mathcal{V}_{\eta}^{\mathrm{ROC}}$ is increasing in π for $\eta > \eta_c^I, \pi \leqslant \widetilde{\pi}_c$.

- For $\eta > \eta_c^I, \pi > \widetilde{\pi}_c$, the first two terms of $\mathcal{V}(\eta, \pi)$ are independent of π. Furthermore the third term increases linearly with π. Hence, $\mathcal{V}$ increases with π.

□

The above proposition can be visualized in Fig. 4.3. It is evident that value of tunable IDS $\mathcal{V}_{\eta}^{\mathrm{ROC}}$ is always positive and increasing in π for $\eta > \eta_c^I$. Moreover $\mathcal{V}_{\eta}^{\mathrm{ROC}}$ is increasing in theft level η.

4.2 Value of Customization

In this section, we briefly define value of customization of the IDS and describe its relevance for choosing the configuration of the IDS.

Value of customization $\mathcal{V}_{\eta}^{\text{cust}}$: The value of customization is defined as the difference in distribution utility's payoff between using the IDS of given configuration (α_0, ρ_0) (from Proposition 1) and using a tunable IDS with optimal false alarm probability for a given ROC curve α^*(from Proposition 2). For a given fraction of fraudulent customers π and level of theft η,

$$\mathcal{V}_{\eta}^{\text{cust}} := u^{\text{D}}(\alpha^*, \gamma^*) - u^{\text{D}}(\beta^{\dagger}, \gamma^{\dagger}) \tag{67}$$

By definition of α^*, we obtain $\mathcal{V}_{cust} \geqslant 0$. Equivalently,

$$\mathcal{V}_{\eta}^{\text{cust}} = \mathcal{V}_{\eta}^{\text{ROC}} - \mathcal{V}_{\eta}^{(\alpha_0, \rho_0)}$$

where $\mathcal{V}_{\eta}^{\text{ROC}}$ is given by Proposition 8 and $\mathcal{V}_{\eta}^{(\alpha_0, \rho_0)}$ is given by Proposition 7.

4.3 Value of Information

In this section, we discuss the value of information that the distribution utility obtains from gaining perfect knowledge of theft level committed by the fraudulent customer. As discussed earlier, the distribution utility in Game $\widehat{\mathcal{G}}_{\eta}^{\text{ROC}}$ finds the optimal configuration of the IDS, i.e. α_0^* in Stage 1 for single theft level while in Game $\mathcal{G}_{\text{P}(\eta)}^{\text{ROC}}$ the distribution utility finds α_0^* in Stage 1 with imperfect information of theft level, i.e. $\text{P}(\eta)$. By studying value of information, we aim to compare the payoff obtained by distribution utility in the two games. In subsequent sections, we first present the dependance of critical false alarm probability on theft level, then evaluate the value of information and its dependence on theft level and finally we visualize the obtained results and discuss their implications.

4.3.1 Critical False Alarm Probability and Theft Level

In this section, we will analyze the relationship between critical false alarm probability and level of theft committed by the fraudulent customer. Recall the definition of ρ_c from (13), the critical probability of detection that makes the fraudulent customer indifferent between theft **T** and no theft **NT**. Firstly, we argue in the following lemma the relationship between critical probability of detection ρ_c and level of theft η.

Lemma 6. *$\rho(\alpha_c, \eta)$ is strictly increasing in η*

Proof. Following (13), for a given $\eta, \mathrm{F}, \mathrm{C}_\eta$ we can write $\rho(\alpha_c, \eta)$

$$\begin{aligned} \rho(\alpha_c, \eta) &= \frac{pq\eta - \mathrm{C}_\eta}{pq\eta + \mathrm{F}}, \\ \partial_\eta \rho(\alpha_c(\eta), \eta) &= \frac{\mathrm{F} + \mathrm{C}_\eta}{(pq\eta + \mathrm{F})^2} > 0 \end{aligned} \tag{68}$$

□

Now for two levels of theft $\eta^{\mathrm{L}} < \eta^{\mathrm{H}}$, using (68) and (13), $\rho(\overline{\alpha_0}^{\mathrm{L}}, \eta^{\mathrm{L}}) < \rho(\overline{\alpha_0}^{\mathrm{H}}, \eta^{\mathrm{H}})$, where $\overline{\alpha_0}^{\mathrm{L}}$ and $\overline{\alpha_0}^{\mathrm{H}}$ are obtained by substituting η^{L} and η^{H} in (13) for ρ_c and,

$$\rho(\overline{\alpha_0}^{\mathrm{L}}, \eta^{\mathrm{L}}) = \frac{pq\eta^{\mathrm{L}} - \mathrm{C}_\eta}{pq\eta^{\mathrm{L}} + \mathrm{F}} \quad \text{and} \quad \rho(\overline{\alpha_0}^{\mathrm{H}}, \eta^{\mathrm{H}}) = \frac{pq\eta^{\mathrm{H}} - \mathrm{C}_\eta}{pq\eta^{\mathrm{H}} + \mathrm{F}} \tag{69}$$

Recall that ROC curve for normal distribution as a function of α and η,

$$\rho(\alpha, \eta) = \phi(\phi^{-1}(\alpha) + \frac{q\eta}{\sigma})$$

where ϕ is the CDF function for standard normal.

Let us define

$$\tau(\eta) := \phi^{-1}(\frac{pq\eta - \mathrm{C}_\eta}{pq\eta + \mathrm{F}}) - \frac{q\eta}{\sigma} = \phi^{-1}(\alpha_c) \tag{70}$$

Note that, since ϕ is strictly increasing, there is a one-to-one relationship between critical false alarm probability and $\tau(\eta)$.

The next lemma focuses on properties of $\tau(\eta)$.

Proposition 9. ***Properties of*** $\tau(\eta)$

Under

$$\frac{\sigma}{q} \leqslant \frac{(\mathrm{F}+pq)^2}{\sqrt{2\pi}pq(\mathrm{C}_\eta+\mathrm{F})\exp\left(\left(\text{erf}^{\text{-}1}(-2\frac{pq-\mathrm{C}_\eta}{pq+\mathrm{F}}+1)\right)^2\right)} \tag{71}$$

and

$$\mathrm{F}+2\mathrm{C}_\eta > pq \tag{72}$$

i. $\exists$ *a unique* $\eta_1 \in (\mathrm{C}_\eta/pq, 1)$ *such that,* $\tau(\eta_1) = \phi^{-1}(\frac{pq-\mathrm{C}_\eta}{pq+\mathrm{F}}) - \frac{q}{\sigma}$

ii. $\exists$ *a unique* $\eta_0 \in (\eta_1, 1)$, *such that,* $\eta_0 := \underset{\eta\in(\mathrm{C}_\eta/pq,1]}{\arg\max}\, \tau(\eta)$

Proof. We first show that $\tau(\eta)$ is concave given (72). Since $\tau(\eta)$ is a continuous, twice differentiable $\mathcal{C}^2$ in $(\mathrm{C}_\eta/pq, 1)$,

$$\frac{\partial^2\tau}{\partial^2\eta} = -2\sqrt{2\pi}(pq)^2\exp{(\zeta^2)}\frac{\mathrm{F}+\mathrm{C}_\eta}{(\mathrm{F}+pq\eta)^3} - \sqrt{2\pi}pq\exp{(2\zeta^2)}\zeta\frac{\mathrm{F}+\mathrm{C}_\eta}{(\mathrm{F}+pq\eta)^2}$$

where

$$\zeta = \text{erf}^{\text{-}1}\left(\frac{(2\mathrm{C}_\eta+\mathrm{F}-pq\eta)}{\mathrm{F}+pq\eta}\right)$$

Note that if $\zeta > 0$ then $\frac{\partial^2\tau}{\partial^2\eta} < 0$. Lastly, $\zeta > 0$ follows from (72).

