RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

SOAR Scholarly Open Access at Rutgers

SOAR showcases Rutgers scholarship and makes it freely accessible to the world

Incentive Schemes for Privacy-Sensitive Consumers

Rutgers University has made this article freely available. Please share how this access benefits you.
Your story matters. [https://rucore.libraries.rutgers.edu/rutgers-lib/48445/story/]

This work is an **ACCEPTED MANUSCRIPT (AM)**

This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.

Citation to Publisher Version: Huang, Chong, Sankar, Lalitha & Sarwate, Anand D. (2015). Incentive Schemes for Privacy-Sensitive Consumers. *Lecture Notes in Computer Science 9406*, 358-369.http://dx.doi.org/10.1007/978-3-319-25594-1_21.

Citation to *this* Version: Huang, Chong, Sankar, Lalitha & Sarwate, Anand D. (2015). Incentive Schemes for Privacy-Sensitive Consumers. *Lecture Notes in Computer Science 9406*, 358-369. Retrieved from doi:10.7282/T3VQ34ND.

The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-25594-1

Terms of Use: Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

Article begins on next page

SOAR is a service of RUcore, the Rutgers University Community Repository
RUcore is developed and maintained by Rutgers University Libraries

Incentive Schemes For Privacy-Sensitive Consumers

Chong Huang[1], Lalitha Sankar[1], and Anand D. Sarwate[2]

[1] Arizona State University
[2] Rutgers University

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 simple model and results to consumers with multiple privacy-sensitivity states as well as coupon-dependent state transition probabilities.

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, a retailer could infer a consumer's pregnancy [1]. Such violations 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 balances 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 "indifferent" through "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.

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. The simple model above is extended to multiple consumer 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. Our main results can be summarized as follows:

1. There exists an optimal, stationary, threshold-based policy for offering coupons such that an 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 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, i.e., the retailer stands to benefit by offering HP coupons for a larger set of beliefs about the consumer’s state.
3. For the coupon-dependent Markov model for the consumer, the threshold is lower than for the non-coupon dependent case which encapsulates the fact that highly sensitive consumers will force the retailers to behave more conservatively.

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.

Related Work: Several economic studies have examined consumer’s attitudes towards privacy via surveys and data analysis including studies on the benefits and costs of using private data (e.g., Aquisti and Grossklags in [2]). On the other hand, formal methods such as differential privacy are finding use in modeling the value of private data for market design [3] and for the problem of partitioning goods with private valuation function amongst the agents [4]. In these models the goal is to elicit private information from individuals. Venkitasubramaniam [5] recently used an MDP model to study data sharing in control systems with time-varying state. He minimizes the weighted sum of the utility (benefit) that the system achieves by sharing data (e.g., with a data collector) and the resulting privacy leakage, quantified using the information-theoretic equivocation function. In our work we do not quantify *privacy loss* directly; instead we model *privacy-sensitivity* 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 has not been studied before; in particular, our work focuses on explicitly considering the consequence to the retailer of the consumers’ awareness of privacy violations.

Markov decision processes (MDPs) have been widely used for decades across many fields [6, 7]; in particular, our model is related to problems in control with communication constraints [8,9] where state estimation has a cost. Our costs are action and state dependent and we consider a different optimization problem. Classical target-search problems [10] 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 [11], which was more recently used by Laourine and Tong to study the Gilbert-Elliot channel in wireless communications [12], 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.

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. In the conclusion we describe several other interesting avenues for future work.

2 System Model

We model interactions between a retailer and a consumer via a discrete-time system (see Fig. 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). The consumer responds 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 refusing to shop 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 Consumer Model

Modeling Assumption 1 ***(Consumer's state)*** *We assume the consumer is in one of a finite set of states that determine their response to coupons – each state corresponds to a type of consumer behavior in terms of purchasing. The consumer's state evolves according to a Markov process.*

For this paper, we primarily focus on the two-state case; the consumer may be Normal or Alerted. Later we will extend this model to multiple consumer states. 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 very likely to use coupons to make purchases. However, in the Alerted state, the consumer is less likely to use coupons, 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 (Fig. 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 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 Alerted state and $\lambda_{A,A} = Pr[G_t = \mathsf{Alerted} | G_{t-1} = \mathsf{Alerted}]$ to be the probability of staying in Alerted state when the previous state is also Alerted. The transition matrix $\mathbf{\Lambda}$ of the Markov chain can be written as

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

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 $T(p_t) = (1-p_t)\lambda_{N,A} + p_t\lambda_{A,A}$, represents the belief that the consumer is in Alerted state at time $t+1$ given p_t, the Alerted state belief at time t.

Modeling 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.*

To guarantee Assumption 2 we consider transition matrices in (1) satisfying $\lambda_{A,A} \geq 1-\lambda_{A,A}$, $1-\lambda_{N,A} \geq \lambda_{N,A}$, and $\lambda_{N,A} \geq 1-\lambda_{A,A}$. Thus, by combining the above three inequalities, we have $\lambda_{A,A} \geq \lambda_{N,A}$.

