Journal of Privacy and Confidentiality (2015–2016) **7**, Number 1, 99–127

Designing Incentive Schemes for Privacy-Sensitive Users

Chong Huang*, Lalitha Sankar†, and Anand D. Sarwate‡

Abstract. Businesses (*retailers*) often wish to offer personalized advertisements (*coupons*) to individuals (*consumers*), but run the risk of strong reactions from consumers who want a customized shopping experience but feel their privacy has been violated. Existing models for privacy such as differential privacy or information theory try to *quantify* privacy risk but do not capture the subjective experience and heterogeneous *expression* of privacy-sensitivity. We propose a Markov decision process (MDP) model to capture (i) different consumer privacy sensitivities via a time-varying state; (ii) different coupon types (action set) for the retailer; and (iii) the action-and-state-dependent cost for perceived privacy violations. For the simple case with two states ("Normal" and "Alerted"), two coupons (targeted and untargeted) model, and consumer behavior statistics known to the retailer, we show that a stationary threshold-based policy is the optimal coupon-offering strategy for a retailer that wishes to minimize its expected discounted cost. The threshold is a function of all model parameters; the retailer offers a targeted coupon if their belief that the consumer is in the "Alerted" state is below the threshold. We extend this two-state model to consumers with multiple privacy-sensitivity states as well as coupon-dependent state transition probabilities. Furthermore, we study the case with imperfect (noisy) cost feedback from consumers and uncertain initial belief state.

Keywords-Privacy, Markov decision processes, retailer-consumer interaction, optimal policies.

1 Introduction

Programs such as retailer "loyalty cards" allow companies to automatically track a customer's financial transactions, purchasing behavior, and preferences. They can then use this information to offer customized incentives, such as discounts on related goods. Consumers may benefit from retailer's knowledge by using more of these targeted discounts or coupons while shopping. However, in some cases the coupon offer implies that the retailer has learned something sensitive or private about the consumer. For example, as noted by Hill (2012), a retailer could infer a consumer's pregnancy. Such violations

*Electrical Engineering Department, Arizona State University, Tempe, AZ, `mailto:chuang83@asu.edu`.

†Department of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, `mailto:lalithasankar@asu.edu`.

‡Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, New Brunswick, NJ, `mailto:asarwate@ece.rutgers.edu`.

The work of Chong Huang and Lalitha Sankar was supported in part by the National Science Foundation under award NSF CAREER CCF-1350914. The work of Anand D. Sarwate was supported in part by the National Science Foundation under award NSF CAREER CCF-1453432.

may make consumers skittish about purchasing from such retailers.

However, modeling the privacy-sensitivity of a consumer is not always straightforward: widely-studied models for quantifying privacy risk using differential privacy or information theory do not capture the subjective experience and heterogeneous *expression* of consumer privacy. The goal of this paper is to introduce a framework to model the consumer-retailer interaction problem and better understand how retailers can develop coupon-offering policies that balance their revenue objectives while being sensitive to consumer privacy concerns. The main challenge for the retailer is that the consumer's responses to coupons are not known a priori; furthermore, consumers do not "add noise" to their purchasing behavior as a mechanism to stay private. Rather, the offer of a coupon may provoke a reaction from the consumer, ranging from "unaffected" to "ambiguous" or "partially concerned" to "creeped out." This reaction is mediated by the consumer's sensitivity level to privacy violations, and it is these levels that we seek to model via a Markov decision process. These privacy-sensitivity states of the consumers are often revealed to the retailer through their purchasing patterns. In the simplest case, they may accept or reject a targeted coupon. We capture these aspects in our model and summarize our main contributions below.

1.1 Main Contributions

We propose a partially-observed Markov decision process (POMDP) model for this problem in which the consumer's state encodes their privacy sensitivity, and the retailer can offer different levels of privacy-violating coupons. The simplest instance of our model is one with two states for the consumer, denoted as "Normal" and "Alerted," and two types of coupons: untargeted *low privacy* (LP) or targeted *high privacy* (HP). At each time, the retailer may offer a coupon and the consumer transitions from one state to another according to a Markov chain that is independent of the offered coupon. The retailer suffers a cost that depends both on the type of coupon offered and the state of the consumer. The costs reflect the advantage of offering targeted HP coupons relative to untargeted LP ones while simultaneously capturing the risk of doing so when the consumer is already "Alerted."

Under the assumption that the retailer (via surveys or prior knowledge) knows the statistics of the consumer Markov process, i.e., the likelihoods of becoming "Alerted" and staying "Alerted," and a belief about the initial consumer state, we study the problem of determining the optimal coupon-offering policy that the retailer should adopt to minimize the long-term discounted costs of offering coupons. We extend the simple model above to multiple states and coupon-dependent transitions. We model the latter via two Markov processes for the consumer, one for each type (HP or LP) of coupon such that a persnickety consumer who is easily "Alerted" will be more likely to do so when offered an HP (relative to LP) coupon. Furthermore, for noisy costs, we propose a heuristic method to compute the decision policy. Moreover, if the initial belief state is unknown to the retailer, we use a Bayesian model to estimate the belief state. Our main results can be summarized as follows:

1. There exists an optimal, stationary, threshold-based policy for offering coupons such that a HP coupon is offered only if the belief of being in the "Alerted" state at each interaction time is below a certain threshold; this threshold is a function of all the model parameters. This structural result holds for multiple states and coupon-dependent transitions.

2. The threshold for offering a targeted HP coupon increases in the following cases:

   (a) once "Alerted," the consumer remains so for a while – the retailer is more willing to take risks since the the consumer takes a while to transition to "Normal";

   (b) the consumer is very unlikely to get "Alerted";

   (c) the cost of offering an untargeted LP coupon is high and close to the cost of offering a targeted HP coupon to an "Alerted" consumer; and

   (d) when the retailer does not discount the future heavily (future rewards nearly as important as present), the retailer stands to benefit by offering HP coupons for a larger set of beliefs about the consumer's state. Conversely, when the retailer discounts the future heavily, it values the present rewards more than future rewards. Thus, the retailer tends to play conservatively so that it will not "creep out" the consumer in the present.

3. For the coupon-dependent Markov model for the consumer, the threshold is smaller than for the non-coupon dependent case which encapsulates the fact that highly sensitive consumers will force the retailers to behave more conservatively.

4. By adopting a heuristic threshold policy computed by the mean value of costs, the retailer can minimize the discounted cost effectively even if costs are noisy. Moreover, the Bayesian approach helps the retailer to estimate the consumer state when the initial belief state is unknown.

Our results use many fundamental tools and techniques from the theory of MDPs through appropriate and meaningful problem modeling. We briefly review the related literature in consumer privacy studies as well as MDPs.

1.2 Related Work

Several economic studies have examined consumers' attitudes towards privacy via surveys and data analysis including studies on the benefits and costs of using private data (e.g., Aquisti, 2010). However, to date, no formal model has been proposed that captures consumers' privacy sensitivity. Most computational and theoretical frameworks for addressing privacy-utility tradeoffs with economic actors involve users trying to control the amount of private information disclosed. For example, differential privacy (see Dwork, 2011) has been used for modeling the value of private data for market design (see Ghosh et al., 2013) and partitioning goods with private valuation functions (e.g., Hsu et al., 2013). Information-theoretic measures of privacy leakage have been used to study competition among data holders in a model for utility companies (see Sankar et al., 2011). However, the private-information disclosure model neither captures interactions between consumers and retailers nor temporal dynamics of consumers' (subjective) privacy sensitivities.

In our model, consumers express their privacy sensitivity through behavior, as opposed to explicitly controlling privacy leakage through disclosure of private information. We model interactions between a retailer and a consumer via a discrete-time system. At each time, the consumer has a discrete-valued state and the retailer may offer different types of coupons. The consumer responds by imposing a cost on the retailer that depends on the coupon offered and its own state. This model captures different consumer privacy sensitivities and external retailer action via a time-varying state and different coupon types (action set), respectively. Moreover, the subjective experience and heterogeneous expression of privacy sensitivity is captured via the action-and-state-dependent cost for perceived privacy violations.

Markov decision processes (MDPs) are common discrete time mathematical models for decision making when observable outputs are partially dependent on internal states and exterior inputs. It has been widely used for decades across many fields (see Feinberg et al., 2002 and Puterman, 2005); in particular, our model is related to problems in control with communication constraints (e.g., Lipsa et al., 2011 and Nayyar, et al., 2013) where state estimation has a cost. Our costs are action and state dependent and we consider a different optimization problem. Classical target-search problems (e.g., Macphee et al., 1995) also have optimal policies that are thresholds, but in our model the retailer goal is not to estimate the consumer state but to minimize cost. The model we use is most similar to Ross's model of product quality control with deterioration (see Ross, 1971), which was more recently used by Laourine and Tong (2010) to study the Gilbert-Elliot channel in wireless communications, in which the channel has two states and the transmitter has two actions (transmit or not). We cannot apply their results directly due to our different cost structure, but use ideas from their proofs. Furthermore, we go beyond these works to study privacy-utility tradeoffs in consumer-retailer interactions with more than two states and action-dependent transition probabilities. We apply more general MDP analysis tools to address our formal behavioral model for privacy-sensitive consumers. In the context of privacy, MDPs have been used by Venkitasubramaniam (2013) to study privacy and utility trade-off in control systems with time-varying state by quantifying privacy via the information-theoretic equivocation function. However, in his paper, the state is really the state of a control system rather than the state of privacy sensitivity of a consumer. In our work we do not quantify *privacy loss* directly; instead we model *privacy-sensitivity states* and resulting user behavior via MDPs to determine interaction policies that can benefit both consumers and retailers. To the best of our knowledge, a formal model for consumer-retailer interactions and the related privacy issues have not been studied before; in particular, our work focuses on explicitly considering the consequence to the retailer of the consumers' awareness of privacy violations.

While the MDP model used in this paper is simple, its application to the problem of revenue maximization with privacy-sensitive consumers is novel. We show that the optimal stationary policy exists and it is a threshold on the probability of the consumer being alerted. We extend the model to cases of consumers with multiple states and consumers with coupon-dependent transition probabilities. Our basic model assumes the probability of the consumer being alerted can be inferred from the received costs. When costs are stochastic, we use a Bayesian estimator to track this probability and propose a heuristic coupon offering policy for this setting. In the conclusion we describe several other interesting avenues for future work.

The paper is organized as follows: Section 2 introduces the system model and its extensions. The main result for known consumer statistics is presented in Section 3. Section 4 and 5 discuss optimal stationary policy results for consumers with coupon dependent response and noisy costs with unknown initial belief, respectively. Finally, some concluding remarks and future work are provided in Section 6.

2 System Model

We model interactions between a retailer and a consumer via a discrete-time system (Figure 1). At each time t, the consumer has a discrete-valued state and the retailer may offer one of two coupons: high privacy risk (HP) or low privacy risk (LP). We assume a sophisticated consumer who can distinguish whether a coupon is HP or LP and responds to the personalized coupon by imposing a cost on the retailer that depends on the coupon offered and its own state. For example, a consumer who is "Alerted" (privacy-aware) may respond to an HP coupon by imposing a high cost to the retailer, such as reducing purchases at the retailer. The retailer's goal is to decide which type of coupon to offer at each time t to minimize its cost.