Next, we show that $\tau'(1) < 0$. Note that,

$$\tau'(\eta) = \sqrt{2\pi}pq\frac{\mathrm{C}_\eta+\mathrm{F}}{(\mathrm{F}+pq\eta)^2}\exp\left(\left(\text{erf}^{\text{-}1}(-2\frac{pq\eta-\mathrm{C}_\eta}{pq\eta+\mathrm{F}}+1)\right)^2\right) - \frac{q}{\sigma} \tag{73}$$

From (73) and (71), $\tau'(\eta)|_{\eta=1} < 0$.

i. Note that $\lim_{\eta\to\mathrm{C}_\eta/pq}\tau(\eta) \to -\infty$ which is strictly less than finite $\tau(1)$. Since $\tau(\eta)$ is a continuous concave function with $\tau'(\eta)|_{\eta=1} < 0$, $\exists\eta$ such that $\tau(\eta) > \tau(1)$. Hence, using Intermediate Value Theorem(IVT), there exists a $\eta_1 \neq 1$ such that $\tau(\eta_1) = \tau(1)$. Lastly, the uniqueness follows from the concavity of $\tau(\eta)$.

ii. From 73, $\lim_{\eta \to \mathrm{C}_\eta/pq} \tau'(\eta) = \infty$ and $\tau'(\eta)|_{\eta=1} < 0$ as shown earlier. As a result, by IVT, there exists η such that $\tau'(\eta) = 0$ proving the existence of η_0. Lastly, the uniqueness of η_0 follows from concavity of $\tau(\eta)$.

□

Discussion

The above proposition for existence and uniqueness of η_0, η_1 hold for "sufficiently" high fine and "sufficiently" small standard deviation of electricity consumption. Note that the conditions on fine and standard deviation are sufficient but not necessary, i.e. we can have unique η_0, η_1 for "small" fine and "large" variance. Furthermore the above conditions are technical (but realistic assumptions). For example, the distribution utility does impose stricter laws (or "large" fine) against electricity theft and electricity customers, after accounting for external factors, do not exhibit large variance in consumption.

Proposition 10. *For given $\eta^{\mathrm{L}} < \eta^{\mathrm{H}}$, with $\tau(\eta)$ satisfying conditions (72),(71) and $\overline{\alpha_0}^{\mathrm{L}}, \overline{\alpha_0}^{\mathrm{H}}$ given by (69), Then,*

i. In the regime $\eta^{\mathrm{L}} \in (\mathrm{C}_\eta/pq, \eta_1)$, $\overline{\alpha_0}^{\mathrm{L}} < \overline{\alpha_0}^{\mathrm{H}}$.

ii. In the regime $\eta^{\mathrm{L}} \in [\eta_0, 1)$, $\overline{\alpha_0}^{\mathrm{L}} > \overline{\alpha_0}^{\mathrm{H}}$.

iii. In the regime $\eta^{\mathrm{L}} \in [\eta_1, \eta_0)$, we define $\eta_c > \eta^{\mathrm{L}}$ as,

$$\phi^{-1}(\frac{pq\eta_c - \mathrm{C}_\eta}{pq\eta_c + \mathrm{F}}) - \frac{q\eta_c}{\sigma} = \phi^{-1}(\frac{pq\eta^{\mathrm{L}} - \mathrm{C}_\eta}{pq\eta^{\mathrm{L}} + \mathrm{F}}) - \frac{q\eta^{\mathrm{L}}}{\sigma}$$

Then, $\eta_c > \eta^{\mathrm{L}}$. Furthermore,

(a) if $\eta_{\mathrm{H}} \in (\eta^{\mathrm{L}}, \eta_c)$, $\overline{\alpha_0}^{\mathrm{L}} < \overline{\alpha_0}^{\mathrm{H}}$.

(b) if $\eta_{\mathrm{H}} \in [\eta_c, 1)$, $\overline{\alpha_0}^{\mathrm{L}} \geqslant \overline{\alpha_0}^{\mathrm{H}}$.

where η_0, η_1 as defined in Proposition 9.

Proof. We know from (72),(71) in Proposition 10, $\tau(\eta)$ is concave and reaches maximum at η_0 such that $\tau(\eta_0) \geqslant \tau(\eta_1) = \tau(1)$

i. In the regime $\eta^{\mathrm{L}} \in (\mathrm{C}_\eta/pq, \eta_1]$, either $\eta^{\mathrm{H}} \leqslant \eta_1$ or $\eta^{\mathrm{H}} > \eta_1$. From the concavity of $\tau(\eta)$, definition of η_0 ie $\tau(\eta_0) > \tau(\eta_1) = \tau(1)$ and $\lim_{\eta \to \mathrm{C}_\eta/pq} \tau(\eta) = -\infty$, we know that $\tau(\eta)$ is increasing in $\eta < \eta_1$ and equivalently, $\overline{\alpha_0}^{\mathrm{L}} < \overline{\alpha_0}^{\mathrm{H}}$. Furthermore for $\eta > \eta_1$ since $\tau(\eta)$ is concave, $\nexists\ \eta \in (\eta_1, 1)$ such that $\tau(\eta) < \tau(\eta_1)$. Finally, since $\tau(\eta^{\mathrm{H}}) > \tau(\eta_1) > \tau(\eta^{\mathrm{L}})$, $\overline{\alpha_0}^{\mathrm{L}} < \overline{\alpha_0}^{\mathrm{H}}$.

ii. In the regime $\eta^{\mathrm{L}} \in [\eta_0, 1)$. Following from concavity of $\tau(\eta)$, it is decreasing in η for $\eta \in [\eta_0, 1)$. Hence, $\overline{\alpha_0}^{\mathrm{L}} > \overline{\alpha_0}^{\mathrm{H}}$.

iii. In the regime $\eta^{\mathrm{L}} \in [\eta_1, \eta_0)$, there exists a unique η_c such that $\tau(\eta_c) = \tau(\eta^{\mathrm{L}})$. This follows from concavity of $\tau(\eta)$ and that $\tau(1) < \tau(\eta^{\mathrm{L}})$. Furthermore, from definition of η_0, we know $\tau(\eta_c) < \tau(\eta_0)$. Hence, $\eta_c > \eta_0$.