2.2 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 both 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 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 Alerted consumer should incur a high cost (low reward) since an 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.

2.3 The 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:

$$C(p_t, u_t) = \begin{cases} C_L & \text{if } u_t = \mathsf{LP} \\ (1-p_t)C_{HN} + p_tC_{HA} & \text{if } u_t = \mathsf{HP} \end{cases}. \quad (2)$$

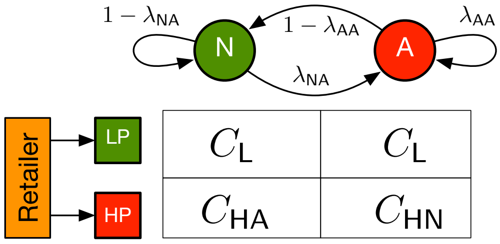


Fig. 1: Markov state transition model for a two-state consumer.

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) [13].

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 $\beta \in (0, 1)$. 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 time horizon. A policy π for the retailer is a rule that selects a coupon to offer at each time. 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

$$V_\beta^{\pi,t}(p_t) = \mathbb{E}_\pi \left[\sum_{i=t}^{\infty} \beta^i C(p_i, A_i) | p_t \right], \tag{3}$$

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

$$V_\beta^t(p_t) = \min_\pi V_\beta^{\pi,t}(p_t) \text{ for all } p_t \in [0, 1]. \tag{4}$$

We define 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 [13]:

$$V_\beta^t(p_t) = \min_{u_t \in \{\mathsf{HP}, \mathsf{LP}\}} \{V_{\beta,u_t}^t(p_t)\}, \tag{5}$$

$$V_{\beta,u_t}^t(p_t) = \beta^t C(p_t, u_t) + V_\beta^{t+1}(p_{t+1} | p_t, u_t). \tag{6}$$

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. We define $\mathcal{P} = \{[0, 1]\}$ to be the belief space and $\mathcal{U} = \{\mathsf{LP}, \mathsf{HP}\}$ to be the action space. In the context of our model, the optimal stationary policy is a deterministic function mapping $\mathcal{P}$ into $\mathcal{U}$. Since the problem is an infinite-horizon, finite state and finite action MDP with discounted cost, by [14], there exists an optimal stationary policy π^* such that starting from time t,

$$V_\beta^t(p_t) = V_\beta^{\pi^*,t}(p_t). \tag{7}$$

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 (5) and (6), 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,

$$V_{\beta,\mathsf{HP}}^t(p_t) = \beta^t C(p_t, \mathsf{HP}) + (1-p_t)V_\beta^{t+1}(\lambda_{N,A}) + p_t V_\beta^{t+1}(\lambda_{A,A}). \tag{8}$$

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

$$\begin{aligned} V_{\beta,\mathsf{LP}}^t(p_t) &= \beta^t C(p_t, \mathsf{LP}) + V_\beta^{t+1}(p_{t+1}) \\ &= \beta^t C_L + V_\beta^{t+1}(T(p_t)). \end{aligned} \tag{9}$$

Correspondingly, the minimum cost function is given by

$$V_\beta^t(p_t) = \min\{V_{\beta,\mathsf{LP}}^t(p_t), V_{\beta,\mathsf{HP}}^t(p_t)\}. \tag{10}$$

3 Optimal Stationary Policies

The first main result is a theorem providing the optimal stationary policy for the two-state basic model in Section 2.

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

$$\pi^*(p_t) = \begin{cases} \mathsf{LP} & \text{if } \tau \le p_t \le 1 \\ \mathsf{HP} & \text{if } 0 \le p_t \le \tau \end{cases}. \tag{11}$$

More precisely, assume that $\delta = C_{HA} - C_{HN} + \beta(V_\beta(\lambda_{A,A}) - V(\lambda_{N,A}))$,

$$\tau = \begin{cases} \frac{C_L - (1-\beta)(C_{HN} + \beta V_\beta(\lambda_{N,A}))}{(1-\beta)\delta} & T(\tau) \ge \tau \\ \frac{C_L + \beta\lambda_{N,A}(C_{HA} + \beta V_\beta(\lambda_{A,A})) - (1-\beta(1-\lambda_{N,A}))(C_{HN} + \beta V_\beta(\lambda_{N,A}))}{(1-(\lambda_{A,A} - \lambda_{N,A})\beta)\delta} & T(\tau) < \tau \end{cases}, \tag{12}$$

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

$$V_\beta(\lambda_{N,A}) = V_\beta(\lambda_{A,A}) = \frac{C_L}{1-\beta} \tag{13}$$