2.1 Two-State Consumer with Coupon Independent Transitions.

Consumer Model

Model Assumption 1. ***(Consumer's state)*** *We model the consumer's response to coupons by assuming them to be in one of several states. Each state corresponds to a type of consumer behavior in terms of purchasing (Privacy sensitivity).*

For this paper, we first focus on the two-state case; the consumer may be Normal or $\mathsf{Alerted}$. Later we will extend this model to multiple consumer states, consumer with coupon dependent response, and unknown initial consumer state cases. The consumer state at time t is denoted by $G_t \in \{\mathsf{Normal}, \mathsf{Alerted}\}$. If a consumer is in Normal state, the consumer is less sensitive to coupons from the retailer in terms of privacy. However, in the $\mathsf{Alerted}$ state, the consumer is likely to be more sensitive to coupons offered by the retailer, since it is more cautious about revealing information to the retailer. The evolution of the consumer state is modeled as an infinite-horizon discrete time Markov chain (Figure 1). The consumer starts out in a random initial state unknown to the retailer and the transition of the consumer state is independent of the action of the retailer. A *belief state* is a probability distribution over possible states in which the consumer could be. The belief of the consumer being in $\mathsf{Alerted}$ state at time t is denoted by p_t. We define $\lambda_{N,A} = Pr[G_t = \mathsf{Alerted}|G_{t-1} = \mathsf{Normal}]$ to be the transition probability from Normal state to $\mathsf{Alerted}$ state and $\lambda_{A,A} = Pr[G_t = \mathsf{Alerted}|G_{t-1} = \mathsf{Alerted}]$ to be the probability of staying in $\mathsf{Alerted}$ state when the previous state is also $\mathsf{Alerted}$. The transition matrix $\mathbf{\Lambda}$ of the Markov chain can be written as

$$\mathbf{\Lambda} = \begin{pmatrix} 1-\lambda_{N,A} & \lambda_{N,A} \\ 1-\lambda_{A,A} & \lambda_{A,A} \end{pmatrix}. \tag{1}$$

We assume the transition probabilities are known to the retailer; this may come from statistical analysis such as a survey of consumer attitudes. The one step transition function, defined by

$$T(p_t) = (1-p_t)\lambda_{N,A} + p_t\lambda_{A,A}, \tag{2}$$

represents the belief that the consumer is in $\mathsf{Alerted}$ state at time $t+1$ given p_t, the $\mathsf{Alerted}$ state belief at time t.

Model Assumption 2. ***(State transitions)*** *Consumers have an inertia in that they tend to stay in the same state. Moreover, once consumers feel their privacy is violated, it will take some time for them to come back to* Normal *state.*

The above assumption implies $\lambda_{A,A} \geq 1-\lambda_{A,A}$, $1-\lambda_{N,A} \geq \lambda_{N,A}$, and $\lambda_{N,A} \geq 1-\lambda_{A,A}$. Thus, by combining the above three inequalities, we have $\lambda_{A,A} \geq \lambda_{N,A}$.

Retailer Model

At each time t, the retailer can take an *action* by offering a coupon to the consumer. We define the action at time t to be $u_t \in \{\mathsf{HP}, \mathsf{LP}\}$, where HP denotes offering a high privacy risk coupon (e.g., a targeted coupon) and LP denotes offering a low privacy risk coupon (e.g., a generic coupon). The retailer's utility is modeled by a *cost* (negative revenue) which depends on the consumer's state and the type of coupon being offered. If the retailer offers an LP coupon, it suffers a cost C_L independent of the consumer's state: offering LP coupons does not reveal anything about the state. However, if the retailer offers an HP coupon, then the cost is C_{HN} or C_{HA} depending on whether the consumer's state is Normal or $\mathsf{Alerted}$. Offering an HP (high privacy risk, targeted) coupon to a Normal consumer should incur a low cost (high reward), but offering an HP coupon to an $\mathsf{Alerted}$ consumer should incur a high cost (low reward) since an $\mathsf{Alerted}$ consumer is privacy-sensitive. Thus, we assume $C_{HN} \leq C_L \leq C_{HA}$.

Under these conditions, the retailer's objective is to choose u_t at each time t to minimize the total cost inccured over the entire time horizon. The HP coupon reveals information about the state through the cost, but is risky if the consumer is alerted, creating a tension between cost minimization and acquiring state information.

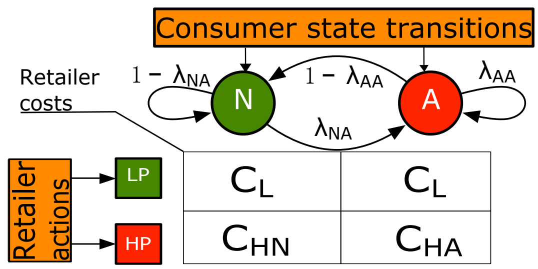


Figure 1: Table of costs to the retailer for offering LP and HP coupons in each privacy sensitive state of the consumer between which the state transitions under a Markov model

Minimum Cost Function

We define $C(p_t, u_t)$ to be the expected cost acquired from an individual consumer at time t where p_t is the probability that the consumer is in Alerted state and u_t is the retailer's action:

$$C(p_t, u_t) = \begin{cases} C_L & \text{if } u_t = \mathsf{LP} \\ (1-p_t)C_{HN} + p_t C_{HA} & \text{if } u_t = \mathsf{HP} \end{cases}. \tag{3}$$

Since the retailer knows the consumer state from the incurred cost only when an HP coupon is offered, the state of the consumer may not be directly observable to the retailer. Therefore, the problem is actually a Partially Observable Markov Decision Process (POMDP).

We model the cost of violating a consumer's privacy as a short term effect. Thus, we adopt a discounted cost model with discount factor $\beta \in (0, 1)$. We define $\mathcal{P} = \{[0, 1]\}$ and $\mathcal{U} = \{\mathsf{LP}, \mathsf{HP}\}$ to be the belief space and the action space, respectively. At each time t, the retailer has to choose which action u_t to take in order to minimize the expected discounted cost over infinite horizon. A policy π for the retailer is a rule that selects a coupon to offer at each time, i.e., $\pi : \mathcal{P} \to \mathcal{U}$. Thus, given that the belief of the consumer being in Alerted state at time t is p_t and the policy is π, the infinite-horizon discounted cost starting from t is

$$V_\beta^{\pi,t}(p_t) = \mathbb{E}_\pi \left[\sum_{i=t}^{\infty} \beta^i C(p_i, u_i) | p_t \right], \tag{4}$$

where $\mathbb{E}_\pi$ indicates the expectation over the policy π. The objective of the retailer is equivalent to minimizing the discounted cost over all possible policies. Thus, we define the minimum cost function starting from time t over all policies to be

$$V_\beta^t(p_t) = \min_\pi V_\beta^{\pi,t}(p_t) \text{ for all } p_t \in [0, 1]. \tag{5}$$

We define $V_{\beta,u_t}^t(p_t)$ to be the infinite-horizon discounted cost starting from t with initial action u_t and p_{t+1} to be the belief of the consumer being in Alerted state at time

$t+1$. The minimum cost function $V_\beta^t(p_t)$ satisfies the Bellman equation (see Bertsekas, 1995):

$$V_\beta^t(p_t) = \min_{u_t \in \{\mathsf{HP},\mathsf{LP}\}} \{V_{\beta,u_t}^t(p_t)\}, \tag{6}$$

$$V_{\beta,u_t}^t(p_t) = \beta^t C(p_t, u_t) + V_\beta^{t+1}(p_{t+1}|p_t, u_t). \tag{7}$$

An optimal policy is *stationary* if it is a deterministic function of states, i.e., the optimal action at a particular state is the optimal action in this state at all times. In the context of our model, the optimal stationary policy is a deterministic and time invariant function mapping $\mathcal{P}$ into $\mathcal{U}$. Since the problem is an infinite-horizon, finite state and finite action POMDP with discounted cost, finding an optimal strategy to this problem is equivalent to solving an associated MDP problem in belief space (see Bonet, 2002), which is an infinite-horizon discounted MDP with finite action space and uncountably infinite state space. By Theorem 6.3 and its generalization in Ross (1992), there exists an optimal stationary policy π^* in the belief space such that starting from time t,

$$V_\beta^t(p_t) = V_\beta^{\pi^*,t}(p_t). \tag{8}$$

Thus, only the optimal stationary policy is considered because it is tractable and achieves the same minimum cost as any optimal non-stationary policy.

By (6) and (7), the minimum cost function evolves as follows. If an HP coupon is offered at time t, the retailer can perfectly infer the consumer state based on the incurred cost. Therefore,

$$V_{\beta,\mathsf{HP}}^t(p_t) = \beta^t C(p_t, \mathsf{HP}) + (1-p_t)V_\beta^{t+1}(\lambda_{N,A}) + p_t V_\beta^{t+1}(\lambda_{A,A}). \tag{9}$$

If an LP coupon is offered at time t, the retailer cannot infer the consumer state from the cost since both Normal and Alerted consumer impose the same cost C_L. Hence, the discounted cost function can be written as

$$\begin{aligned} V_{\beta,\mathsf{LP}}^t(p_t) &= \beta^t C(p_t, \mathsf{LP}) + V_\beta^{t+1}(p_{t+1}) \\ &= \beta^t C_L + V_\beta^{t+1}(T(p_t)). \end{aligned} \tag{10}$$

Correspondingly, the minimum cost function is given by

$$V_\beta^t(p_t) = \min\{V_{\beta,\mathsf{LP}}^t(p_t), V_{\beta,\mathsf{HP}}^t(p_t)\}. \tag{11}$$

In the sequel, we also consider the following value functions in addition to those defined above. For notational clarity, we define them all here.

- $V_\beta^{t\sim k}(p)$: the minimum cost when the decision horizon starts from t and only spans k stages with initial belief p at time t.
- $V_{\beta,u_t}^{t\sim k}(p)$: the minimum cost when the decision horizon starts from t and only spans k stages with initial belief p and initial action u_t.
- $V_\beta(p)$: the minimum cost function starting from $t = 0$.

We now describe some simple extensions of this basic model.

2.2 Consumer with Multi-Level Alerted States

In this section, the case that the consumer has multiple $\mathsf{Alerted}$ states is studied. Without loss of generality, we define $G_t \in \{\mathsf{Normal}, \mathsf{Alerted}_1, \ldots \mathsf{Alerted}_K\}$ to be the consumer state at time t. If the consumer is in $\mathsf{Alerted}_k$ state, it is even more cautious about coupons than in $\mathsf{Alerted}_{k-1}$ state. Beliefs of the consumer being in Normal, $\mathsf{Alerted}_1, \ldots, \mathsf{Alerted}_K$ state at time t are defined by $\bar{\mathbf{p}}_t = (p_{N,t}, p_{A_1,t}, \ldots, p_{A_K,t})^T$. At each time t, the retailer can offer either an HP or an LP coupon. Costs of the retailer when an HP coupon is offered while the state of the consumer is Normal, $\mathsf{Alerted}_1, \ldots, \mathsf{Alerted}_K$ are defined by $\bar{\mathbf{C}} = (C_{HN}, C_{HA_1}, \ldots, C_{HA_K})^T$. If an LP coupon is offered, no matter in which state, the retailer gets a cost of C_L. We assume that $C_{HA_K} \geq \cdots \geq C_{HA_1} \geq C_L \geq C_{HN}$. The minimum cost function evolves as follows:

$$V_\beta^t(\bar{\mathbf{p}}_t) = \min\{V_{\beta,\mathsf{LP}}^t(\bar{\mathbf{p}}_t), V_{\beta,\mathsf{HP}}^t(\bar{\mathbf{p}}_t)\}, \tag{12}$$

where $V_{\beta,\mathsf{LP}}^t(\bar{\mathbf{p}}_t) = \beta^t C_L + V_\beta^{t+1}(\bar{\mathbf{p}}_{t+1})$ and $V_{\beta,\mathsf{HP}}^t(\bar{\mathbf{p}}_t) = \beta^t \bar{\mathbf{p}}_t^T \bar{\mathbf{C}} + V_\beta^{t+1}(\bar{\mathbf{p}}_{t+1})$ represents the cost of offering an LP and an HP coupon, respectively. This model can be generalized to consumer with finitely many states.