   (a) $\eta^{\mathrm{H}} \in (\eta^{\mathrm{L}}, \eta_c)$: From our argument in (i), since $\tau(\eta)$ is concave, $\nexists\ \eta \in (\eta^{\mathrm{L}}, \eta_c)$ such that $\tau(\eta) < \tau(\eta^{\mathrm{L}})$. Hence, $\overline{\alpha_0}^{\mathrm{L}} < \overline{\alpha_0}^{\mathrm{H}}$

   (b) $\eta^{\mathrm{H}} \in [\eta_c, 1)$: We know that $\tau(1) = \tau(\eta_1) < \tau(\eta^{\mathrm{L}}) = \tau(\eta_c)$ as $\eta_1 < \eta^{\mathrm{L}}$. Furthermore since $\eta_0 < \eta_c$ and by concavity of $\tau(\eta)$, we know that $\tau(\eta)$ is decreasing in η or equivalently, $\tau(\eta^{\mathrm{H}}) < \tau(\eta^{\mathrm{L}})$ and $\overline{\alpha_0}^{\mathrm{L}} > \overline{\alpha_0}^{\mathrm{H}}$.

□

Discussion

We can visualize the proposition in the following example:

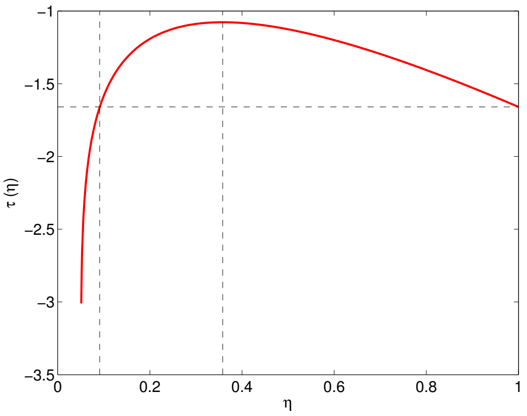


Figure 4.4: Variation of $\tau(\eta)$with η

The variation of critical false alarm probability α_0^c, or equivalently $\tau(\eta)$ with η is crucial to understand the properties of the IDS and ultimately argue for optimal configuration of the default IDS for heterogeneous fraudulent customers. As we have seen in Lemma 6, the critical probability of detection increases with increase in theft level η. Equivalently, it means that the distribution utility increases the probability of detection of the IDS as the fraudulent customer commits more theft η. This is intuitive because the critical probability of detection is the probability that makes the fraudulent customer indifferent between theft **T** and no theft **NT**. Furthermore, we know from the definition of ROC curve (more specifically for normal distribution) that, for a given level of false alarm probability, the probability of detection increases with increase in theft level η. We also know that the probability of detection increases with increase in false alarm probability. Consider two levels of theft η^{L} and η^{H} such that $\eta^{\mathrm{L}} < \eta^{\mathrm{H}}$. Also let $\overline{\alpha_0}^{\mathrm{L}}$, $\overline{\alpha_0}^{\mathrm{H}}$ be the critical false alarm probability for low level theft η^{L}, high level theft η^{H} respectively. Two cases can arise, as shown in Figure 4.5.

- As theft level increases from η^{L} to η^{H}, the probability of detection for $\overline{\alpha_0}^{\mathrm{L}}$ increases from $\rho(\overline{\alpha_0}^{\mathrm{L}}, \eta^{\mathrm{L}})$ to $\rho(\overline{\alpha_0}^{\mathrm{L}}, \eta^{\mathrm{H}})$. However the distribution utility needs an even higher detection probability to make the fraudulent customer indifferent between theft and no theft actions. Consequently, the distribution utility chooses a higher false alarm probability, i.e., $\overline{\alpha_0}^{\mathrm{L}} < \overline{\alpha_0}^{\mathrm{H}}$. This corresponds to case (i), case(ii)(a) in Proposition 10.

- As theft level increases from η^{L} to η^{H}, the probability of detection for $\overline{\alpha_0}^{\mathrm{L}}$ increases from $\rho(\overline{\alpha_0}^{\mathrm{L}}, \eta^{\mathrm{L}})$ to $\rho(\overline{\alpha_0}^{\mathrm{L}}, \eta^{\mathrm{H}})$. However the distribution utility needs a lower detection probability to make the fraudulent customer indifferent between theft and no theft actions. Consequently, the distribution utility chooses a lower false alarm probability, i.e., $\overline{\alpha_0}^{\mathrm{L}} > \overline{\alpha_0}^{\mathrm{H}}$. This corresponds to case(ii)(b), case(iii) in Proposition 10.

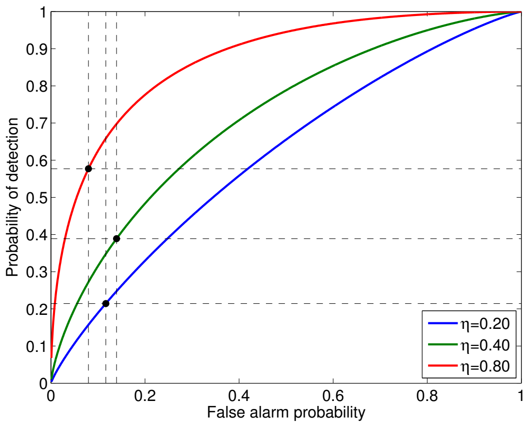


Figure 4.5: ROC and False Alarm Probability with theft level η

Finally, we justify studying the behavior of critical false alarm probability as the distribution utility chooses it as the equilibrium false alarm probability (from Theorem 2) for fraction of fraudulent customers π beyond $\widetilde{\pi}_c$(22).Consequently, it helps us argue for actions of the distribution utility independent of π.

4.3.2 Definition and Discussion

Intuitively, value of information is the amount a decision maker would be willing to pay for information prior to making a decision. Before we define the value of information, recall the definition of two games $\widehat{\mathcal{G}}_{\eta}^{\text{ROC}}$, $\mathcal{G}_{\text{P}(\eta)}^{\text{ROC}}$

i. **Perfect Information $\widehat{\mathcal{G}}_{\eta}^{\text{ROC}}$**: In Stage 1, the distribution utility sets a fixed configuration of the IDS α_0^* with their respective ROC curves $\rho(\alpha, \eta^{\text{L}})$ and $\rho(\alpha, \eta^{\text{H}})$. In Stage2, the distribution utility and fraudulent customer play a simultaneous game $(\beta^{\dagger}, \gamma^{\dagger})$ with complete knowledge of theft level η and prior π (Theorem 2).

ii. **Imperfect Information $\mathcal{G}_{\text{P}(\eta)}^{\text{ROC}}$**: In Stage 1, the distribution utility sets a fixed configuration of the IDS α_0^* using probability distribution P over theft level. In Stage2, the distribution utility and fraudulent customer play a simultaneous game $(\beta^{\dagger}, \gamma^{\dagger})$ with complete knowledge of theft level η and prior π (Theorem 3).