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

$$V_\beta(\lambda_{N,A}) = (1-\lambda_{N,A})[C_{HN} + \beta V_\beta(\lambda_{N,A})] + \lambda_{N,A}[C_{HA} + \beta V_\beta(\lambda_{A,A})], \tag{14}$$

$$V_\beta(\lambda_{A,A}) = \min_{n\ge 0}\{G(n)\}, \tag{15}$$

where

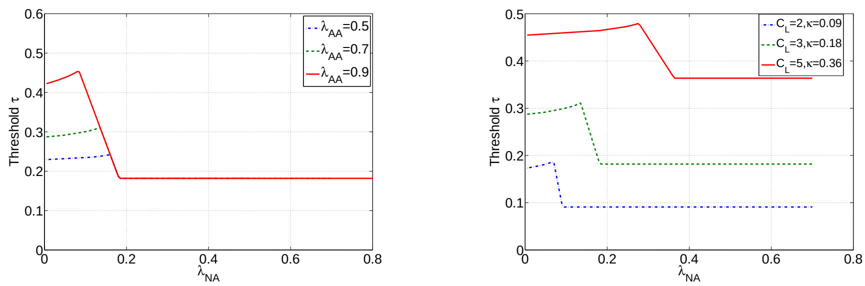
$$G(n) = \frac{C_L \frac{1-\beta^n}{1-\beta} + \beta^n[\bar{T}^n(\lambda_{A,A})(C_{HN} + C(\lambda_{N,A})) + T^n(\lambda_{A,A})C_{HA}]}{1 - \beta^{n+1}[\bar{T}^n(\lambda_{A,A})\frac{\lambda_{N,A}\beta}{1-(1-\lambda_{N,A})\beta} + T^n(\lambda_{A,A})]}, \tag{16}$$

$$T^n(\lambda_{A,A}) = \frac{(\lambda_{A,A} - \lambda_{N,A})^{n+1}(1 - \lambda_{A,A}) + \lambda_{N,A}}{1 - (\lambda_{A,A} - \lambda_{N,A})} \tag{17}$$

$$\bar{T}^n(\lambda_{A,A}) = 1 - T^n(\lambda_{A,A}) \tag{18}$$

$$C(\lambda_{N,A}) = \beta \frac{(1 - \lambda_{N,A})C_{HN} + \lambda_{N,A}C_{HA}}{1 - (1 - \lambda_{N,A})\beta}. \tag{19}$$

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



(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: $\lambda_{A,A} = 0.7, \beta = 0.9, C_{HN} = 1, C_{HA} = 12$.)

Fig. 2: Threshold τ vs. β for different values of $\lambda_{A,A}$ and $\lambda_{N,A}$

Corollary 1. *A sufficient condition for offering an* HP *coupon when transition probabilities* $(\lambda_{N,A}, \lambda_{A,A})$ *are unknown to the retailer is* $p_t \leq \frac{C_L - C_{HN}}{C_{HA} - C_{HN}}$.

The proof of Corollary 1 is provided in the Appendix. The ratio

$$\kappa = \frac{C_L - C_{HN}}{C_{HA} - C_{HN}}. \tag{20}$$

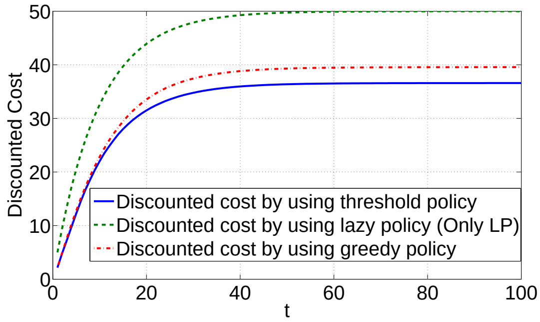
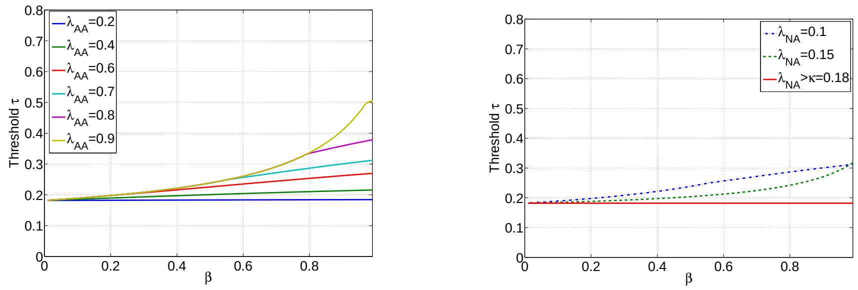


Fig. 3: Discounted cost resulted by using different decision policies

represents 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.

To illustrate the performance of the proposed threshold policy, we compare the discounted cost resulted from the threshold policy with the greedy policy which minimize the instantaneous cost at each decision epoch as well as a lazy policy 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 Fig. 3. The illustration demonstrates that the proposed threshold policy performs better than the greedy policy and the lazy policy. Fig. 2a 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 consumers is 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}, Fig 2b 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.

The relationship between the discount factor β and the threshold τ as functions of transition probabilities is shown in Fig. 4. It can be seen in Fig. 4a 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. Fig. 4b 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



(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: $\lambda_{A,A} = 0.7, C_L = 3, C_{HN} = 1, C_{HA} = 12$.)

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

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 Fig. 4b that the threshold τ equals to κ after $\lambda_{N,A}$ exceeds the ratio κ. This is consistent with results shown in Fig. 2.