2.3 Consumer with Coupon Dependent Transitions

In the previous formulations, we assume that the consumer's state transition is independent of the retailer's action. A natural extension is the case where the action of the retailer can affect the dynamics of the consumer state evolution (Figure 2). Generally, a consumer's reactions to HP and LP coupons are different. For example, a consumer is likely to feel less comfortable when being offered a coupon on medication (HP) than food (LP). Thus, in Section 4, we assume that the Markov transition probabilities are dependent on the coupon offered with transition matrix given by $\mathbf{\Lambda}_{\mathsf{LP}}(\mathbf{\Lambda}_{\mathsf{HP}})$, where $\mathbf{\Lambda}_{\mathsf{LP}}$ and $\mathbf{\Lambda}_{\mathsf{HP}}$ are defined as:

$$\mathbf{\Lambda}_{\mathsf{LP}} = \begin{pmatrix} 1-\lambda_{N,A} & \lambda_{N,A} \\ 1-\lambda_{A,A} & \lambda_{A,A} \end{pmatrix}, \mathbf{\Lambda}_{\mathsf{HP}} = \begin{pmatrix} 1-\lambda'_{N,A} & \lambda'_{N,A} \\ 1-\lambda'_{A,A} & \lambda'_{A,A} \end{pmatrix}. \tag{13}$$

Thus, the minimum cost function is given by (11), where $V_{\beta,\mathsf{LP}}^t(p_t) = \beta^t C(p_t, \mathsf{LP}) + V_\beta^{t+1}(T(p_t))$ and $V_{\beta,\mathsf{HP}}^t(p_t) = \beta^t C(p_t, \mathsf{HP}) + (1-p_t)V_\beta^{t+1}(\lambda'_{N,A}) + p_t V_\beta^{t+1}(\lambda'_{A,A})$ denotes the cost function of using an LP coupon and an HP coupon, respectively. $T(p_t)$ is the one step transition given by $T(p_t) = \lambda_{N,A}(1-p_t) + \lambda_{A,A}p_t$.

2.4 Policies under Noisy Cost Feedback and Uncertain Initial Belief

Consider a setting in which the feedback regarding the cost may be noisy, e.g., the cost incurred by the consumer's response to the coupon is not deterministic. For each individual consumer, the state transition is independent of the action of the retailer. For given state G_t and action u_t, define the distribution of observing a cost $C_t = c$ to be $f(c|G_t, u_t)$. In this case, the threshold policy computed using costs might not be

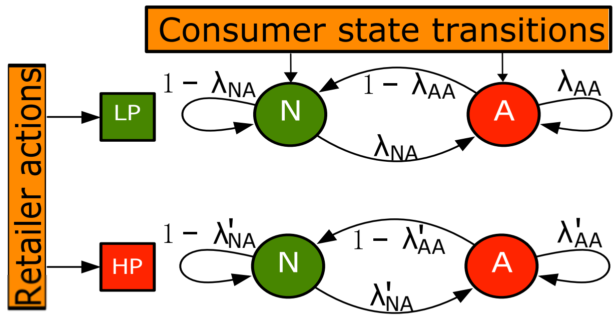


Figure 2: Coupon type (HP or LP) dependent Markov state transition model for the consumer.

optimal. Moreover, if the initial belief is unknown to the retailer, it has to estimate the consumer state before making decision. Thus, we propose some alternative approaches to decide which coupon to offer when those costs are random. A heuristic approach to deal with the randomized cost is to use the threshold τ computed by the mean value of costs. Furthermore, the estimation of consumer belief state p_t or the actual state G_t is updated by the maximum a posteriori rule (see Gelman et al., 2014). After the estimation process, the retailer decides which coupon to offer based on the threshold policy given in Section 3.

2.5 Summary of Main Results

For the problems described in Subsection 2.1, 2.2, and 2.3, given all system parameters, we show the following:

- there exists an optimal stationary solution which has a single threshold property and
- the threshold only depends on the system parameters, i.e., transition probabilities and instantaneous cost associated with each type of coupon.

This means by adopting the optimal policy, the retailer will offer an HP coupon if p_t is less than some threshold and offer an LP if p_t is above the threshold.

For the model described in Subsection 2.3, we assume that cost feedbacks are noisy and consumer belief state is unknown to the retailer. For this model:

- we design a heuristic threshold policy when the received costs are noisy.
- a Bayesian estimation approach is proposed to estimate the actual state or the belief state of the consumer when the initial state is unknown to the retailer.

3 Optimal Policies with Known Consumer Statistics

In this section, we consider the basic formulation as well as the first three extensions. First, we assume that there is only one retailer and one consumer in the system and the state transition of the consumer is independent of the coupon offered. The evolution of the minimum cost function is given in (9), (10), and (11).

3.1 Properties of Minimum Cost Function

Lemma 1. *Notice that $V_\beta^{t\sim k}(p)$ is the minimum cost when the decision horizon starts from t and only spans k stages with initial belief p at time t, given a time invariant action set $u_i \in \mathcal{U} = \{\mathsf{LP}, \mathsf{HP}\}$, for any $i = 0, 1, \ldots$, $V_\beta^{t\sim k}(p) = \beta V_\beta^{t-1\sim k}(p)$.*

Proof. By (5) and $u_i \in \{\mathsf{LP}, \mathsf{HP}\}$ for any $i = 0, 1, \ldots$.

$$\begin{aligned} V_\beta^{t\sim k}(p) &= \min_\pi \mathbb{E}_\pi \left[\sum_{i=t}^{t+k-1} \beta^i C(p_i, u_i) | p_t = p \right] \\ &= \beta \min_\pi \mathbb{E}_\pi \left[\sum_{i=t-1}^{t+k-2} \beta^i C(p_i, u_i) | p_{t-1} = p \right] \\ &= \beta V_\beta^{t-1\sim k}(p). \end{aligned} \tag{14}$$

By using induction on t, we can easily prove $V_\beta^{t\sim k}(p) = \beta V_\beta^{t-1\sim k}(p) = \cdots = \beta^t V_\beta^{0\sim k}(p)$. □

Lemma 2. *The minimum cost function $V_\beta^t(p)$ is a concave and non-decreasing function of p.*

Proof. We prove these properties by induction. Remember that $V_{\beta,u_t}^{t\sim k}(p)$ is the minimum cost when the decision horizon starts from t and only spans k stages with initial belief p and initial action u_t. For $k = 1$,

$$V_\beta^{t\sim k}(p) = \min\{C_L, (1-p)C_{HN} + pC_{HA}\}, \tag{15}$$

which is a concave function of p. For $k = n - 1$, assume that $V_\beta^{t\sim k}(p)$ is a concave function. Then, for $k = n$, since $V_\beta^{t\sim n-1}(p)$ is concave and $V_{\beta,\mathsf{LP}}^{t\sim k}(p) = \beta^t C_L + V_\beta^{t+1\sim n-1}(T(p))$, by the definition of concavity and Lemma 1, we can conclude that $V_{\beta,\mathsf{LP}}^{t\sim k}(p)$ is concave. Furthermore, $V_{\beta,\mathsf{HP}}^{t\sim k}(p)$ is an affine function of p, so $V_\beta^{t\sim k}(p) = \min\{V_{\beta,\mathsf{LP}}^{t\sim k}(p), V_{\beta,\mathsf{HP}}^{t\sim k}(p)\}$ is a concave function of p. Taking $k \to \infty, V_\beta^{t\sim k}(p) \to V_\beta^t(p)$, which implies $V_\beta^t(p)$ is a concave function.

Next, we prove the non-decreasing property of the minimum cost function. For $k = 1$, as shown in Equation (15), it is a non-decreasing function of p. Assume that

$V_\beta^{t\sim k}(p)$ is a non-decreasing function for $k = n-1$. For $k = n$, Let $p_1 \geq p_2$,

$$\begin{aligned}
&V_{\beta,\mathsf{LP}}^{t\sim k}(p_1) - V_{\beta,\mathsf{LP}}^{t\sim k}(p_2) && (16)\\
&= \beta(V_\beta^{t\sim n-1}(T(p_1)) - V_\beta^{t\sim n-1}(T(p_2))) && (17)\\
&= \beta(V_\beta^{t\sim n-1}((\lambda_{A,A} - \lambda_{N,A})p_1 + \lambda_{N,A})\\
&\qquad - V_\beta^{t\sim n-1}((\lambda_{A,A} - \lambda_{N,A})p_2 + \lambda_{N,A}))) && (18)\\
&\geq 0. && (19)
\end{aligned}$$

By using the same technique, we can prove that given $p_2 - p_1 \leq 0, C_{HN} - C_{HA} \leq 0$ and $V_\beta^{t\sim k-1}(\lambda_{N,A}) - V_\beta^{t\sim k-1}(\lambda_{A,A}) \leq 0$,

$$V_{\beta,\mathsf{HP}}^{t\sim k}(p_1) - V_{\beta,\mathsf{HP}}^{t\sim k}(p_2) \geq 0. \tag{20}$$

Since $V_\beta^{t\sim k}(p_t) = \min\{V_{\beta,\mathsf{LP}}^{t\sim k}(p), V_{\beta,\mathsf{HP}}^{t\sim k}(p)\}$, it is the minimum of two non-decreasing functions. Therefore, $V_\beta^{t\sim k}(p)$ is non-decreasing. By taking $k \to \infty, V_\beta^{t\sim k}(p) \to V_\beta^t(p)$. Thus, $V_\beta^t(p)$ is a non-decreasing function. □

Lemma 3. *Let Φ_{HP} be the set of values of p_t for which offering an* HP *coupon is the optimal action at time t. Then, Φ_{HP} is a convex set.*

Proof. Since $\Phi_{\mathsf{HP}} = \{p \in [0,1], V_\beta^t(p) = V_{\beta,\mathsf{HP}}^t(p)\}$, assume that $p_t = ap_{t,1} + (1-a)p_{t,2}$ in which $p_{t,1}, p_{t,2} \in \Phi_{\mathsf{HP}}$ and $a \in [0,1]$, $V_\beta^t(p_t)$ can be written as:

$$\begin{aligned}
V_\beta^t(p_t) &= V_\beta^t(ap_{t,1} + (1-a)p_{t,2}) && (21)\\
&\geq aV_\beta^t(p_{t,1}) + (1-a)V_\beta^t(p_{t,2}) && (22)\\
&= aV_{\beta,\mathsf{HP}}^t(p_{t,1}) + (1-a)V_{\beta,\mathsf{HP}}^t(p_{t,2}) && (23)\\
&= a[(1-p_{t,1})[\beta^t C_{HN} + \beta V_\beta^t(\lambda_{N,A})] + p_{t,1}[\beta^t C_{HA} + \beta V_\beta^t(\lambda_{A,A})]]\\
&\qquad + (1-a)[(1-p_{t,2})[\beta^t C_{HN} + \beta V_\beta^t(\lambda_{N,A})] + p_{t,2}[\beta^t C_{HA} + \beta V_\beta^t(\lambda_{A,A})]] && (24)\\
&= V_{\beta,\mathsf{HP}}^t(ap_{t,1} + (1-a)p_{t,2}). && (25)
\end{aligned}$$