Like before, consider two theft levels $\eta^{\mathrm{L}} < \eta^{\mathrm{H}}$ with mixing probability $\lambda^{\mathrm{L}}, \lambda^{\mathrm{H}}(= 1 - \lambda^{\mathrm{L}}$ respectively. Furthermore, recall from (58), the definition of ${\alpha_0}^{\mathrm{L}}$ and ${\alpha_0}^{\mathrm{H}}$ such that ${\alpha_0}^{\mathrm{L}}$ is the optimal false alarm probability (Game $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$) for low theft level η^{L} and ${\alpha_0}^{\mathrm{H}}$ is the optimal false alarm probability (Game $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$) for high theft level. Furthermore, similar to Theorem 3, $u_{\mathrm{L}}^{\mathrm{D}}$ and $u_{\mathrm{H}}^{\mathrm{D}}$ is the distribution utility's payoff for low theft level and high theft level respectively.

Definition 1. Value of Information $\mathcal{V}_{\text{info}}^{\mathrm{ROC}}$:
The value of information is defined as the difference in distribution utility's payoff between using an optimally configured IDS under imperfect information $\mathcal{G}_{\mathrm{P}(\eta)}^{\mathrm{ROC}}$ and an optimally configured IDS under perfect information $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$.

$$\begin{aligned}\mathcal{V}_{info}(\pi, \mathrm{P}) :=& \lambda^{\mathrm{H}} \left(u_{\mathrm{H}}^{\mathrm{D}}(({\alpha_0}^{\mathrm{H}}, \beta^{\mathrm{H}}), \gamma^{\mathrm{H}}) - u_{\mathrm{H}}^{\mathrm{D}}((\alpha_0^*, \beta^{\mathrm{H}}), \gamma^{\mathrm{H}})\right) \\ &+ \lambda^{\mathrm{L}} \left(u_{\mathrm{L}}^{\mathrm{D}}(({\alpha_0}^{\mathrm{L}}, \beta^{\mathrm{L}}), \gamma^{\mathrm{L}}) - u_{\mathrm{L}}^{\mathrm{D}}((\alpha_0^*, \beta^{\mathrm{L}}), \gamma^{\mathrm{L}})\right) \qquad (74)\end{aligned}$$

where P is the bernoulli probability distribution over η ($\{\eta^{\mathrm{L}}, \eta^{\mathrm{H}}\}$ with $\{\lambda^{\mathrm{L}}, \lambda^{\mathrm{H}}\}$ probability), $(\beta^{\mathrm{H}}, \gamma^{\mathrm{H}})$ and $(\beta^{\mathrm{L}}, \gamma^{\mathrm{L}})$ is the subgame equilibrium given α_0 chosen in stage 1 for high theft level and low theft level respectively.(Refer Theorem 3)

Remark 8. The value of information is always non-negative, $\mathcal{V}_{info} \geqslant 0$. Equivalently, the distribution utility obtains a higher payoff from Game 3a $\widehat{\mathcal{G}}_{\eta}^{\mathrm{ROC}}$ as compared to Game 3b $\mathcal{G}_{\mathrm{P}(\eta)}^{\mathrm{ROC}}$ for all π.

Proof. Assume the contrary, i.e. there exists ${\alpha_0}^{\mathrm{L}}, {\alpha_0}^{\mathrm{H}}, \alpha_0^*$ such that $\mathcal{V}_{info} < 0$. Then we necessarily need, from (74), either $u_{\mathrm{H}}^{\mathrm{D}}({\alpha_0}^{\mathrm{H}}, \pi, \eta^{\mathrm{H}}) < u_{\mathrm{H}}^{\mathrm{D}}(\alpha_0^*, \pi, \eta^{\mathrm{H}})$ or $u_{\mathrm{L}}^{\mathrm{D}}({\alpha_0}^{\mathrm{L}}, \pi, \eta^{\mathrm{L}}) < u_{\mathrm{L}}^{\mathrm{D}}(\alpha_0^*, \pi, \eta^{\mathrm{L}})$. But from (58), we know that $u_{\mathrm{L}}^{\mathrm{D}}({\alpha_0}^{\mathrm{L}}, \pi, \eta^{\mathrm{L}}) \geqslant u_{\mathrm{L}}^{\mathrm{D}}(\alpha_0, \pi, \eta^{\mathrm{L}})$ and $u_{\mathrm{H}}^{\mathrm{D}}({\alpha_0}^{\mathrm{H}}, \pi, \eta^{\mathrm{L}}) \geqslant u_{\mathrm{H}}^{\mathrm{D}}(\alpha_0, \pi, \eta^{\mathrm{L}})$ for any α_0. A contradiction.

□

Proposition 11. ***Value of Information***
Given two levels of theft $\eta^{\mathrm{H}} > \eta^{\mathrm{L}}$ with probability $\lambda^{\mathrm{H}}, \lambda^{\mathrm{L}}(= 1 - \lambda^{\mathrm{H}})$ respectively and (72), (71) hold, then for $\eta^{\mathrm{L}} > \eta_1$, there exists η^{H} such that value of customization

$\mathcal{V}_{info} = 0$ *for* $\pi > \widetilde{\pi}_\Gamma$ *where*

$$\widetilde{\pi}_\Gamma = \max_{\eta \in \{\eta^{\mathrm{L}}, \eta^{\mathrm{H}}\}} \frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \rho'(\alpha_0^c(\eta), \eta)(pq\eta + \mathrm{F} - \mathrm{C}_\rho)}$$

where $\eta_1 \neq 1$ *is, from Proposition 10,*

$$\phi^{-1}(\frac{pq\eta_1 - \mathrm{C}_\eta}{pq\eta_1 + \mathrm{F}}) - \frac{q\eta_1}{\sigma} = \phi^{-1}(\frac{pq - \mathrm{C}_\eta}{pq + \mathrm{F}}) - \frac{q}{\sigma}$$

Proof. Recall from Proposition 10, that $\tau(\eta)$ is non-monotonic and concave for $\eta \in [\eta_1, 1]$ with $\tau(\eta_1) = \tau(1)$. Hence, there exists $\eta^{\mathrm{L}}, \eta^{\mathrm{H}} \in (\eta_1, 1)^2$ such that $\tau(\eta^{\mathrm{L}}) = \tau(\eta^{\mathrm{H}})$. Furthermore since, $\alpha_0^c = \phi(\tau(\eta))$, both η^{L} and η^{H} theft level will have same critical false alarm probability α_0^c. From Theorem 2, the equilibrium response of the distribution utility for $\pi > \widetilde{\pi}_c$ in both low level theft η^{L} case and high level theft η^{H} case is α_0^c, i.e. ${\alpha_0}^{\mathrm{L}} = \alpha_0^c$ and ${\alpha_0}^{\mathrm{H}} = \alpha_0^c$. From proof of part(i), we know that $\alpha_0^* \in [{\alpha_0}^{\mathrm{L}}, {\alpha_0}^{\mathrm{H}}]$ or $\alpha_0^* = \alpha_0^c$. Equivalently, $u_{\mathrm{L}}^{\mathrm{D}}(\alpha_0^*, \pi, \eta^{\mathrm{L}}) = u_{\mathrm{L}}^{\mathrm{D}}({\alpha_0}^{\mathrm{L}}, \pi, \eta^{\mathrm{L}})$ and $u_{\mathrm{H}}^{\mathrm{D}}({\alpha_0}^{\mathrm{H}}, \pi, \eta^{\mathrm{H}}) = u_{\mathrm{H}}^{\mathrm{D}}(\alpha_0^*, \pi, \eta^{\mathrm{H}})$. Hence, $\mathcal{V}_{info} = 0$. Finally the fraction of fraudulent customers π should be greater than, from (22),