The effect of an LP coupon cost on the threshold for different discount factors is plotted in Fig. 5. 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.

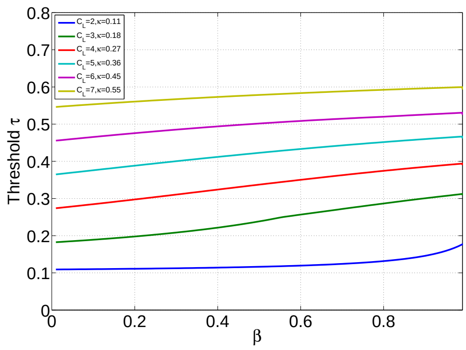


Fig. 5: Threshold τ vs. β for different values of C_L. (Parameters: $\lambda_{N,A} = 0.1, \lambda_{A,A} = 0.9, C_{HN} = 1, C_{HA} = 12$.)

4 Consumer with Multi-Level Alerted States

In this section, the case that the consumer has multiple Alerted states is studied. We define $G_t \in \{\mathsf{Normal}, \mathsf{Alerted}_1, \ldots \mathsf{Alerted}_K\}$ to be the consumer state at time t. If the consumers is in $\mathsf{Alerted}_k$ state, it is even more cautious about coupons than in $\mathsf{Alerted}_{k-1}$ state.

The transition matrix is defined to be

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_{N,N} & \lambda_{N,A_1} & \ldots & \lambda_{N,A_K} \\ \lambda_{A_1,N} & \lambda_{A_1,A_1} & \ldots & \lambda_{A_1,A_K} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{A_K,N} & \lambda_{A_K,A_1} & \ldots & \lambda_{A_K,A_K} \end{pmatrix}. \tag{21}$$

We denote $\bar{\mathbf{e}}_i$ to be the i^{th} row of the transition matrix (21). At each time t, the retailer can offer either an HP or an LP coupon. We define $C_{HN}, C_{HA_1}, \ldots, C_{HA_K}$ to be the costs of the retailer when an HP coupon is offered while the state of the consumer is Normal, $\mathsf{Alerted}_1, \ldots, \mathsf{Alerted}_K$, respectively. 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 belief of the consumer being in Normal, $\mathsf{Alerted}_1, \ldots, \mathsf{Alerted}_K$ state at time t is defined by $p_{N,t}, p_{A_1,t}, \ldots, p_{A_K,t}$, respectively. The expected cost at time t has the following expression:

$$C(\bar{\mathbf{p}}_t, u_t) = \begin{cases} C_L & \text{if } u_t = \mathsf{LP} \\ \bar{\mathbf{p}}_t^T \bar{\mathbf{C}} & \text{if } u_t = \mathsf{HP} \end{cases}, \tag{22}$$

where $\bar{\mathbf{p}}_t = (p_{N,t}, p_{A_1,t}, \ldots, p_{A_K,t})^T$ and $\bar{\mathbf{C}} = (C_{HN}, C_{HA_1}, \ldots, C_{HA_K})^T$. Assume

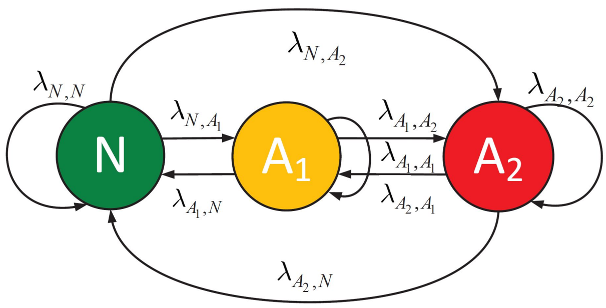


Fig. 6: Markov state transition model for a consumer with three states.

that the retailer has perfect information about the belief of the consumer state, the cost function evolves as follows: by using an LP coupon at time t,

$$V^t_{\beta,\mathsf{LP}}(\bar{\mathbf{p}}_t) = \beta^t C_L + V^{t+1}_\beta(\bar{\mathbf{p}}_{t+1}) = \beta^t C_L + V^{t+1}_\beta(T(\bar{\mathbf{p}}_t)), \tag{23}$$

where $T(\bar{\mathbf{p}}_t) = \bar{\mathbf{p}}_t^T \mathbf{\Lambda}$ is the one step Markov transition function. By using an HP coupon at time t,

$$V^t_{\beta,\mathsf{HP}}(\bar{\mathbf{p}}_t) = \beta^t \bar{\mathbf{p}}_t^T \bar{\mathbf{C}} + \bar{\mathbf{p}}_t^T \begin{pmatrix} V_\beta^{t+1}(\bar{\mathbf{e}}_1) \\ V_\beta^{t+1}(\bar{\mathbf{e}}_2) \\ \vdots \\ V_\beta^{t+1}(\bar{\mathbf{e}}_{K+1}) \end{pmatrix}. \tag{24}$$

