Dynamic Contract Design with Learning

Shiyuan Wang
Department of Management Science and Engineering, School of Economics and Management, Tsinghua University, Beijing, China, 100084, wangshiy20@mails.tsinghua.edu.cn

Yong Liang
Research Center for Contemporary Management, Key Research Institute of Humanities and Social Sciences at Universities, School of Economics and Management, Tsinghua University, Beijing, China, 100084, liangyong@sem.tsinghua.edu.cn

Peng Sun
Fuqua School of Business, Duke University, Durham, North Carolina, 27516, psun@duke.edu

We investigate a dynamic moral hazard problem in which the agent's capability is unknown to both the principal and the agent in the beginning and can be learned over time. Specifically, the agent can exert effort in order to generate random arrivals that are beneficial to the principal. The probability of arrival, however, is determined by the agent's capability. Not knowing the agent's type as well as whether the agent exerts effort or not makes it extremely hard to identify the optimal dynamic contract. Therefore, we focus on designing dynamic contracts that achieve the best regret rate. In a discrete-time setting with two possible agent's capability types, we propose two types of contracts, both of which achieve the regret rate of $O(\ln T)$ for a time horizon T. Our contracts are "history-independent," and therefore easy to implement for the principal. They motivate the agent to always exert effort before termination, and therefore are easy for the agent to respond to. We also establish a regret lower bound, which is $\Omega(\ln T)$ among all contracts, including the ones that do allow the agent to shirk. This implies that our regret rate is the best possible. We further extend these results to continuous-time settings.

Key words: dynamic contract design, learning, moral hazard, online optimization, principal-agent model

1. Introduction

Designing dynamic contracts to address moral hazard problems over time has been a topic of study in the recent operations research literature (see, for example, Sun and Tian 2018, Dawande et al. 2019, Gupta et al. 2023, Zorc et al. 2023). These papers generally assume that the principal and agent both know the probability of output from the agent's effort, even though the effort itself is not observable by the principal. In many settings, however, even if the agent exerts effort, its effect is not fully understood and needs to be learned over time. For example, consider that a firm (principal) hires a sales representative (agent) to generate customers in a new market or for a new product. The novelty of the market/product implies that neither the principal nor the agent knows the exact arrival rate of customers in the beginning, beyond some prior beliefs. If the agent sustains effort, after multiple arrivals have occurred, the true arrival rate will gradually reveal itself. In this process, the principal needs to pay the corresponding rent to motivate the agent's effort, unless the arrival rate in their beliefs becomes so low that it is no longer worth hiring the agent. Therefore, the principal needs to decide how much to pay the agent for each arrival and whether/when to terminate the contract, not only to motivate the agent's effort, but also to learn about the arrival rate.

Learning arrival rate in dynamic moral hazard problems could also arise in settings where the players do not know the agent's capability in the beginning. Consider an example to which readers of this paper may be able to relate. When a newly graduated Ph.D. (agent) starts a faculty job at a university (principal), neither party may know the professor's productivity, which is gradually revealed with papers getting accepted over time.

In this paper, we study dynamic contracts over a discrete finite planning horizon, during which an "arrival" (say, customers) may occur if the agent exerts effort in a period. Each arrival brings the principal a fixed revenue, while effort is costly to the agent and unobservable to the principal. Furthermore, the probability of arrival may be either high or low. The high arrival probability means that the market/product is profitable, or the agent is sufficiently capable to be worth hiring. The low arrival probability, on the other hand, means that the market/product is not profitable, or the agent is not worth hiring. For convenience, we refer to high and low arrival probability as the agent's type, which can be "capable" or "incapable." At the beginning of the time horizon, the principal and agent do not know the exact type, and share a common prior probability on the type being capable. The principal, who can commit to a long-term contract, utilizes both a payment schedule and a contract termination criterion to incentivize the agent to exert effort and hedge against the risk of facing an incapable agent.

Simultaneously managing learning and moral hazard over time can be quite challenging. If the players believe that the arrival rate tends to be low (but is still worth pursuing), the principal needs to pay the agent more for each arrival in order to sustain the agent's effort. However, if the agent has shirked in the past, the agent knows that the relatively low number of arrivals is due to the lack of effort, rather than the arrival rate being inherently low. The principal, who may not know about past shirking, however, believes in a lower arrival rate, and hence feels obliged to pay a higher rent than necessary. Moreover, the divergence of beliefs persists until the very end of the planning horizon. Consequently, this persistent divergence of beliefs, and the fact that a lower arrival rate implies higher rent payment, potentially leads to the agent's strategic shirking behavior, which significantly complicates contract design.

This complication means that the traditional approach of using dynamic program/optimal control to obtain optimal contracts does not work in our setting. In fact, we show in the paper that the corresponding dynamic program formulation involves a high-dimensional state space. This implies that the model is hard to solve even numerically, and has little hope to yield practical contracts or useful insights. Therefore, we utilize the regret-minimization approach that is popular in the learning literature (see, e.g. Shalev-Shwartz 2012, Slivkins et al. 2019, Lattimore and Szepesvári 2020).

Our contributions are threefold. First, from a theoretical perspective, our paper contributes to the dynamic moral hazard literature by considering learning the arrival rate. In particular, we introduce the regret-minimization approach to a dynamic contract design setting in which both the principal and the agent learn about the arrival rate of good events. Second, from a practical perspective, our results provide prescriptive guidance on designing easy-to-implement dynamic contracts that allow the players to learn the market condition/agent's capability. Our proposed contracts are easy for the principal to implement because the contracts depend on realized history only through the total number of arrivals and the time period. These contracts are also easy for the agent to follow because they create the incentive for the agent to always exert effort before a potential termination. In terms of performance, our contracts achieve the optimal regret rate. Third, we extend the results to a continuous-time counterpart of the discrete-time setting. We have not seen the regret-minimization approach used in a continuous-time setting in the learning literature before.