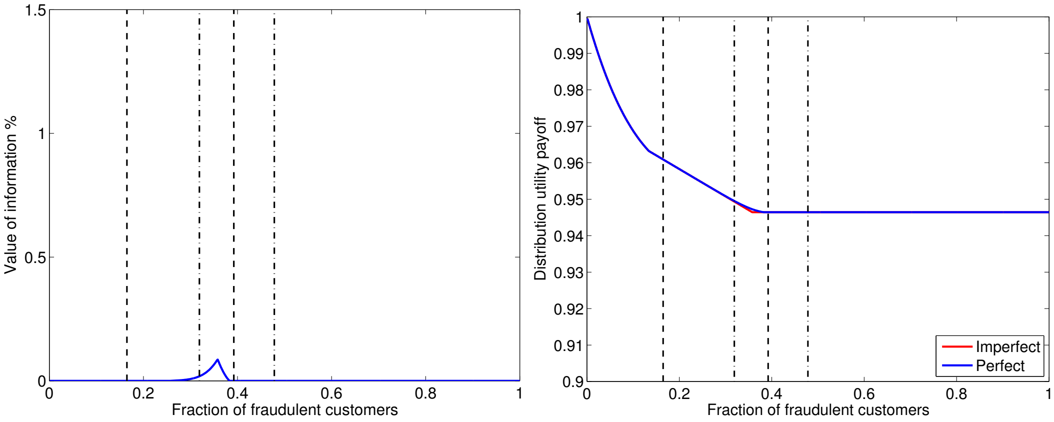
$$\begin{aligned}\widetilde{\pi}_\Gamma = \max\{&\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \rho'(\alpha_0^c, \eta^{\mathrm{L}})(pq\eta^{\mathrm{L}} + \mathrm{F} - \mathrm{C}_\rho)},\\ &\frac{(\mathrm{C}_\alpha + \mathrm{C}_\rho)}{(\mathrm{C}_\alpha + \mathrm{C}_\rho) + \rho'(\alpha_0^c, \eta^{\mathrm{H}})(pq\eta^{\mathrm{H}} + \mathrm{F} - \mathrm{C}_\rho)}\}\end{aligned}$$

□

Simulations

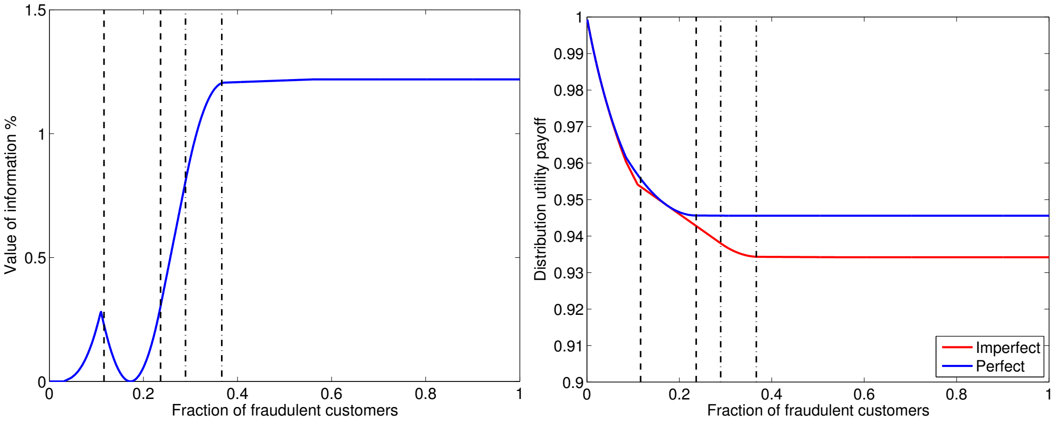
In the plots below, we show the % value of information and payoff of the distribution utility for both complete and incomplete information scenarios in part(a) and part(b) respectively.

i. When critical false alarm probability is same for both level of thefts i.e. $\eta^{\mathrm{L}} = 0.15, \eta^{\mathrm{H}} = 0.724$: We can visualize Proposition 11 for the above theft levels. Like we discussed, we observe that the value of customization is zero after a critical fraction of fraudulent customers, Figure 4.6b, 4.6a. This case emphasizes the



(a) (b)

Figure 4.6: Value of Information for $\eta^{\mathrm{L}} = 0.15, \eta^{\mathrm{H}} = 0.724$



(a) (b)

Figure 4.7: Value of Information for $\eta^{\mathrm{L}} = 0.5, \eta^{\mathrm{H}} = 0.9$

idea of optimally using the IDS with default configuration for distinct theft levels.

ii. When critical false alarm probability for low theft is more than critical false alarm probability for hight theft i.e. $\eta^{\mathrm{L}} = 0.5, \eta^{\mathrm{H}} = 0.9$: This case emphasizes the non-monotonic behavior of α_0^c with level of theft η. From Figure 4.7a, it is evident that the distribution utility obtains a greater payoff from the perfect information case and the incremental payoff($\sim$2%) is constant for fraction of fraudulent customers greater than a critical fraction. Furthermore, the incremental payoff is negligible for small fractions of fraudulent customers. This is intuitive as the distribution utility is hesitant to use the IDS (equivalently small false alarm probability) for small fractions and hence, the optimal tuning does not create a substantial benefit. However we observe that as the fraction of fraudulent activity increases, the distribution utility is penalized more heavily for installing the non-optimal IDS configuration.

Chapter 5

Manager Implications, Discussions and Extensions

Big data revolution is already making a significant impact in the electric power industry, and is expected to enable fully cyber-enabled operations in many traditional distribution utilities. On one hand routine data collection about customer's consumption patterns and health of grid infrastructure is creating valuable opportunities in the form of demand response programs, outage management and grid management. On the other hand, the emerging cyber-physical security threats can potentially compromise safety critical elements of the grid and/or cause economic losses, e.g. energy diversion, fraud in demand response, and distribution energy generation etc. In this work, we study the economic implications of attacks to vulnerable devices located in customer premises (e.g. AMIs, feeder connecting devices), and specifically focused on distribution utility's incentives to impact fraudulent customers who strategically manipulate their meter readings.