Therefore, the minimum cost function is given by (10). In this problem, since

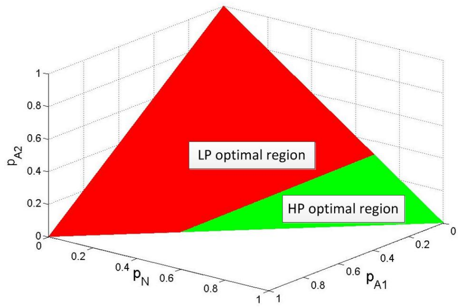


Fig. 7: Optimal policy region for three-state consumer. (Parameters: $\lambda_{N,N} = 0.7, \lambda_{N,A_1} = 0.2, \lambda_{N,A_2} = 0.1; \lambda_{A_1,N} = 0.2, \lambda_{A_1,A_1} = 0.5, \lambda_{A_1,A_2} = 0.3; \lambda_{A_2,N} = 0.1, \lambda_{A_2,A_1} = 0.2, \lambda_{A_2,A_2} = 0.7; \beta = 0.9, C_L = 7, C_{HN} = 1, C_{HA_1} = 10, C_{HA_2} = 20$).

the instantaneous costs are nondecreasing with states 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 for finite many states is guaranteed by Proposition 2 in [15]. The optimal stationary policy for a three-state consumer model (Fig. 6) is illustrated in Fig. **??**. 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.

5 Consumers with Coupon-Dependent Transition

Generally, consumers' reactions 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.

As shown in Fig. 8, if an LP coupon is offered, the state transition follows the Markov chain

$$\mathbf{\Lambda}_{\mathsf{LP}} = \begin{pmatrix} 1-\lambda_{N,A} & \lambda_{N,A} \\ 1-\lambda_{A,A} & \lambda_{A,A} \end{pmatrix}. \tag{25}$$

Otherwise, the state transition follows

$$\mathbf{\Lambda}_{\mathsf{HP}} = \begin{pmatrix} 1-\lambda'_{N,A} & \lambda'_{N,A} \\ 1-\lambda'_{A,A} & \lambda'_{A,A} \end{pmatrix}. \tag{26}$$

According to the model in Section 2, $\lambda_{A,A} > \lambda_{N,A}, \lambda'_{A,A} > \lambda'_{N,A}$. Moreover,

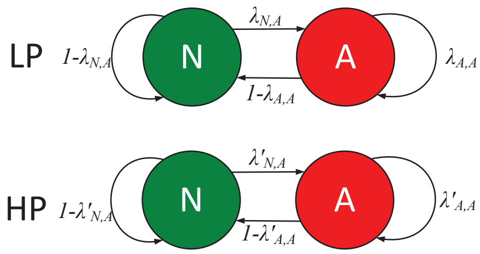


Fig. 8: Markov state transition model for a consumer with coupon dependent transition probabilities.

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,

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

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.

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 is similar to the proof of Theorem 1 except that there are two transition matrices correspond to two different coupons. Thus, it is omitted for brevity.

Fig. 9 shows the effect of costs on the threshold τ. We can see that, for fixed coupon costs, the threshold for offering an HP coupon to a consumer with coupon dependent transition probabilities 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. It can be seen 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.

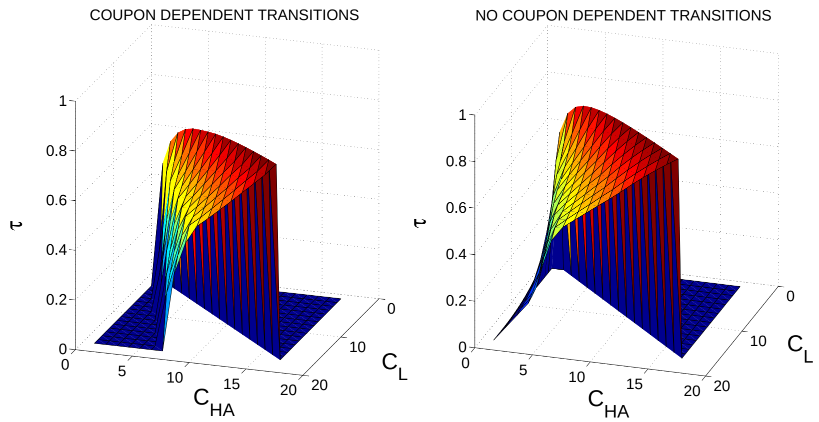


Fig. 9: Optimal policy threshold for consumer with/without coupon dependent transition probabilities. (Parameters: $\lambda_{N,A} = 0.2, \lambda_{A,A} = 0.8, \lambda'_{N,A} = 0.5, \lambda'_{A,A} = 0.9, \beta = 0.9$).

6 Conclusion

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. Our work suggests several interesting directions for future work: cases where retailer has additional uncertainty about the state, for example due to randomness in the received costs, game theoretic models to study the interaction between the retailer and strategic consumers, and more generally, understanding the tension between acquiring information about the consumers and maximizing revenue.