Thus, we have shown that:

$$V_\beta^t(p_t) \geq V_{\beta,\mathsf{HP}}^t(ap_{t,1} + (1-a)p_{t,1}) = V_{\beta,\mathsf{HP}}^t(p_t). \tag{26}$$

By the definition of $V_\beta^t(p_t)$ in (11), $V_\beta^t(p_t) \leq V_{\beta,\mathsf{HP}}^t(p_t)$. Therefore, $V_{\beta,\mathsf{HP}}^t(p_t) = V_\beta^t(p_t)$, which implies Φ_{HP} is convex. □

3.2 Optimal Stationary Policy Structure

Theorem 1. *There exists a threshold $\tau \in [0,1]$ such that the following policy is optimal:*

$$\pi^*(p_t) = \begin{cases} \mathsf{LP} & \text{if } \tau \leq p_t \leq 1 \\ \mathsf{HP} & \text{if } 0 \leq p_t \leq \tau. \end{cases} \tag{27}$$

More precisely, let $\delta \triangleq C_{HA} - C_{HN} + \beta(V_\beta(\lambda_{A,A}) - V_\beta(\lambda_{N,A}))$,

$$\tau = \begin{cases} \frac{C_L-(1-\beta)(C_{HN}+\beta V_\beta(\lambda_{N,A}))}{(1-\beta)\delta} & T(\tau) \geq \tau \\ \frac{C_L+\beta\lambda_{N,A}(C_{HA}+\beta V_\beta(\lambda_{A,A}))}{(1-(\lambda_{A,A}-\lambda_{N,A})\beta)\delta} - \frac{(1-\beta(1-\lambda_{N,A}))(C_{HN}+\beta V_\beta(\lambda_{N,A}))}{(1-(\lambda_{A,A}-\lambda_{N,A})\beta)\delta} & T(\tau) < \tau \end{cases} \tag{28}$$

where for $\lambda_{N,A} \geq \tau$,

$$V_\beta(\lambda_{N,A}) = V_\beta(\lambda_{A,A}) = C_L/(1-\beta) \tag{29}$$

and for $\lambda_{N,A} < \tau$,

$$\begin{aligned} V_\beta(\lambda_{N,A}) &= (1-\lambda_{N,A})[C_{HN} + V_\beta^1(\lambda_{N,A})] \\ &\quad + \lambda_{N,A}[C_{HA} + V_\beta^1(\lambda_{A,A})], \qquad (30) \\ V_\beta(\lambda_{A,A}) &= \min_{n\geq 0}\{G(n)\}, \qquad (31) \end{aligned}$$

where

$$G(n) = \frac{C_L\frac{1-\beta^n}{1-\beta} + \beta^n[\bar{T}^n(\lambda_{A,A})(C_{HN} + C(\lambda_{N,A})) + T^n(\lambda_{A,A})C_{HA}]}{1-\beta^{n+1}[\bar{T}^n(\lambda_{A,A})\frac{\lambda_{N,A}\beta}{1-(1-\lambda_{N,A})\beta} + T^n(\lambda_{A,A})]} \tag{32}$$

$$T^n(\lambda_{A,A}) = \frac{(\lambda_{A,A}-\lambda_{N,A})^{n+1}(1-\lambda_{A,A}) + \lambda_{N,A}}{1-(\lambda_{A,A}-\lambda_{N,A})} \tag{33}$$

$$\bar{T}^n(\lambda_{A,A}) = 1 - T^n(\lambda_{A,A}) \tag{34}$$

$$C(\lambda_{N,A}) = \beta\frac{(1-\lambda_{N,A})C_{HN} + \lambda_{N,A}C_{HA}}{1-(1-\lambda_{N,A})\beta}. \tag{35}$$

The proof of Theorem 1 is provided in the Appendix 1. An immediate consequence of this result is an upper bound on p_t for offering an HP coupon.

We define κ to be the ratio between the gain from offering an HP coupon to a Normal consumer and the loss from offering an HP coupon to a consumer whom the retailer thinks is Normal but is actually Alerted. Thus,

$$\kappa = \frac{C_L - C_{HN}}{C_{HA} - C_{HN}}. \tag{36}$$

For fixed costs, the threshold can be bounded by the following two Corollaries.

Corollary 1. *If $p_t \leq \kappa$, then it is optimal for the retailer to offer an* HP *coupon.*

Corollary 2. *Fix coupon offering costs and $\lambda_{A,A}$, let $\lambda_1 = \frac{C_L-C_{HN}}{C_{HA}-C_{HN}}$ and λ_2 be the solution of $\frac{\lambda_2}{1-(\lambda_{A,A}-\lambda_2)} = \frac{\beta(C_L-C_{HA})\lambda_2+C_L-C_{HN}}{(1-\beta)C_{HA}-C_{HN}+\beta C_L}$. When $\lambda_{N,A} \geq \lambda_2$, the threshold τ in the optimal stationary policy can be written as a closed form expression with respect to $\lambda_{N,A}$: if $\lambda_{N,A} > \lambda_1$,*

$$\tau = \kappa; \tag{37}$$

if $\lambda_2 < \lambda_{N,A} < \lambda_1$,

$$\tau = \frac{\beta(C_L - C_{HA})\lambda_{N,A} + C_L - C_{HN}}{(1-\beta)C_{HA} - C_{HN} + \beta C_L}. \tag{38}$$

Moreover, if $\lambda_{N,A} < \lambda_2$, τ *can be upperbounded by*

$$\bar{\tau} = \frac{\lambda_2}{1 - (\lambda_{A,A} - \lambda_2)}. \tag{39}$$

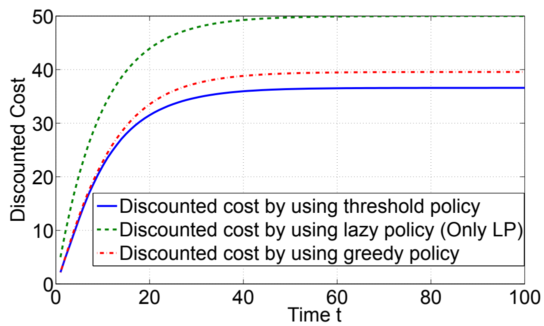
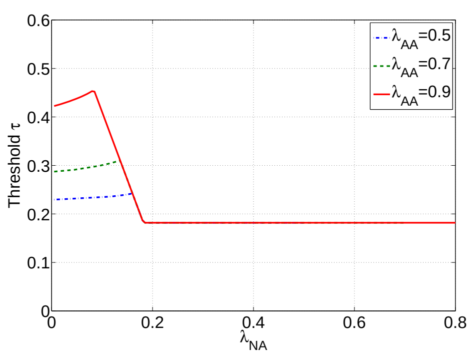
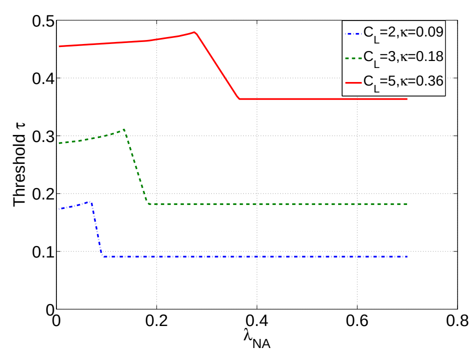


Figure 3: Discounted cost resulted by using different decision policies.

A detailed proof of Corollary 1 and 2 are presented in Appendix 2 and 3, respectively. To illustrate the performance of the proposed threshold policy, we compare the



(a) Threshold τ vs. $\lambda_{N,A}$. (Parameters: $\beta = 0.9, C_L = 3, C_{HN} = 1, C_{HA} = 12, \kappa = 0.18$.)



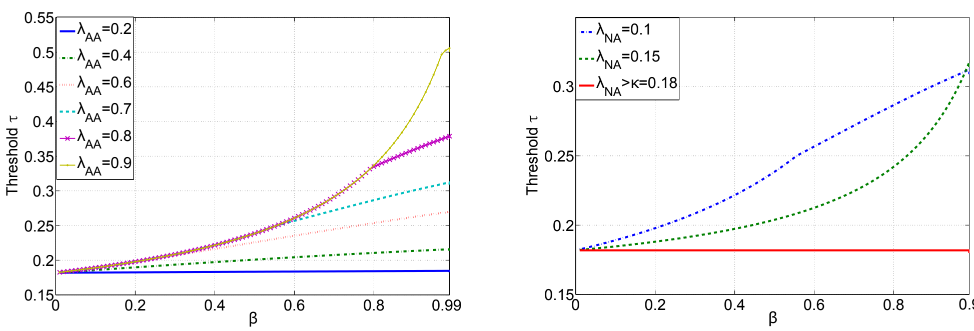
(b) Threshold τ vs. $\lambda_{N,A}$. (Parameters: $\lambda_{A,A} = 0.7, \beta = 0.9, C_{HN} = 1, C_{HA} = 12$.)

Figure 4: Threshold τ vs. β for different values of $\lambda_{A,A}$ and $\lambda_{N,A}$

discounted cost resulted from the threshold policy with the greedy policy which minimizes the instantaneous cost at each decision epoch as well as with a lazy policy in which a retailer only offers LP coupons. We plot the discounted cost averaged over

1000 independent MDPs w.r.t. time t for different decision policies in Figure 3. The illustration demonstrates that the proposed threshold policy performs better than the greedy policy and the lazy policy.

Figure 4a shows the optimal threshold policy with respect to $\lambda_{N,A}$ for three fixed choices of $\lambda_{A,A}$. It can be seen that the threshold is increasing when $\lambda_{N,A}$ is small; this is because for a small $\lambda_{N,A}$ the consumeris less likely to transition from Normal to Alerted. Therefore, the retailer tends to offer an HP coupon to the consumer. When $\lambda_{N,A}$ gets larger, the consumer is more likely to transition from Normal to Alerted. Thus, the retailer tends to play conservatively by decreasing the threshold for offering an LP coupon. When $\lambda_{N,A}$ is greater than κ, the retailer will just use κ to be the threshold for offering an HP coupon. One can also observe that with increasing $\lambda_{A,A}$, the threshold τ decreases. On the other hand, for fixed C_{HN} and C_{HA}, Figure 4b shows that the threshold τ increases as the cost of offering an LP coupon increases, making it more desirable to take a risk and offer an HP coupon.



(a) Threshold τ vs. β for different values of $\lambda_{A,A}$ (Parameters: $\lambda_{N,A} = 0.1, C_L = 3, C_{HN} = 1, C_{HA} = 12, \kappa = 0.18$.)

(b) Threshold τ vs. β for different values of $\lambda_{N,A}$ (Parameters: $\lambda_{A,A} = 0.7, C_L = 3, C_{HN} = 1, C_{HA} = 12$.)

Figure 5: Threshold τ vs. β for different values of $\lambda_{A,A}$ and $\lambda_{N,A}$.