So far data analytics in energy sector is in relative infancy. The distribution utilities in the UD and Europe are primarily focused on using fine-grained data to ensure reliable supply of electricity and execute demand response programs that are tailored to preferences of different customer classes. For example, several data analytics companies such as EnerNOC, Opower, and Bidgely have found business in providing energy management solutions to distribution utilities. Few large utilities, for example PG&E have developed “in-house” analytics teams comprising of experienced data scientists to enhance their business and engineering operations using insights from data. Unfortunately, many distribution utilities are still hesitant to act on the insights/alerts generated by analytics solutions because of insufficient economic incentives or lack of understanding of how valuable these solutions are.

Indeed, there has been some recent research on the profitability and implementation aspects of analytics solutions for the distribution utilities. However, limited work has addressed the question of value of these solutions in strategic environments. In this work, we consider a specific example of energy diversion attacks, that contribute to non-technical losses for distribution utilities. We expect that due to the combined deployment of off-the-shelf IT devices in customer premises and inadequate security levels of these devices, the threats of energy diversion attacks (and similar attacks such as fraud in DR programs) is expected to increase, especially when energy prices are high. Our modeling approach is motivated by these threats, and incorporates the effect of strategic interactions between the distribution utility and customers in the value assessment of IDS/ fraud detection systems.

In this work, we take a game-theoretic approach to assess value of IDS/fraud detection systems by modeling the interactions between a distribution utility and customers. Standard economic principles dictate that without an accurate assessment of value proposition in adoption of new technological solutions, traditional businesses are likely to delay the transition to these technologies. The value of technological solutions such as intrusion/fraud detection technologies is especially difficult due to following reasons: (a) their tuning require accurate understanding of customer behavior (b) they should give reasonable detection accuracy even when the data is noisy and/or manipulated by strategic customers.

Below we provide three main recommendations from our study for inspection of energy diversion attacks and value assessment of IDS.

i. **Is NTL significant for investment in IDS?**

   The distribution utility must first evaluate whether it faces significant non-technical losses. For example, if the average theft level committed by the fraudulent customer is insignificant, or the fraction of fraudulent customers is small, the distribution utility is better off by not investing in costly fraud detection systems but relying on traditional methods of inspection. This is supported by our finding that the losses from more intense inspection may not offset the cost of investigation and the cost of false alarm, even though the cost of IDS is relatively modest.

ii. **Incentives to reduce NTL**

We find that in a range of situations, the distribution utility may not have incentives to investigate fraud, and more broadly, limit non-technical losses. Despite the fact that electricity distribution is a regulated industry, distribution utilities tend to recover NTL's by requesting tariff revisions, which inherently forces the genuine customers to cross-subsidize the fraudulent ones. One of the primary reasons of this problem is that a significant portion of demand is inelastic, and consequently customers do not change their consumption substantially to an increased tariff. Our work provides one way to evaluate the "value of IDS" and "value of information" about theft levels. We believe that these valuations can help in more informed regulatory impositions regarding investment in IDS/fraud detection technologies. In Proposition 7, we show that an IDS of a fixed default configuration , is valuable only if it can guarantee a minimum "quality" threshold, where we define quality of an IDS as the ratio of detection probability and false alarm probability. The threshold accounts for the tradeoff between gain from fraud deterrence and cost of investigation and false alarms. In contrast, Proposition 8 provides that the distribution utility obtains reduced value of information for certain conditions on heterogeneity of theft levels committed by the fraudulent customer.

iii. **Outsourcing vs Insourcing**

As mentioned earlier, the distribution utility can leverage consumption data to address non-payment and energy theft by either outsourcing or insourcing the required data analytics capabilities. By outsourcing, the distribution utility can benefit from the capabilities of analytics service providers and can efficiently address generic energy diversion problems. The outsourcing may be especially beneficial to traditional utilities that may not have the capabilities or lack the requisite expertise to handle large amounts of data. On the contrary, the distribution utility should insource data analytics capabilities if sharing data with third party analytics providers raises data privacy and security issues. The distribution utility should prefer insourcing if it posses the knowledge of specific types of fraud/theft that are typical in distribution networks/feeders that are managed by it. In our framework, IDS with default configuration represents outsourcing by distribution utility who does not possess the means to configure or tune the configuration offered by the analytics services provider. In contrast, IDS of tunable configuration represents another scenario in which the distribution utility chooses an detection probability and false alarm probability combination after accounting for the tradeoff between gain from fraud deterrence and investigation costs, and even the knowledge of average theft level that it faces.

Appendix A

Background Concepts

A.1 Decision Theory

The distribution utility has to decide the underlying distribution ($\mathcal{H}_0$ and $\mathcal{H}_1$) of the customer by observation of the random variable representing electricity consumption. In this section we introduce the two decision making framework in statistics - bayesian and frequentist.

A.1.1 Non-Bayesian Hypothesis Testing

Null Hypothesis Significance Testing

We can run a null hypothesis significance test through the following steps:

- Design an experiment to collect data and choose a test statistic $\mathbf{Y}$ to be computed from the data. The key requirement here is to know the null distribution $p_{y|\mathcal{H}}(\mathbf{Y} = y|\mathcal{H} = \mathcal{H}_0)$ To compute power, one must also know the alternative distribution $p_{y|\mathcal{H}}(\mathbf{Y} = y|\mathcal{H} = \mathcal{H}_1)$.

- Decide if the test is one or two-sided based on $\mathcal{H}_1$ and the form of the null distribution.

- Choose a significance level α for rejecting the null hypothesis

- Run the experiment to collect data $y_1, y_2, ..., y_n$.

- Compute the test statistic $\mathbf{Y}$.

- Compute the p *value* corresponding to $\mathbf{Y}$ using the null distribution.

- If $p < \alpha$, reject the null hypothesis in favor of the alternative hypothesis.

Neyman Pearson Decision Theory

Let us define our two hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ and associated PDFs as $p_{y|\mathcal{H}_0}(y|\mathcal{H}_0)$ and $p_{y|\mathcal{H}_1}(y|\mathcal{H}_1)$ respectively. We define likelihood ratio $\mathcal{L}(y)$ and the Likelihood Ratio Test(LRT) as,

$$\mathcal{L}(y) \triangleq \frac{p_{y|\mathcal{H}}(\mathbf{Y} = y|\mathcal{H} = \mathcal{H}_1)}{p_{y|\mathcal{H}}(\mathbf{Y} = y|\mathcal{H} = \mathcal{H}_0)} \underset{\widehat{H}(y)=\mathcal{H}_0}{\overset{\widehat{H}(y)=\mathcal{H}_1}{\gtrless}} \eta$$

We will choose a decision rule $\widehat{H}$ such that,

$$\max \rho \text{ subject to } \alpha \leqslant \alpha_0$$

where ρ is the probability of detection and α is the probability of false alarm defined as (in continuous case),

$$\rho := \mathbb{P}(\widehat{H}(y) = \mathcal{H}_1|\mathcal{H} = \mathcal{H}_1) = \int_{y_1} p_{y|\mathcal{H}}(y|\mathcal{H}_1)dy$$

$$\alpha := \mathbb{P}(\widehat{H}(y) = \mathcal{H}_1|\mathcal{H} = \mathcal{H}_0) = \int_{y_1} p_{y|\mathcal{H}}(y|\mathcal{H}_0)dy$$

where y_1 represents the region that corresponds $\widehat{H}(y) = \mathcal{H}_1$. We state here without proof that the Neyman-Pearson Lemma states that to maximize ρ subject to the constraint corresponds to using decision rule with Likelihood Ratio above the threshold η, such that,

$$\alpha = \mathbb{P}(L(y) \geqslant \eta|\mathcal{H} = \mathcal{H}_0) = \alpha_0$$

Although we have presented for the continuous case, the same analysis is extensible to discrete distributions.