References

1. K. Hill, "How Target figured out a teen girl was pregnant before her father did," 2012. [Online]. Available: http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-outa-teen-girl-was-pregnant-before-her-father-did/
2. A. Acquisti, "The economics of personal data and the economics of privacy," *Background Paper for OECD Joint WPISP-WPIE Roundtable*, vol. 1, 2010.
3. A. Ghosh and A. Roth, "Selling privacy at auction," *Games and Economic Behavior*, 2013.
4. J. Hsu, Z. Huang, A. Roth, T. Roughgarden, and Z. S. Wu, "Private matchings and allocations," *arXiv preprint arXiv:1311.2828*, 2013.
5. P. Venkitasubramaniam, "Privacy in stochastic control: A Markov decision process perspective." in *Proc. Allerton Conf.*, 2013, pp. 381–388.
6. E. A. Feinberg, A. Shwartz, and E. Altman, *Handbook of Markov decision processes: methods and applications*. Kluwer Academic Publishers Boston, MA, 2002.
7. M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2009, vol. 414.
8. G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order lti systems," *Automatic Control, IEEE Transactions on*, vol. 56, no. 9, pp. 2013–2025, 2011.
9. A. Nayyar, T. Başar, D. Teneketzis, and V. V. Veeravalli, "Optimal strategies for communication and remote estimation with an energy harvesting sensor," *Automatic Control, IEEE Transactions on*, vol. 58, no. 9, pp. 2246–2260, 2013.
10. I. MacPhee and B. Jordan, "Optimal search for a moving target," *Probability in the Engineering and Informational Sciences*, vol. 9, no. 02, pp. 159–182, 1995.
11. S. M. Ross, "Quality control under markovian deterioration," *Management Science*, vol. 17, no. 9, pp. 587–596, 1971.
12. A. Laourine and L. Tong, "Betting on gilbert-elliot channels," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 2, pp. 723–733, 2010.
13. D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific Belmont, MA, 1995, vol. 1, 2, no. 2.
14. S. M. Ross, *Applied probability models with optimization applications*. Courier Dover Publications, 2013.
15. W. S. Lovejoy, "Some monotonicity results for partially observed markov decision processes," *Operations Research*, vol. 35, no. 5, pp. 736–743, 1987.

Appendix:

Lemma 1. *The minimum cost when the decision horizon starts from t and only spans m stages is defined to be* $V_\beta^{t,m}$*. Given a time invariant action set* $A_i = A = \{\mathsf{LP}, \mathsf{HP}\}$*, for any* $i = 0, 1, \dots$ *,* $V_\beta^{t,m}(p) = \beta V_\beta^{t-1,m}(p)$*.*

Proof. By (4) and $A_i = A = \{\mathsf{LP}, \mathsf{HP}\}$ for any $i = 0, 1, \ldots$.

$$\begin{aligned} &V_\beta^{t,m}(p) \\ &= \min_\pi \mathbb{E}_\pi \left[\sum_{i=t}^{t+m-1} \beta^i C(p_i, A_i) | p_t = p \right] \\ &= \beta \min_\pi \mathbb{E}_\pi \left[\sum_{i=t-1}^{t+m-2} \beta^i C(p_i, A_i) | p_{t-1} = p \right] \\ &= \beta V_\beta^{t-1,m}(p). \end{aligned} \tag{27}$$

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

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. We define $V_\beta^{t,m}$ to be the minimum cost when the decision horizon starts from t and only spans k stages. For $k = 1$,

$$V_\beta^{t,k}(p) = \min\{C_L, (1-p)C_{HN} + pC_{HA}\}, \tag{28}$$

which is a concave function of p. For $k = n-1$, assume that $V_\beta^{t,k}(p)$ is a concave function. Then, for $k = n$, since $V_\beta^{t,n-1}(p)$ is concave and $V_{\beta,\mathsf{LP}}^{t,k}(p) = \beta^t C_L + V_\beta^{t+1,n-1}(T(p))$, by the definition of concavity and Lemma 1, we can conclude that $V_{\beta,\mathsf{LP}}^{t,k}(p)$ is concave. Also, $V_{\beta,\mathsf{HP}}^{t,k}(p)$ is an affine function of p, thus $V_\beta^{t,k}(p) = \min\{V_{\beta,\mathsf{LP}}^{t,k}(p), V_{\beta,\mathsf{HP}}^{t,k}(p)\}$ is a concave function of p. Taking $k \to \infty, V_\beta^{t,k}(p) \to V_\beta^t(p)$, 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 (28), it is a non-decreasing function of p. Assume that $V_\beta^{t,k}(p)$ is a non-decreasing function for $k = n-1$. For $k = n$, assume that $p_1 \geq p_2$,

$$\begin{aligned} &V_{\beta,\mathsf{LP}}^{t,k}(p_1) - V_{\beta,\mathsf{LP}}^{t,k}(p_2) &(29)\\ &= \beta(V_\beta^{t,n-1}(T(p_1)) - V_\beta^{n-1}(T(p_2))) &(30)\\ &= \beta(V_\beta^{t,n-1}((\lambda_{A,A} - \lambda_{N,A})p_1 + \lambda_{N,A}) \\ &\quad - V_\beta^{t,n-1}((\lambda_{A,A} - \lambda_{N,A})p_2 + \lambda_{N,A}))) &(31)\\ &\geq 0. &(32) \end{aligned}$$