The relationship between the discount factor β and the threshold τ as functions of transition probabilities is shown in Figure 5. It can be seen in Figure 5a that the threshold increases as β increases. This is because when β is small, the retailer values the present rewards more than future rewards. Therefore, the retailer tends to play conservatively so that it will not "creep out" the consumer in the present. Figure 5b shows that the threshold is high when $\lambda_{A,A}$ is large or $\lambda_{N,A}$ is small. A high $\lambda_{A,A}$ value indicates that a consumer is more likely to remain in Alerted state. The retailer is willing to play aggressively since once the consumer is in alerted state, it can take a very long time to transition back to Normal state. A low $\lambda_{N,A}$ value implies that the consumer is not very privacy sensitive. Thus, the retailer tends to offer HP coupons to reduce cost. One can also observe in Figure 5b that the threshold τ equals to κ after $\lambda_{N,A}$ exceeds the ratio κ. This is consistent with results shown in Figure 4.

The effect of an LP coupon cost on the threshold for different discount factors is

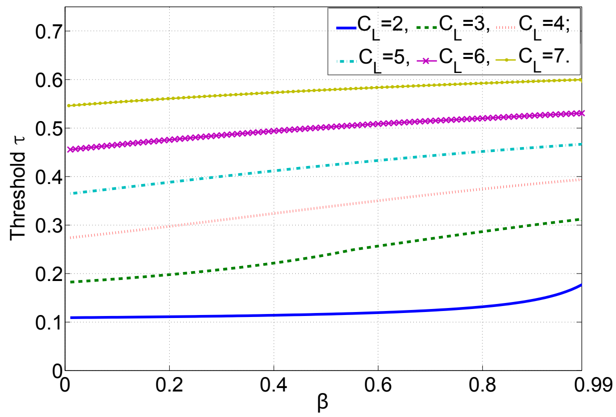


Figure 6: Threshold τ vs. β for different values of C_L. (Parameters: $\lambda_{N,A} = 0.1, \lambda_{A,A} = 0.9, C_{HN} = 1, C_{HA} = 12.$).

plotted in Figure 6. It can be seen that a higher C_L will increase the threshold because the retailer is more likely to offer an HP coupon when the cost of offering an LP coupon is high.

3.3 Consumer with Multi-Level Alerted States

In this section, we study the case that the consumer has multiple Alerted states. Without loss of generality, we define the transition matrix to be

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_{N,N} & \lambda_{N,A_1} & \dots & \lambda_{N,A_K} \\ \lambda_{A_1,N} & \lambda_{A_1,A_1} & \dots & \lambda_{A_1,A_K} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{A_K,N} & \lambda_{A_K,A_1} & \dots & \lambda_{A_K,A_K} \end{pmatrix} \tag{40}$$

and $\bar{\mathbf{e}}_i$ to be the i^{th} row of $\mathbf{\Lambda}$. The expected cost at time t, given belief $\bar{\mathbf{p}}_t$ and action u_t, has the following expression:

$$C(\bar{\mathbf{p}}_t, u_t) = \begin{cases} C_L & \text{if } u_t = \mathsf{LP} \\ \bar{\mathbf{p}}_t^T \bar{\mathbf{C}} & \text{if } u_t = \mathsf{HP} \end{cases} . \tag{41}$$

Assuming that the retailer has perfect information about the belief states, the cost function evolves as follows. By using an LP coupon at time t,

$$V_{\beta,\mathsf{LP}}^t(\bar{\mathbf{p}}_t) = \beta^t C_L + V_\beta^{t+1}(\bar{\mathbf{p}}_{t+1}) = \beta^t C_L + V_\beta^{t+1}(T(\bar{\mathbf{p}}_t)), \tag{42}$$

where $T(\bar{\mathbf{p}}_t) = \bar{\mathbf{p}}_t^T \mathbf{\Lambda}$ is the Markov transition operator generalizing (2). By using an

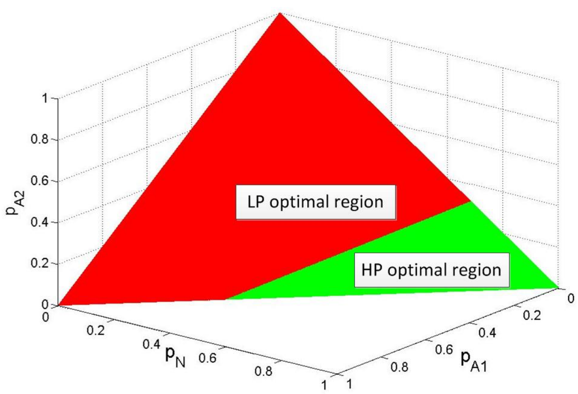


Figure 7: Example of the optimal policy region for three-state consumer. (Parameters: $\lambda_{\mathsf{N,N}} = 0.7, \lambda_{\mathsf{N},A1} = 0.2, \lambda_{\mathsf{N,A2}} = 0.1; \lambda_{\mathsf{A1,N}} = 0.2, \lambda_{\mathsf{A1,A1}} = 0.5, \lambda_{\mathsf{A1,A2}} = 0.3; \lambda_{\mathsf{A2,N}} = 0.1, \lambda_{\mathsf{A2,A1}} = 0.2, \lambda_{\mathsf{A2,A2}} = 0.7; \beta = 0.9, C_{\mathsf{L}} = 7, C_{\mathsf{HN}} = 1, C_{\mathsf{HA1}} = 10, C_{\mathsf{HA2}} = 20$).

HP coupon at time t,

$$V^t_{\beta,\mathsf{HP}}(\bar{\mathbf{p}}_t) = \beta^t \bar{\mathbf{p}}_t^T \bar{\mathbf{C}} + V_\beta^{t+1}(\bar{\mathbf{p}}_{t+1}) = \beta^t \bar{\mathbf{p}}_t^T \bar{\mathbf{C}} + \bar{\mathbf{p}}_t^T \begin{pmatrix} V_\beta^{t+1}(\bar{\mathbf{e}}_1) \\ V_\beta^{t+1}(\bar{\mathbf{e}}_2) \\ \vdots \\ V_\beta^{t+1}(\bar{\mathbf{e}}_{K+1}) \end{pmatrix}. \tag{43}$$

Therefore, by (11), we have $V^t_\beta(\bar{\mathbf{p}}_t) = \min\{V^t_{\beta,\mathsf{LP}}(\bar{\mathbf{p}}_t), V^t_{\beta,\mathsf{HP}}(\bar{\mathbf{p}}_t)\}$.

In this problem, since the instantaneous costs are nondecreasing with the state when the action is fixed and the evolution of belief state is the same for both LP and HP, the existence of an optimal stationary policy with threshold property is guaranteed by Proposition 2 in Lovejoy (1987). The optimal stationary policy for a three-state consumer model is illustrated in Figure 7. For fixed costs, the plot shows the partition of the belief space based on the optimal actions and reveals that offering an HP coupon is optimal when $p_{N,t}$, the belief of the consumer being in Normal state, is high.

4 Consumers with Coupon Dependent Transitions

Generally, consumers' reaction to HP and LP coupons are different. To be more specific, a consumer is likely to feel less comfortable when being offered a coupon on medication (HP) than food (LP). Thus, we assume that the Markov transition probabilities are dependent on the coupon offered. Let p_t denote the belief of a consumer being in the Alerted state at time t.

As shown in Figure 2, by offering an LP coupon, the state transition follows the

Markov chain

$$\mathbf{\Lambda}_{\mathsf{LP}} = \begin{pmatrix} 1-\lambda_{N,A} & \lambda_{N,A} \\ 1-\lambda_{A,A} & \lambda_{A,A} \end{pmatrix}. \tag{44}$$

Otherwise, the state transition follows

$$\mathbf{\Lambda}_{\mathsf{HP}} = \begin{pmatrix} 1-\lambda'_{N,A} & \lambda'_{N,A} \\ 1-\lambda'_{A,A} & \lambda'_{A,A} \end{pmatrix}. \tag{45}$$

According to the model in Section 2, $\lambda_{A,A} > \lambda_{N,A}, \lambda'_{A,A} > \lambda'_{N,A}$. Moreover, we assume that offering an HP coupon will increase the probability of transition to or staying at Alerted state. Therefore, $\lambda'_{A,A} > \lambda_{A,A}$ and $\lambda'_{N,A} > \lambda_{N,A}$. The minimum cost function evolves as follows: for an HP coupon offered at time t, we have

$$V^t_{\beta,\mathsf{HP}}(p_t) = \beta^t C(p_t, \mathsf{HP}) + (1-p_t)V^{t+1}_\beta(\lambda'_{N,A}) + p_t V^{t+1}_\beta(\lambda'_{A,A}).$$

Otherwise,

$$V^t_{\beta,\mathsf{LP}}(p_t) = \beta^t C_L + V^{t+1}_\beta(p_{t+1}) = \beta^t C_L + V^{t+1}_\beta(T(p_t)),$$

where $T(p_t) = \lambda_{N,A}(1-p_t) + \lambda_{A,A}p_t$ is the one step transition defined in Section 2. In this case, the transition probability is just a deterministic function of the retailer action. Thus, finding an optimal strategy to this problem is equivalent to solving an associated MDP problem in belief space. Furthermore, Theorem 6.3 and its generalization in Ross (1992) still hold since the transition probability is a function of the action. Therefore, there exists an optimal stationary policy π^* in the belief space which minimizes the infinite horizon discounted cost.

Theorem 2. *Given action dependent transition matrices $\mathbf{\Lambda}_{\mathsf{LP}}$ and $\mathbf{\Lambda}_{\mathsf{HP}}$, the optimal stationary policy has threshold structure.*

The proof of Theorem 2 is provided in Appendix 4.

Figure 8 shows the effect of costs on the threshold τ. We can see that for a fixed C_L and C_{HA} pair, the threshold for LP coupons for consumers in this model is lower than our original model without coupon-dependent transition probabilities. The retailer can only offer an LP coupon with certain combination of costs; we call this the LP-only region. One can also see that the LP-only region for the coupon-independent transition case is smaller than that for the coupon-dependent transition case since for the latter, the likelihood of being in an Alerted state is higher for the same costs.

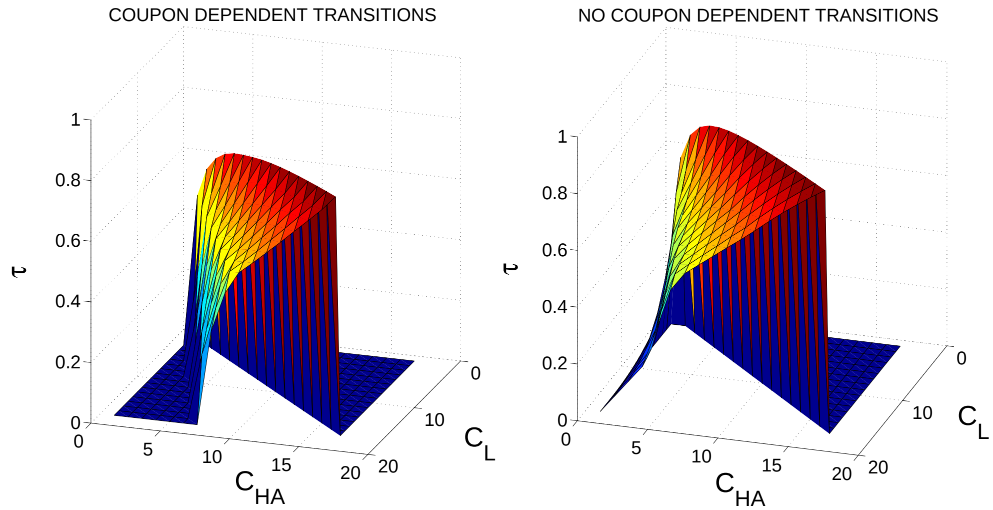


Figure 8: Optimal policy threshold for consumer with/without coupon dependent transition probabilities. (Parameters: $\lambda_{N,A} = 0.2, \lambda_{A,A} = 0.8, \lambda'_{N,A} = 0.5, \lambda'_{A,A} = 0.9, \beta = 0.9$).