A.1.2 Bayesian Hypothesis Testing

Let us define our two hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$ and associated PDFs as $p_{y|\mathcal{H}_0}(y|\mathcal{H}_0)$ and $p_{y|\mathcal{H}_1}(y|\mathcal{H}_1)$ and priors P_0 and P_1 respectively. Define $\widetilde{P}(\mathcal{H}_i, \mathcal{H}_j) \triangleq \widetilde{P}_{ij} =$ "profit" of $\widehat{H} = \mathcal{H}_i$ when the correct hypothesis is $\mathbf{H} = \mathcal{H}_j$. $\therefore$ find the function $\widehat{H}(\mathbf{Y})$

$$\widehat{H}(\mathbf{Y}) = \arg\max_{f()} \varphi(f) \ , \ \varphi(f) \triangleq \mathbb{E}[\widetilde{P}(\mathbf{H}, f(\mathbf{Y}))]$$

By law of iterated expectation, we have,

$$\varphi(f) = \sum E[\widetilde{P}(\mathcal{H}, f(\mathbf{Y}))|\mathbf{Y} = y] \ P_{\mathbf{Y}}(y), \quad \widetilde{\varphi}(f, y) = E[\widetilde{P}(\mathcal{H}, f(\mathbf{Y}))|\mathbf{Y} = y]$$

Furthermore since we know that $P_{\mathbf{Y}}(y) \geqslant 0$, maximizing $\varphi(f)$ to find $\widehat{H}(y)$ is equivalent to maximizing $\widetilde{\varphi}(f, y)$ to find the optimal $\widehat{H}$ at all y.

$$\widehat{H}(y) = \mathcal{H}_0, \ \widetilde{\varphi}(H_0, y) = \widetilde{P}_{01} \ p_{H|Y}(H_1|y) + \widetilde{P}_{00} \ p_{H|Y}(H_0|y)$$

$$\widehat{H}(y) = \mathcal{H}_1, \ \widetilde{\varphi}(H_1, y) = \widetilde{P}_{11} \ p_{H|Y}(H_1|y) + \widetilde{P}_{10} \ p_{H|Y}(H_0|y)$$

The defender will maximize the expected profit for every value of y by choosing $\max\{\widetilde{\varphi}(H_0, y), \widetilde{\varphi}(H_1, y)\}$

A.2 Game Theory

A.2.1 Strategic Form Game

Definition 1.(Strategic Form Game) A strategic form game is a triplet $< \mathcal{I}, (S_i), (u_i)_{i\in\mathcal{I}} >$ where

- $\mathcal{I}$ is a finite set of players, $\mathcal{I} = 1, ..., I$.
- S_i is a non-empty set of available actions for player i.
- $u_i : S \to \mathbb{R}$ is the utility (payoff) function of player i where $S = \prod_{i\in\mathcal{I}} Si$.

A.2.2 Nash Equilibrium / Mixed Strategy Nash Equilibrium

Definition 2. (Nash Equilibrium). A (pure strategy) Nash equilibrium of a strategic form game $< \mathcal{I}, (S_i), (u_i)_{i \in \mathcal{I}} >$ is a strategy profile $s^* \in S$ such that for all $i \in \mathcal{I}$, we have

$$u_i(s_i^*, s_{-i}^*) \geqslant u_i(s_i, s_{-i}^*) \quad for\ all\ s_i \in S_i$$

Definition 3. (Mixed Strategy Nash Equilibrium). A mixed strategy profile σ^* is a mixed strategy Nash equilibrium if for each player i,

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geqslant u_i(\sigma_i, \sigma_{-i}^*) \quad for\ all\ \sigma_i \in \Sigma_i$$

where Σ_i is the set of probability measures over the pure strategy (action) set S_i.

A.2.3 Bayesian Game

Definition 4. (Bayesian Games). A Bayesian game consists of

- $\mathcal{I}$ is a finite set of players, $\mathcal{I} = 1, ..., I$.
- S_i is a non-empty set of available actions for player i.
- A set of types for each player $i : \theta_i \in \Theta_i$
- A payoff function for each player $i : u_i(s_1, ..s_I, \theta_1, ..\theta_I)$
- A (joint) probability distribution $p(\theta_1, \theta_2, ..., \theta_I)$

Importantly, throughout in Bayesian games, the strategy spaces, the payoff functions, possible types, and the prior probability distribution are assumed to be common knowledge.

Definition 5. (Bayesian Nash Equilibrium): The strategy profile $s()$ is a (pure strategy) Bayesian Nash equilibrium if for all $i \in \mathcal{I}$ and for all $\theta_i \in \Theta_i$, we have that

$$s_i(\theta_i) \in \arg\max_{s_i{}' \in S_i} \sum_{\theta_{-i}} p(\theta_{-i}|\theta_i) u_i(s_i{}', s_{-i}(\theta_{-i}), \theta_i, \theta_{-i}) \tag{75}$$

or in non-finite case, we have,

$$s_i(\theta_i) \in \arg\max_{s_i{}' \in S_i} \int u_i(s_i{}', s_{-i}(\theta_{-i}), \theta_i, \theta_{-i}) P(d\theta_{-i}|\theta_i) \tag{76}$$

Theorem 4. *: Consider a finite incomplete information (Bayesian) game. Then a mixed strategy Bayesian Nash equilibrium exists.*

Bibliography

[1] Saurabh Amin, Galina A Schwartz, Alvaro A Cardenas, and S Shankar Sastry. Game theoretic models of electricity theft detection in smart utility networks. 2015.

[2] Saurabh Amin, Galina A Schwartz, and Hamidou Tembine. Incentives and security in electricity distribution networks. In *International Conference on Decision and Game Theory for Security*, pages 264–280. Springer, 2012.

[3] Pedro Antmann. Reducing technical and non-technical losses in the power sector. Technical report, World Bank, July 2009.

[4] Andrew Appel. Security seals on voting machines: A case study. *ACM Transactions on Information and Systems Security*, 14:1–29, 2011.

[5] Rudolf Avenhaus, Bernhard Von Stengel, and Shmuel Zamir. Inspection games. *Handbook of game theory with economic applications*, 3:1947–1987, 2002.