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

$$V_{\beta,\mathsf{HP}}^{t,k}(p_1) - V_{\beta,\mathsf{HP}}^{t,k}(p_2) \geq 0. \tag{33}$$

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

Lemma 3. *We define Φ_{HP} to 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:

$$
\begin{aligned}
V_{\beta}^{t}(p_t) &= V_{\beta}^{t}(ap_{t,1} + (1-a)p_{t,2}) && (34)\\
&\geq aV_{\beta}^{t}(p_{t,1}) + (1-a)V_{\beta}^{t}(p_{t,2}) && (35)\\
&= aV_{\beta,\mathsf{HP}}^{t}(p_{t,1}) + (1-a)V_{\beta,\mathsf{HP}}^{t}(p_{t,2}) && (36)\\
&= a[(1-p_{t,1})[\beta^t C_{HN} + \beta V_{\beta}^{t}(\lambda_{N,A})] + p_{t,1}[\beta^t C_{HA} + \beta V_{\beta}^{t}(\lambda_{A,A})]]\\
&\quad + (1-a)[(1-p_{t,2})[\beta^t C_{HN} + \beta V_{\beta}^{t}(\lambda_{N,A})] + p_{t,2}[\beta^t C_{HA} + \beta V_{\beta}^{t}(\lambda_{A,A})]] && (37)\\
&= V_{\beta,\mathsf{HP}}^{t}(ap_{t,1} + (1-a)p_{t,2}). && (38)
\end{aligned}
$$

Thus, we have shown that:

$$V_{\beta}^{t}(p_t) \geq V_{\beta,\mathsf{HP}}^{t}(ap_{t,1} + (1-a)p_{t,1}) = V_{\beta,\mathsf{HP}}^{t}(p_t). \quad (39)$$

By the definition of $V_{\beta}^{t}(p_t)$ in (10), $V_{\beta}^{t}(p_t) \leq V_{\beta,\mathsf{HP}}^{t}(p_t)$. Therefore, $V_{\beta,\mathsf{HP}}^{t}(p_t) = V_{\beta}^{t}(p_t)$, which implies Φ_{HP} is convex.

Proof of Theorem 1

Proof. We define p_F be the stationary distribution of the Markov transition. Thus, $p_F = \lambda_{A,A}p_F + (1-p_F)\lambda_{N,A}$, which implies $p_F = \frac{\lambda_{N,A}}{1-\lambda_{A,A}+\lambda_{N,A}}$. Remember that the threshold τ is the solution to $V_{\beta,\mathsf{LP}}^{t}(p_t) = V_{\beta,\mathsf{HP}}^{t}(p_t)$. Therefore, we have:

$$
\begin{aligned}
&\beta^t C_L + V_{\beta}^{t+1}(T(\tau))\\
&= (1-\tau)[\beta^t C_{HN} + V_{\beta}^{t+1}(\lambda_{N,A})] + \tau[\beta^t C_{HA} + V_{\beta}^{t+1}(\lambda_{A,A})].
\end{aligned} \quad (40)
$$

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

If $T(\tau) \geq \tau$, which is equivalent to $p_F \geq \tau$, then $V_{\beta}^{t+1}(T(\tau)) = V_{\beta,\mathsf{LP}}^{t+1}(T(\tau))$. Therefore,

$$V_{\beta,\mathsf{LP}}^{t}(\tau) = \lim_{n\to\infty}\left\{\beta^t \frac{1-\beta^n}{1-\beta} C_L + \beta^n V_{\beta}^{t+1}(T^n(\tau))\right\} \quad (41)$$

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 $V_{\beta,\mathsf{LP}}^{t}(\tau) = \beta^t \frac{C}{1-\beta}$. Substitute this into (40) yields:

$$\frac{C_L}{1-\beta} = (1-\tau)C_{HN} + \tau C_{HA} + \beta(\tau V_{\beta}(\lambda_{A,A}) + (1-\tau)V_{\beta}(\lambda_{N,A})). \quad (42)$$

By rearranging terms in the above expression, we have

$$\tau = \frac{\frac{C_L}{1-\beta} - C_{HN} - \beta V_\beta(\lambda_{N,A})}{(C_{HA} - C_{HN}) + \beta(V_\beta(\lambda_{A,A}) - V_\beta(\lambda_{N,A}))}. \tag{43}$$

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

$$V_{\beta,\mathsf{LP}}^t(\tau) = \beta^t C_L + V_\beta^{t+1}(T(\tau)) = \beta^t C_L + V_{\beta,\mathsf{HP}}^{t+1}(T(\tau)) = V_{\beta,\mathsf{HP}}^t(\tau). \tag{44}$$

In this case,

$$C_L + \beta V_{\beta,\mathsf{HP}}(T(\tau)) = V_{\beta,\mathsf{HP}}(\tau). \tag{45}$$