5 Policies under Noisy Cost Feedback and Uncertain Initial Belief

In this section, we study the case in which the received costs are random. In the previous sections, if the retailer offered an HP coupon at time t, then it could learn the state of the consumer at time t based on whether the received cost was C_{HN} or C_{HA}. If the cost feedback is random, then the retailer may not be able to infer the consumer's state exactly. We describe policy heuristics for this setting that perform Bayesian estimation of the quantity p_t used in the threshold policy earlier. This approach is also useful when the initial value p_0 is not known to the retailer.

We model the noisy cost feedback by assuming the received cost C_t is random. The distribution of C_t is given by a conditional probability density $f(c|G_t, u_t)$ on a bounded subset of $\mathbb{R}$, where G_t is the state of the consumer and u_t is the action taken by the retailer at time t. To match the previous model, we further take $f(c|G_t = \mathsf{Alerted}, u_t = \mathsf{LP}) = f(c|G_t = \mathsf{Normal}, u_t = \mathsf{LP})$ to indicate that the received cost conveys no information about the state under an LP coupon. Let $f(c|u_t = \mathsf{LP}) = f(c|G_t = \mathsf{Alerted}, u_t = \mathsf{LP})$. For a given value $p_t = p$, define the likelihood of observing a cost $C_t = c$ under the two coupons:

$$\ell(c|\mathsf{LP}, p) = f(c|\mathsf{Alerted}, \mathsf{LP}) \tag{46}$$

$$\ell(c|\mathsf{HP}, p) = f(c|\mathsf{Normal}, \mathsf{HP})(1-p) + f(c|\mathsf{Alerted}, \mathsf{HP})p. \tag{47}$$

These likelihoods will be useful in defining the two estimators.

In both approaches in this section the retailer computes an estimate $\hat{p}_t$ of the probability p_t that $G_t = \mathsf{Alerted}$. It then uses (27) to decide which coupon to offer at time t by comparing $\hat{p}_t$ to a version of the threshold in (28). Define $\mathcal{C_L}, \mathcal{C_{HN}}$, and $\mathcal{C_{HA}}$ to be the feasible cost sets $\{c : f(c|\mathsf{LP}) > 0\}$, $\{c : f(c|\mathsf{Alerted}, \mathsf{HP}) > 0\}$, and $\{c : f(c|\mathsf{Normal}, \mathsf{HP}) > 0\}$, respectively. Since τ involves costs C_L, C_{HN} and C_{HA}, there are several ways to compute an approximate threshold under the cost uncertainty.

Firstly, we can set C_L, C_{HN} and C_{HA} to be the expected costs:

$$C_L = \int_{\mathbb{R}} c f(c|\mathsf{LP}) dc \tag{48}$$

$$C_{HN} = \int_{\mathbb{R}} c f(c|\mathsf{Normal}, \mathsf{HP}) dc \tag{49}$$

$$C_{HA} = \int_{\mathbb{R}} c f(c|\mathsf{Alerted}, \mathsf{HP}) dc. \tag{50}$$

Plugging these into (28) gives the mean threshold τ_{avg}. Since τ is monotonically increasing in C_L and C_{HA} and monotonically decreasing in C_{HN}, we can compute and upper bound on τ by setting $C_L = \max\{c : c \in \mathcal{C_L}\}$, $C_{HA} = \max\{c : c \in \mathcal{C_{HA}}\}$, and $C_{HN} = \max\{c : c \in \mathcal{C_{HN}}\}$. These values give the upper bound threshold τ_{max}. Similarly, by setting C_L and C_{HA} to the lower bounds on the support and C_{HN} to the upper bound, we obtain a lower bound threshold τ_{min}. Finally, we computed a robust version of threshold τ_{R} as $\tau_{\mathsf{R}} = \{\tau : \max_{C_L, C_{HN}, C_{HA}} \{\min_{\pi(p_t)} V_\beta^t(p_t)\}\}$, where $(C_L, C_{HN}, C_{HA}) \in \mathcal{C_L} \times \mathcal{C_{HN}} \times \mathcal{C_{HA}}$, is the This threshold policy is the largest (cost case) threshold over all possible combination of costs. Thus, it gives the $max - min$ value of the total discounted cost. We can see that the total discounted cost induced by this robust version of threshold is close to that induced by using the upper bound of costs.

5.1 Estimation of the Consumer State

In the previous model, if $u_t = \mathsf{HP}$ the retailer could infer G_t based on C_t, so p_{t+1} is given by the state transitions of the Markov chain. With noisy costs this exact inference is no longer possible. A simple heuristic for the retailer is to try to infer G_t based on the random cost C_t, compute an estimate of p_t, and then use the previous strategy.

At time $t = 1$, given an initial p_0 we estimate $\hat{p}_1 = T(p_0)$. The retailer then applies the threshold policy (27) with input $\hat{p}_1$ to offer a coupon. For times $t = 2, 3, \ldots$ the retailer treats the estimate $\hat{p}_{t-1}$ as an estimate of the probability that $G_{t-1} = \mathsf{Alerted}$. If $u_{t-1} = \mathsf{LP}$, then the retailer sets $\hat{p}_t = T(\hat{p}_{t-1})$. If $u_{t-1} = \mathsf{HP}$ then the retailer uses a maximum a posteriori probability (MAP) detection rule to estimate the state G_{t-1} based on the received cost C_{t-1}. That is, it sets $\hat{G}_{t-1} = \mathsf{Normal}$ if

$$\frac{f(C_{t-1}|\mathsf{Normal}, \mathsf{HP})(1 - \hat{p}_{t-1})}{f(C_{t-1}|\mathsf{Alerted}, \mathsf{HP})\hat{p}_{t-1}} > 1 \tag{51}$$

and $\hat{G}_{t-1} = \mathsf{Alerted}$ otherwise, where C_{t-1} is the received cost at time $t - 1$. It then

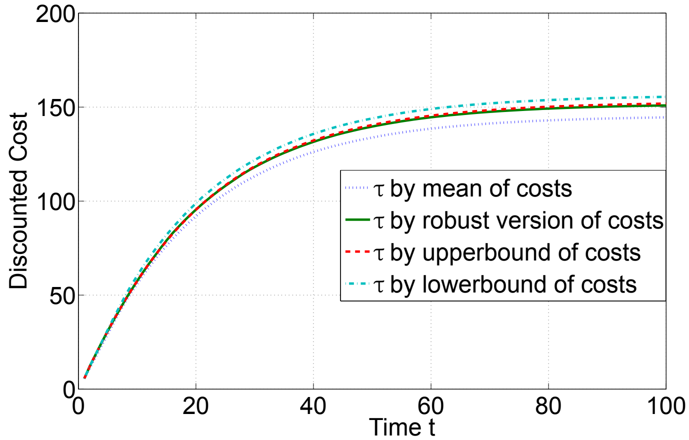


Figure 9: Temporal discounted costs for different heuristics on computing thresholds. (Parameters: $\lambda_{N,A} = 0.2$, $\lambda_{A,A} = 0.8$, $p_0 = 0.2, \beta = 0.95$, $f(c|\mathsf{LP}) = \mathsf{Unif}[6, 10]$, $f(c|\mathsf{Normal}, \mathsf{HP}) = \mathsf{Unif}[0.2, 5.8]$, and $f(c|\mathsf{Alerted}, \mathsf{HP}) = \mathsf{Unif}[12, 20]$). The discounted cost is averaged over 1000 independent runs.

uses the following estimate p_t at time t:

$$\hat{p}_t = \begin{cases} \lambda_{N,A} & \text{if } \hat{G}_t = \mathsf{Normal} \\ \lambda_{A,A} & \text{if } \hat{G}_t = \mathsf{Alerted}. \end{cases} \tag{52}$$

Essentially, the retailer uses MAP estimation to infer G_{t-1} after receiving the cost C_{t-1} from the action $u_{t-1} = \mathsf{HP}$. If the densities $f(c|\mathsf{Normal}, \mathsf{HP})$ and $f(c|\mathsf{Alerted}, \mathsf{HP})$ have disjoint supports, then the inference of G_{t-1} is error free, so $\hat{G}_{t-1} = G_{t-1}$ and the estimate $\hat{p}_t$ is correct. Figure 9 shows the discounted cost as a function of time for some different variants of the threshold in (28). In this example the cost distributions are uniformly distributed in disjoint intervals. The plot shows that the mean threshold yields a total discounted cost that is slightly less than the upper and lower bound thresholds.

5.2 Bayesian Estimation of State Probabilities

In the previous approach, the retailer estimates the underlying state and then uses this to form an estimate of the probability p_t that $G_t = \mathsf{Alerted}$. A different approach is to form a Bayes estimate of p_t: the retailer computes a probability distribution on $[0, 1]$ representing its uncertainty about p_t. To choose an action u_t it can use a point estimate of p_t to use in (27) with one of the thresholds described before.

In this formulation, the estimator of p_t is a probability distribution. Let $q_{t-1}(p)$ be the estimator of p_{t-1}. The retailer treats this as a prior distribution. Upon receiving the cost C_{t-1} it computes a posterior estimate on p_{t-1} using Bayes rule. If $u_{t-1} = \mathsf{HP}$, it sets

$$q_{t-1}(p|C_{t-1}) = \frac{\ell(C_{t-1}|\mathsf{HP}, p)q_{t-1}(p)}{\int_0^1 \ell(C_{t-1}|\mathsf{HP}, p')q_{t-1}(p')dp'}. \tag{53}$$

If $u_{t-1} = \mathsf{LP}$ then from (46) we can see that $\ell(C_{t-1}|\mathsf{LP}, p)$ does not depend on p, so the posterior $q_{t-1}(p|C_{t-1}) = q_{t-1}(p)$ in this case. Given the posterior estimate $q_{t-1}(p|C_{t-1})$ the retailer then evolves the state distribution through the Markov chain governing the state to form the prior distribution $q_t(p)$ for estimating p_t at time t. That is, if P_{t-1} is a random variable with distribution $q_{t-1}(p|C_{t-1})$, then $q_t(p)$ is the distribution of $T(P_{t-1})$. Let $Q_{t-1}(p|C_{t-1}) = \int_0^p q_{t-1}(p'|C_{t-1})$ be the cumulative distribution function of P_{t-1}. Then

$$\mathbb{P}\left(T(P_{t-1}) \leq p\right) = \mathbb{P}\left(P_{t-1} \leq \frac{p - \lambda_{N,A}}{\lambda_{A,A} - \lambda_{N,A}}\right) = Q_{t-1}\left(\frac{p - \lambda_{N,A}}{\lambda_{A,A} - \lambda_{N,A}}\big|C_{t-1}\right) \tag{54}$$

so

$$q_t(p) = \frac{1}{\lambda_{A,A} - \lambda_{N,A}} q_{t-1}\left(\frac{p - \lambda_{N,A}}{\lambda_{A,A} - \lambda_{N,A}}\big|C_{t-1}\right). \tag{55}$$

The retailer then uses $q_t(p)$ to form a point estimate $\hat{p}_t$ of p_t suitable for applying the threshold policy in (27) and (28). We consider two such point estimates which we call the mean and max estimators, respectively:

$$\hat{p}_{t,\mathsf{mean}} = \int_0^1 p q_t(p) dp \tag{56}$$

$$\hat{p}_{t,\mathsf{MAP}} = \underset{p \in [0,1]}{\operatorname{argmax}}\, q_t(p). \tag{57}$$

Figure 10 shows the discounted cost versus time for uniformly distributed costs with overlapping support. The decision is made by following the optimal stationary policy computed by the mean threshold in Section 5. We illustrate the result for four algorithms: the solid curve and the dash-dot curve are the MAP and mean strategy described above, respectively; the dashed curve is a policy in which costs are random but the algorithm is given side information about G_t after choosing $u_t = \mathsf{HP}$ (perfect state information); finally, the curve with cross is the MAP estimate of actual state G_t described in Section 5.1. In this example, as one can expect, decision making with perfect state information has the minimum discounted cost. MAP estimation of G_t results in an 0.82% increase in total discounted cost compared to the case in which the retailer receives perfect information about consumer state. However, the MAP and mean policy to estimate belief state p_t only have 2.9% and 4.29% increases, respectively. Thus, the MAP for estimating belief perfoms slightly better than the Mean policy. Effectively, the lack of initial belief knowledge does not affect the discouted cost very much on average. This is because offering an HP coupon allows the retailer to learn the actual state from the cost feedback, and thus, reset the belief state.