[6] CJ Bandim, JER Alves Jr, AV Pinto Jr, FC Souza, MRB Loureiro, CA Magalhaes, and F. Galvez-Durand. Identification of energy theft and tampered meters using a central observer meter: a mathematical approach. In *Transmission and Distribution Conference and Exposition, 2003 IEEE PES*, volume 1, pages 163–168. IEEE, 2003.

[7] D.C. Bergman, Dong Jin, J.P. Juen, N. Tanaka, C.A. Gunter, and A.K. Wright. Distributed non-intrusive load monitoring. In *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pages 1 –8, jan. 2011.

[8] Markus Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jorg Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pages 93–104. ACM, 2000.

[9] B.E. Brodsky and B.S. Darkhovsky. *Non-Parametric Methods in Change-Point Problems.* Kluwer Academic Publishers, 1993.

[10] Alvaro A. Cardenas, Saurabh Amin, Zong-Syun Lin, Yu-Lun Huang, Chi-Yen Huang, and Shankar Sastry. Attacks against process control systems: Risk assessment, detection, and response. In *Proceedings of ACM Symposium on Information, Computer and Communications Security (AsiaCCS 2011)*. ACM, 2010.

[11] Alvaro A Cárdenas, Saurabh Amin, Galina Schwartz, Roy Dong, and Shankar Sastry. A game theory model for electricity theft detection and privacy-aware control in ami systems. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1830–1837. IEEE, 2012.

[12] Huseyin Cavusoglu, Birendra Mishra, and Srinivasan Raghunathan. The value of intrusion detection systems in information technology security architecture. *Information Systems Research*, 16(1):28–46, 2005.

[13] Huseyin Cavusoglu, Srinivasan Raghunathan, and Wei T Yue. Decision-theoretic and game-theoretic approaches to it security investment. *Journal of Management Information Systems*, 25(2):281–304, 2008.

[14] Mike Davis. Smartgrid device security. adventures in a new medium. http://www.blackhat.com/presentations/bh-usa-09/MDAVIS/BHUSA09-Davis-AMI-SLIDES.pdf, July 2009.

[15] E. De Buda. System for accurately detecting electricity theft. US Patent Application 12/351978, Jan. 2010.

[16] S.S.S.R. Depuru, Lingfeng Wang, and V. Devabhaktuni. Support vector machine based data classification for detection of electricity theft. In *Power Systems Conference and Exposition (PSCE), 2011 IEEE/PES*, pages 1 –8, march 2011.

[17] ECI Telecom. *Fighting Electricity Theft with Advanced Metering Infrastructure*, March 2011.

[18] Chet Geschickter. *The Emergence of Meter Data Management (MDM): A Smart Grid Information Strategy Report.* GTM Research, 2010.

[19] Peter J Huber and Volker Strassen. Minimax tests and the neyman-pearson lemma for capacities. *The Annals of Statistics*, pages 251–263, 1973.

[20] Rong Jiang, Rongxing Lu, Ye Wang, Jun Luo, Changxiang Shen, and Xuemin Sherman Shen. Energy-theft detection issues for advanced metering infrastructure in smart grid. *Tsinghua Science and Technology*, 19(2):105–120, 2014.

[21] Tamim Saleh Jon Brock, Stephan Lehrke and Nadija Yousif. Making big data work: Retail energy. https://www.bcgperspectives.com/content/articles/energy_environment_technology_strategy_making_big_data_work_retail_energy/, July 2014.

[22] G. Kalogridis, C. Efthymiou, S.Z. Denic, T.A. Lewis, and R. Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 232 –237, oct. 2010.

[23] Brian Krebs. FBI: smart meter hacks likely to spread. http://krebsonsecurity.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/, April 2012.

[24] Michael LeMay and Carl A Gunter. Cumulative attestation kernels for embedded systems. *Smart Grid, IEEE Transactions on*, 3(2):744–760, 2012.

[25] Adam Lesser. When big IT goes after big data on the smart grid. http://gigaom.com/cleantech/when-big-it-goes-after-big-data-on-the-smart-grid-2/, March 2012.

[26] Yu Liu, Cristina Comaniciu, and Hong Man. A bayesian game approach for intrusion detection in wireless ad hoc networks. In *Proceeding from the 2006 Workshop on Game Theory for Communications and Networks*, GameNets '06, New York, NY, USA, 2006. ACM.

[27] Daisuke Mashima and Alvaro A Cárdenas. Evaluating electricity theft detectors in smart grid networks. In *Research in Attacks, Intrusions, and Defenses*, pages 210–229. Springer, 2012.

[28] S. McLaughlin, D. Podkuiko, and Patrick McDaniel. Energy theft in the advanced metering infrastructure. In *Proceedings of CRITIS 09, 4th International Conference on Critical Information Infrastructures Security*, 2009.

[29] S. McLaughlin, D. Podkuiko, S. Miadzvezhanka, A. Delozier, and Patrick McDaniel. Multi-vendor penetration testing in the advanced metering infrastructure. In *Proceedings of the Annual Computer Security Applications Conference (ACSAC)*, Dec. 2010.

[30] Stephen McLaughlin, Brett Holbert, Ahmed Fawaz, Robin Berthier, and Saman Zonouz. A multi-sensor energy theft detection framework for advanced metering infrastructures. *Selected Areas in Communications, IEEE Journal on*, 31(7):1319–1330, 2013.

[31] Stephen E. McLaughlin, Patrick McDaniel, and William Aiello. Protecting consumer privacy from electric load monitoring. In *ACM Conference on Computer and Communications Security*, pages 87–98, 2011.

[32] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad. Nontechnical loss detection for metered customers in power utility using support vector machines. *IEEE Transactions on Power Delivery Systems*, 25(2):1162–1171, April 2010.

[33] A. Nizar and Z. Dong. Identification and detection of electricity customer behaviour irregularities. *Power Systems Conference and Exposition (PSCE)*, pages 1–10, March 2009.

[34] Doug Peeples. India plans 21.6 billion spend on smart grid to stop theft. http://www.smartgridnews.com/story/india-plans-216-billion-spend-smart-grid-stop-theft/2015-01-21, January 2015.

[35] J. Peralta, A. Cheung, M. Pedley, and J. Tao. BC hydro's methodology for energy losses assessment in distribution systems. In *Electrical Power Energy Conference (EPEC), 2009 IEEE*, pages 1 –6, oct. 2009.

[36] Dale Peterson. AppSecDC in review: Real-world backdoors on industrial devices. http://www.digitalbond.com/2012/04/11/appsecdc-in-review/, April 2012.

[37] Smart Grid Interoperability Panel, editor. *NISTIR 7628. Guidelines for Smart Grid Cyber Security*. NIST, August 2010.

[38] W Alan Snook. Why its time to get tougher on energy theft. http://grid2020.com/2014/03/time-get-tougher-energy-theft/, March 2014.

[39] Robin Sommer and Vern Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *IEEE Symposium on Security and Privacy*, 2010.