Substitute (8) and (1)into (45), we have

$$\begin{aligned}\tau = &\frac{C_L - (1-\beta(1-\lambda_{N,A}))(C_{HN} + \beta V_\beta(\lambda_{N,A}))}{(1-(\lambda_{A,A}-\lambda_{N,A})\beta)(C_{HA} - C_{HN} + \beta(V_\beta(\lambda_{A,A}) - V(\lambda_{N,A})))} \\ &+ \frac{\beta\lambda_{N,A}(C_{HA} + \beta V_\beta(\lambda_{A,A}))}{(1-(\lambda_{A,A}-\lambda_{N,A})\beta)(C_{HA} - C_{HN} + \beta(V_\beta(\lambda_{A,A}) - V(\lambda_{N,A})))}.\end{aligned} \tag{46}$$

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 assumption, $\lambda_{A,A} \geq \lambda_{N,A} \geq \tau$ and $p_F \geq \lambda_{N,A} \geq \tau$. Thus, both $\lambda_{A,A}$ and $\lambda_{N,A}$ are in Φ_{LP}, therefore,

$$V_\beta(\lambda_{N,A}) = V_\beta(\lambda_{A,A}) = \frac{C_L}{1-\beta}. \tag{47}$$

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

$$V_\beta(\lambda_{N,A}) = (1-\lambda_{N,A})[C_{HN} + V_\beta^1(\lambda_{N,A})] + \lambda_{N,A}[C_{HA} + V_\beta^1(\lambda_{A,A})]. \tag{48}$$

$$V_\beta(\lambda_{A,A}) = \min_{A_t \in \{\mathsf{HP},\mathsf{LP}\}} V_{\beta,A_t}(\lambda_{A,A}) \tag{49}$$

$$= \min\{C_L + V_\beta^1(T(\lambda_{A,A})), V_{\mathsf{HP}}(\lambda_{A,A})\} \tag{50}$$

$$= \min\{C_L\frac{1-\beta^N}{1-\beta}, \min_{0\leq n\leq N-1}\{C_L\frac{1-\beta^n}{1-\beta} + V_{\beta,\mathsf{HP}}^n(T^n(\lambda_{A,A}))\}\}. \tag{51}$$

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

$$V_\beta(\lambda_{A,A}) = \min_{n>0}\left\{C_L\frac{1-\beta^n}{1-\beta} + \beta^n V_{\beta,\mathsf{HP}}(T^n(\lambda_{A,A}))\right\}. \tag{52}$$

we have:

$$V_\beta(\lambda_{A,A}) = \min_{n\geq 0}\left\{\frac{C_L\frac{1-\beta^n}{1-\beta} + \beta^n[\bar{T}^n(\lambda_{A,A})(C_{HN} + C(\lambda_{N,A})) + T^n(\lambda_{A,A})C_{HA}]}{1-\beta^{n+1}[\bar{T}^n(\lambda_{A,A})\frac{\lambda_{N,A}\beta}{1-(1-\lambda_{N,A})\beta} + T^n(\lambda_{A,A})]}\right\}. \tag{53}$$

where

$$T^n(\lambda_{A,A}) = T(T^{n-1}(\lambda_{A,A})) = \frac{(\lambda_{A,A} - \lambda_{N,A})^{n+1}(1-\lambda_{A,A}) + \lambda_{N,A}}{1-(\lambda_{A,A} - \lambda_{N,A})}, \quad (54)$$

$$\bar{T}^n(\lambda_{A,A}) = 1 - T^n(\lambda_{A,A}) \quad (55)$$

$$C(\lambda_{N,A}) = \beta \frac{(1-\lambda_{N,A})C_{HN} + \lambda_{N,A}C_{HA}}{1-(1-\lambda_{N,A})\beta}. \quad (56)$$

Proof of Corollary 1

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

$$\begin{aligned} &\beta^t C_L + \beta V_\beta^t(T(p_t)) \geq \\ &(1-p_t)[\beta^t C_{HN} + \beta V_\beta^t(\lambda_{N,A})] + p_t[\beta^t C_{HA} + \beta V_\beta^t(\lambda_{A,A})]. \end{aligned} \quad (57)$$

According to Lemma 2, $V_\beta^t(p_t)$ is a concave function. Thus,

$$\begin{aligned} V_\beta^t(T(p_t)) &= V_\beta^t(\lambda_{N,A}(1-p_t) + \lambda_{A,A}p_t) \\ &\geq (1-p_t)V_\beta^t(\lambda_{N,A}) + p_t V_\beta^t(\lambda_{A,A}). \end{aligned} \quad (58)$$

By substituting (58) into (57), we can simplify inequality (57) to $(1-p_t)C_{HN} + p_t C_{HA} \leq C_L$, which implies $p_t \leq \frac{C_L - C_{HN}}{C_{HA} - C_{HN}}$ when $V_{\mathsf{LP}}^t(p_t) \geq V_{\mathsf{HP}}^t(p_t)$.