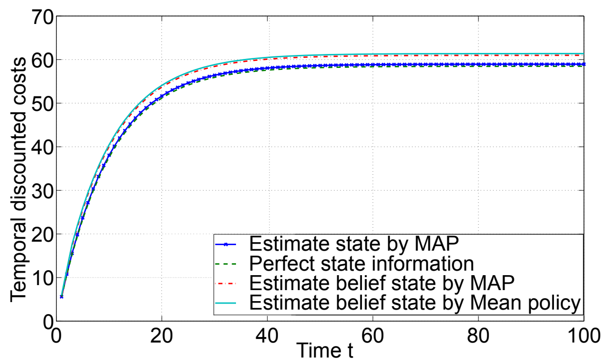


Figure 10: Temporal discounted costs for different estimation mechanisms. (Parameters: $\lambda_{N,A} = 0.2, \lambda_{A,A} = 0.8$, $p_0 = 0.2, \beta = 0.9$, $f(c|\mathsf{LP}) = \mathsf{Unif}[3,9]$, $f(c|\mathsf{Normal}, \mathsf{HP}) = \mathsf{Unif}[0.25, 7.75]$, $f(c|\mathsf{Alerted}, \mathsf{HP}) = \mathsf{Unif}[6, 18]$). The discounted cost is averaged over 1000 independent runs.

6 Conclusions

We proposed a POMDP model to capture the interactions between a retailer and a privacy-sensitive consumer in the context of personalized shopping. The retailer seeks to minimize the expected discounted cost of violating the consumer's privacy. We showed that the optimal coupon-offering policy is a stationary policy that takes the form of an explicit threshold that depends on the model parameters. In summary, the retailer offers an HP coupon when the Normal to Alerted transition probability is low or the probability of staying in Alerted state is high. Furthermore, the threshold optimal policy also holds for consumers whose privacy sensitivity can be captured via multiple alerted states as well as for the case in which consumers exhibit coupon-dependent transition. For the case in which cost feedbacks from the consumer are noisy, we have introduced a heuristic method using the mean value of costs to compute the decision threshold. Furthermore, under noisy cost feedbacks scenario, we have introduced a Bayesian data analysis approach for decision making which includes estimating consumer belief state when the initial belief state is unknown to the retailer.

To the best of our knowledge, our proposed model is the first model to capture the subjective experience and heterogeneous expression of privacy sensitivity in designing incentive schemes for privacy-sensitive consumers. This is just a first step, and our model has several important limitations. First of all, we only consider the problem from the retailer's perspective. In reality, a consumer may be strategic and either accept or refuse to use the coupon in order to maximize its reward. Secondly, our model has a significant number of parameters that must be estimated by a retailer. In particular, the transition probabilities must be estimated through population surveys or prior data. Finally, we consider a single consumer and single retailer in this work, which can only capture a limited section of the rich space of possible interactions between populations.

Our work suggests several interesting future directions: one straightfoward extension of our work is to model uncertainties in the statistical model for the consumer transition probabilities. Extending our framework to competitive settings with strategic users and multiple retailers could lead to interesting insights into the potential benefits of respecting consumer privacy. This might involve developing game theoretic models. Finally, it would be interesting to develop experimental paradigms to test our model and derive new modifications.

1 Proof of Theorem 1

Proof. Let p_F be the stationary distribution of the Markov transition. Then $p_F = \lambda_{A,A}p_F + (1-p_F)\lambda_{N,A}$, which implies $p_F = \frac{\lambda_{N,A}}{1-\lambda_{A,A}+\lambda_{N,A}}$. If $V^t_{\beta,\mathsf{LP}}(p_t) > V^t_{\beta,\mathsf{HP}}(p_t)$ Remember that the threshold is the solution to $V^t_{\beta,\mathsf{LP}}(p_t) = V^t_{\beta,\mathsf{HP}}(p_t)$. Let τ be the threshold value, we have:

$$\begin{aligned}&\beta^t C_L + V^{t+1}_\beta(T(\tau))\\&= (1-\tau)[\beta^t C_{HN} + V^{t+1}_\beta(\lambda_{N,A})] + \tau[\beta^t C_{HA} + V^{t+1}_\beta(\lambda_{A,A})].\end{aligned} \tag{58}$$

By the definition of $V^t_\beta(p_t)$, we know that $V^t_\beta(p_t) = \beta^t V_\beta(p_t)$. Thus $V^t_\beta(\lambda_{N,A}) = \beta^t V_\beta(\lambda_{N,A})$ and $V^t_\beta(\lambda_{A,A}) = \beta^t V_\beta(\lambda_{A,A})$.

If $T(\tau) \geq \tau$, which is equivalent to $p_F \geq \tau$, then $V^{t+1}_\beta(T(\tau)) = V^{t+1}_{\beta,\mathsf{LP}}(T(\tau))$. Therefore, $V^t_{\beta,\mathsf{LP}}(\tau) = \lim_{n\to\infty}\{\beta^t\frac{1-\beta^n}{1-\beta}C_L + \beta^n V^{t+1}_\beta(T^n(\tau))\}$ where $T^n(\tau) = T(T^{n-1}(\tau)) = p_F(1-(\lambda_{A,A}-\lambda_{N,A})^n) + (\lambda_{A,A}-\lambda_{N,A})^n\tau$. Taking $n\to\infty$, we have $V^t_{\beta,\mathsf{LP}}(\tau) = \beta^t\frac{C}{1-\beta}$. Substitute this into (58) yields:

$$\frac{C_L}{1-\beta} = (1-\tau)C_{HN} + \tau C_{HA} + \beta(\tau V_\beta(\lambda_{A,A}) + (1-\tau)V_\beta(\lambda_{N,A})). \tag{59}$$

By rearranging terms in the above expression, we have

$$\tau = \frac{\frac{C_L}{1-\beta} - C_{HN} - \beta V_\beta(\lambda_{N,A})}{(C_{HA}-C_{HN}) + \beta(V_\beta(\lambda_{A,A}) - V_\beta(\lambda_{N,A}))}. \tag{60}$$

If $p_F \leq \tau$, then $T(\tau)\leq\tau$. Therefore $V^{t+1}_\beta(T(\tau)) = V^{t+1}_{\beta,\mathsf{HP}}(T(\tau))$, which implies

$$V^t_{\beta,\mathsf{LP}}(\tau) = \beta^t C_L + V^{t+1}_\beta(T(\tau)) = \beta^t C_L + V^{t+1}_{\beta,\mathsf{HP}}(T(\tau)) = V^t_{\beta,\mathsf{HP}}(\tau). \tag{61}$$

In this case,

$$C_L + \beta V_{\beta,\mathsf{HP}}(T(\tau)) = V_{\beta,\mathsf{HP}}(\tau). \tag{62}$$

Substitute (1) and (9) into (62), we have

$$\begin{aligned}\tau =& \frac{C_L - (1-\beta(1-\lambda_{N,A}))(C_{HN}+\beta V_\beta(\lambda_{N,A}))}{(1-(\lambda_{A,A}-\lambda_{N,A})\beta)(C_{HA}-C_{HN}+\beta(V_\beta(\lambda_{A,A})-V(\lambda_{N,A})))}\\&+\frac{\beta\lambda_{N,A}(C_{HA}+\beta V_\beta(\lambda_{A,A}))}{(1-(\lambda_{A,A}-\lambda_{N,A})\beta)(C_{HA}-C_{HN}+\beta(V_\beta(\lambda_{A,A})-V(\lambda_{N,A})))}.\end{aligned} \tag{63}$$

Next, we present how to compute $V_\beta(\lambda_{N,A})$ and $V_\beta(\lambda_{A,A})$.

Case 1: If $\lambda_{N,A} \geq \tau$, then by Model Assumption 2, $\lambda_{A,A} \geq \lambda_{N,A} \geq \tau$ and $p_F \geq \lambda_{N,A} \geq \tau$. Thus, both $\lambda_{A,A}$ and $\lambda_{N,A}$ are in Φ_{LP}, therefore,

$$V_\beta(\lambda_{N,A}) = V_\beta(\lambda_{A,A}) = \frac{C_L}{1-\beta}. \tag{64}$$

Case 2: If $\lambda_{N,A} \leq \tau$, we have $V_\beta(\lambda_{N,A}) = V_{\beta,\mathsf{HP}}(\lambda_{N,A})$. Therefore,

$$V_\beta(\lambda_{N,A}) = (1-\lambda_{N,A})[C_{HN} + V_\beta^1(\lambda_{N,A})] + \lambda_{N,A}[C_{HA} + V_\beta^1(\lambda_{A,A})]. \tag{65}$$

$$V_\beta(\lambda_{A,A}) = \min_{u_t \in \{\mathsf{HP},\mathsf{LP}\}} V_{\beta,u_t}(\lambda_{A,A}) \tag{66}$$

$$= \min\{C_L + V_\beta^1(T(\lambda_{A,A})), V_{\mathsf{HP}}(\lambda_{A,A})\} \tag{67}$$

$$= \min\{C_L \frac{1-\beta^N}{1-\beta}, \min_{0\leq n\leq N-1}\{C_L \frac{1-\beta^n}{1-\beta} + V_{\beta,\mathsf{HP}}^n(T^n(\lambda_{A,A}))\}\}. \tag{68}$$

Since $N \to \infty$ and $0 \leq \beta \leq 1$,

$$V_\beta(\lambda_{A,A}) = \min_{n>0}\{C_L \frac{1-\beta^n}{1-\beta} + \beta^n V_{\beta,\mathsf{HP}}(T^n(\lambda_{A,A}))\}. \tag{69}$$

we have:

$$V_\beta(\lambda_{A,A}) = \min_{n\geq 0}\{\frac{C_L \frac{1-\beta^n}{1-\beta} + \beta^n[\bar{T}^n(\lambda_{A,A})(C_{HN} + C(\lambda_{N,A})) + T^n(\lambda_{A,A})C_{HA}]}{1 - \beta^{n+1}[\bar{T}^n(\lambda_{A,A})\frac{\lambda_{N,A}\beta}{1-(1-\lambda_{N,A})\beta} + T^n(\lambda_{A,A})]}\}. \tag{70}$$

where

$$T^n(\lambda_{A,A}) = T(T^{n-1}(\lambda_{A,A})) = \frac{(\lambda_{A,A} - \lambda_{N,A})^{n+1}(1-\lambda_{A,A}) + \lambda_{N,A}}{1 - (\lambda_{A,A} - \lambda_{N,A})}, \tag{71}$$

$$\bar{T}^n(\lambda_{A,A}) = 1 - T^n(\lambda_{A,A}) \tag{72}$$

$$C(\lambda_{N,A}) = \beta\frac{(1-\lambda_{N,A})C_{HN} + \lambda_{N,A}C_{HA}}{1-(1-\lambda_{N,A})\beta}. \tag{73}$$

Next, we prove the uniqueness of τ. Noticing that $V_{\beta,\mathsf{LP}}^t(p)$ is a concave and non-decreasing function of p and $V_{\beta,\mathsf{HP}}^t(p)$ is an affine and non-decreasing function of p (see Lemma 2). Thus, both $V_{\beta,\mathsf{LP}}^t(p)$ and $V_{\beta,\mathsf{HP}}^t(p)$ are continuous functions (every concave/affine function is continuous). Furthermore, if $p = 0$, the optimal action will be offering HP since the retailer is sure that the state of consumer is Normal and $C_{HN} < C_L$. This implies

$$V_{\beta,\mathsf{LP}}^t(p=0) > V_{\beta,\mathsf{HP}}^t(p=0). \tag{74}$$

Likewise, when $p = 1$, the optimal action will be offering LP since the retailer is sure that the state of consumer is $\mathsf{Alerted}$ and $C_L < C_{HA}$. Thus, we have

$$V^t_{\beta,\mathsf{LP}}(p=1) < V^t_{\beta,\mathsf{HP}}(p=1). \tag{75}$$

Thus, no action is uniformly better than the other in this model. Therefore, by (74),(75) and continuity and concavity of $V^t_{\beta,\mathsf{LP}}(p)$ and $V^t_{\beta,\mathsf{HP}}(p)$, there is a unique solution to $V^t_{\beta,\mathsf{LP}}(p) = V^t_{\beta,\mathsf{HP}}(p)$ for $p \in [0,1]$. □

2 Proof of Corollary 1

Proof. By setting $V_{\mathsf{LP}}(p_t) \leq V_{\mathsf{HP}}(p_t)$, we have

$$\beta^t C_L + \beta V^t_\beta(T(p_t)) \leq (1-p_t)[\beta^t C_{HN} + \beta V^t_\beta(\lambda_{N,A})] + p_t[\beta^t C_{HA} + \beta V^t_\beta(\lambda_{A,A})]. \tag{76}$$

By Lemma 2, $V^t_\beta(p_t)$ is a concave function. Thus,

$$\begin{aligned} V^t_\beta(T(p_t)) &= V^t_\beta(\lambda_{N,A}(1-p_t) + \lambda_{A,A}p_t) \\ &\geq (1-p_t)V^t_\beta(\lambda_{N,A}) + p_t V^t_\beta(\lambda_{A,A}). \end{aligned} \tag{77}$$

By substituting (77) into (76), we can simplify inequality (76) to $(1-p_t)C_{HN} + p_t C_{HA} \geq C_L$, which implies $p_t \geq \frac{C_L - C_{HN}}{C_{HA} - C_{HN}} = \kappa$ when $V^t_{\mathsf{LP}}(p_t) \leq V^t_{\mathsf{HP}}(p_t)$. Thus, $p_t < \kappa$ implies $V_{\mathsf{LP}}(p_t) > V_{\mathsf{HP}}(p_t)$. □

3 Proof of Corollary 2

Proof. Assume that $\lambda_{N,A} \geq \tau$, we have $\lambda_{A,A} > p_F = \frac{\lambda_{N,A}}{1-(\lambda_{A,A}-\lambda_{N,A})} > \lambda_{N,A} \geq \tau$. In this case, By (60) and (64), we have

$$\tau = \frac{C_L - C_{HN}}{C_{HA} - C_{HN}} = \kappa. \tag{78}$$

Thus, $\tau = \kappa$ if $\lambda_{N,A} > \kappa$. Assume that $\lambda_{N,A} < \tau$, then there are two cases for p_F:
Case 1: $p_F > \tau$, then $\lambda_{A,A} > p_F > \tau$, which implies

$$V_\beta(\lambda_{A,A}) = V_{\beta,\mathsf{LP}}(\lambda_{A,A}) = \frac{C_L}{1-\beta}. \tag{79}$$

By (60), (65), and (79), we have

$$\tau = \frac{\beta(C_L - C_{HA})\lambda_{N,A} + C_L - C_{HN}}{(1-\beta)C_{HA} - C_{HN} + \beta C_L}. \tag{80}$$

Therefore, $\tau = \frac{\beta(C_L-C_{HA})\lambda_{N,A}+C_L-C_{HN}}{(1-\beta)C_{HA}-C_{HN}+\beta C_L}$ if $p_F = \frac{\lambda_{N,A}}{1-(\lambda_{A,A}-\lambda_{N,A})} \geq \tau = \frac{\beta(C_L-C_{HA})\lambda_{N,A}+C_L-C_{HN}}{(1-\beta)C_{HA}-C_{HN}+\beta C_L}$ and $\lambda_{N,A} < \frac{\beta(C_L-C_{HA})\lambda_{N,A}+C_L-C_{HN}}{(1-\beta)C_{HA}-C_{HN}+\beta C_L}$.

Case 2: $p_F < \tau$, τ can be computed by (63), (65), and (70). Moreover, for fixed $\lambda_{A,A}$, (63) is a non-decreasing function w.r.t. $\lambda_{N,A}$. Thus, let $\tau^+ = \frac{\lambda_{N,A}}{1-(\lambda_{A,A}-\lambda_{N,A})} = \frac{\beta(C_L - C_{HA})\lambda_{N,A}+C_L-C_{HN}}{(1-\beta)C_{HA}-C_{HN}+\beta C_L}$, $\tau \leq \tau^+$ in Case 2. Therefore, τ^+ is an upperbound for the optimal action in Case 2.

Since (63) is non-decreasing, (80) is decreasing and intersects with (78) at $\lambda_{N,A} = \frac{C_L - C_{HN}}{C_{HA}-C_{HN}}$, we have proved Corollary 2. □

4 Proof of Theorem 2

Proof. Assume that τ is the threshold of offering either HP or LP coupons, then we have $V^t_{\beta,\mathsf{LP}}(\tau) = V^t_{\beta,\mathsf{HP}}(\tau)$. Noticing that the state of the consumer is revealed to the retailer through cost when an HP coupon is offered, we have

$$\begin{aligned} &V^t_{\beta,\mathsf{LP}}(\tau) - V^t_{\beta,\mathsf{HP}}(\tau) \\ &= \beta^t(C_L - (1-\tau)C_{HN} - \tau C_{HA}) + [V^{t+1}_\beta(T(\tau)) - (1-\tau)V^{t+1}_\beta(\lambda'_{N,A}) - \tau V^{t+1}_\beta(\lambda'_{A,A})] \\ &= 0. \end{aligned} \tag{81}$$

The above equation is similar to (58) with $V^{t+1}_\beta(\lambda_{N,A})$ and $V^{t+1}_\beta(\lambda_{A,A})$ replaced by $V^{t+1}_\beta(\lambda'_{N,A})$ and $V^{t+1}_\beta(\lambda'_{A,A})$, respectively. Thus, Lemmas 1-3 still hold. Therefore, the proof follows the same argument for proving Theorem 1; we omit it for brevity. □

References

[1] Acquisti, A. (2010). Background Paper for OECD Joint WPISP-WPIE Roundtable, 1.

[2] Acquisti, A. and Grossklags, J. (2005). *IEEE Security & Privacy*, 2:24–30.

[3] Bertsekas, D. P. (1995). Dynamic programming and optimal control, Vol. 1, 2. Athena Scientific Belmont, MA.

[4] Commission, F. T. and others. (2010). Federal Trade Commission.

[5] Dwork, C. (2011). In *Encyclopedia of Cryptography and Security.* Springer. 338–340.

[6] Feinberg, E. A., Shwartz, A., and Altman, E. (2002). *Handbook of Markov Decision Processes: Methods and Applications.* Kluwer Academic Publishers Boston, MA.

[7] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*, Vol. 2. Taylor & Francis.

[8] Ghosh, A. and Roth, A. (2013). *Games and Economic Behavior.*

[9] Hill, K. (2012). Available at: http://www. forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-outa-teen-girl-was-pregnant-before-her-father-did/ (Accessed July 4th, 2012).

[10] Hsu, J., Huang, Z., Roth, A., Roughgarden, T., and Wu, Z. S. (2013). arXiv preprint arXiv:1311.2828.

[11] Izadi, M. T. and Precup, D. (2005). In *Machine Learning: ECML 2005,* Springer. 593–600.

[12] Laourine, A. and Tong, L. (2010). In *IEEE Transactions on Wireless Communications,* 9(2):723–733.

[13] Lipsa, G. M. and Martins, N. C. (2011). In *IEEE Transactions on Automatic Control,* 56(9):2013–2025.

[14] Lovejoy, W. S. (1987). *Operations Research*, 35(5):736–743.

[15] MacPhee, I. and Jordan, B. (1995). *Probability in the Engineering and Informational Sciences*, 9(02):159–182.

[16] Mansourifard, P. and Javidi, T. (2014). *Tracking of real-valued continuous markovian random processes with asymmetric cost and observation.* Available at:http://anrg.usc.edu/www/papers/Mansourifard_ACC2015.pdf/.

[17] Nayyar, A., Basar, T., Teneketzis, D., and Veeravalli, V. V. (2013). *IEEE Transactions on Automatic Control,* 58(9):2246–2260.

[18] Puterman, M. L. (2009). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Vol. 414, John Wiley & Sons.

[19] Ross, S. M. (2013). *Applied Probability Models with Optimization Applications,* Courier Dover Publications.

[20] Ross, S. M. (1971). *Management Science,* 17(9):587–596.

[21] Rust, R. T., Kannan, P., and Peng, N. (2002). *Journal of the Academy of Marketing Science,* 30(4):455–464.

[22] Sankar, L., Kar, S., Tandon, R., and Poor, H. V. (2011). In *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm),* IEEE. 220–225.

[23] Sarwate, A. and Chaudhuri, K. (2013). *IEEE Signal Processing Magazine*, 30(5):86–94.

[24] Sondik, E. J. (1978). *Operations Research*, 26(2):282–304.

[25] Taylor, C. R. (2004). *RAND Journal of Economics*, 631–650.

[26] Venkitasubramaniam, P. (2013). In *Proc. Allerton Conf.*, 381–388.

[27] Wang, H., Lee, M. K., and Wang, C. (1998). *Communications of the ACM*, 41(3):63–70.

[28] Wu, Y. and Krishnamachari, B. (2012). In *2012 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, IEEE. 27–32..