1.1. Literature Review

Our study is closely related to the extensive stream of literature on moral hazard problems, starting from the seminal works of Holmström (1979) and Grossman and Hart (1983) on contract theory. We focus on studying dynamic moral hazard problems, in which the agent takes private actions over time. Many early studies of dynamic moral hazard problems consider discrete-time models (Rogerson 1985, Spear and Srivastava 1987, Gibbons and Murphy 1992, Holmström 1999). Spear and Srivastava (1987) proposed a key idea of formulating dynamic contract design problems recursively as discrete-time stochastic dynamic programming models using agents' promised utility as a state variable. This modeling framework is extended to continuous-time settings and analyzed using stochastic optimal control techniques (see, for example, Sannikov 2008, Biais et al. 2010, for uncertainties modeled as Brownian motions and Poisson processes, respectively). We study a dynamic moral hazard setting in which an agent exerts effort to increase the arrival rate of either a Bernoulli process in discrete time, or a Poisson process in continuous time. Our setting in continuous time and with the agent's type known is similar to the one studied in Sun and Tian (2018).

Demarzo and Sannikov (2017) and He et al. (2017) study dynamic contract design problems in which the principal and the agent both learn some underlying "fundamentals" following a Brownian motion. Players in both papers observe some signals that also follow a Brownian motion, whose drift is affected by the agent's actions. The Gaussian-Gaussian structure implies that one can use Kalman filtering to analyze the system. Both papers study the system in a "steady state", such that the overall volatility of the posterior dynamics remains a constant over time in equilibrium. In contrast, the agent's type in our setting is fixed over time. Hence, steady-state and the corresponding analytical approach are irrelevant in our setting. A key difference between these two papers is that in Demarzo and Sannikov (2017) the agent has limited liability, which is similar to our setting, while in He et al. (2017) the agent is risk averse whose utility is an exponential function on consumption decisions. In contrast, Prat and Jovanovic (2014) is focused on finding an optimal dynamic optimal contract with persistent and unknown agent's type, similar to ours. However, similar to He et al. (2017) and different from ours, the agent in Prat and Jovanovic (2014) is risk averse with an exponential utility function. All these three continuous-time papers utilize a first-order method, which describes a necessary condition for incentive compatibility, and verify sufficiency after obtaining the corresponding optimal solution. As pointed out by Demarzo and Sannikov (2017), "[the first-order approach] is a powerful approach, but analytical results are still difficult to establish formally in discrete time."

As mentioned earlier and elaborated in more detail in Section 2.2, the dynamic programming/optimal control approach used in the previous literature is not tractable in our setting. Therefore, we follow the regret-minimization approach from the extensive line of literature on online optimization. We refer readers to Bubeck (2011), Hazan (2016) and the references therein for recent surveys. The general setting of online optimization problems assumes that there is uncertainty in the problem data. The decision maker needs to determine a sequence of decisions to optimize an expected value over a long planning horizon. Data is revealed incrementally after each decision is implemented. The decision maker learns the uncertainties from the data on the fly. The key to designing an effective online optimization algorithm is to balance exploration and exploitation: exploring suboptimal decisions to gather informative data, while exploiting solutions that align with current data and estimates. In our setting, exploration corresponds to continuously motivating the agent to work and gradually reveal its type; exploitation can be thought of as terminating the agent if it appears to be the incompetent type.

The MAB problem is a classic problem that exemplifies the exploration–exploitation trade-off (see, e.g. Lattimore and Szepesvári 2020, Slivkins et al. 2019). In a standard MAB setting, the decision maker faces a fixed set of actions, with each one termed as an "arm." In each period, the decision maker chooses an arm and receives a reward. The reward is drawn from some fixed but unknown distribution that depends only on the chosen arm. The decision maker tries to maximize the total collected reward over a fixed number of periods. The regret is often defined as the difference between the performance of knowing the best arm versus the proposed algorithm. Traditional MAB problems do not involve a strategic agent to interact with, as we do.

A recent learning literature considers principal-agent settings as well. The agent in Zhu et al. (2023) is privately informed about its own type and takes hidden actions. Furthermore, their agent is myopic. In contrast, neither our principal nor the agent knows the agent's type, and our agent is strategic over the entire time horizon. Amin et al. (2013) and Amin et al. (2014) assume a long-term strategic agent, who knows its own type, but does not take hidden actions. That is, they consider adverse selection settings, while our study deals with moral hazard and learning. Under a somewhat related but different context, Zhao et al. (2022) study a supply chain contract design problem in which neither the supplier nor the retailer knows the demand distribution. The retailer, who has no private information other than observed demands, acts according to certain learning algorithms rather than taking the best response to the contract, and the supplier responds to the inventory decisions of the retailer. There is no hidden action in the setting of Zhao et al. (2022).

The remainder of this paper is organized as follows. Section 2 introduces the model and the regret-minimization formulation. In particular, Section 2.3 illustrates why the traditional recursive formulation for dynamic program/optimal control faces the curse of dimensionality issue. Section 3 derives the regret lower bound of any contract, either incentive compatible or not. Then, Section 4 develops two discrete-time online dynamic contracts whose regret-rate upper bounds match the lower bound from Section 3. Section 5 extends the discrete-time setting into a continuous-time one. Finally, Section 6 concludes the paper and discusses some potential future research directions and their challenges. All the proofs and technical materials are presented in the Appendix.

Before closing this section, we introduce some mathematical notations to be used in the rest of the paper. For an integer T, notation $[T]$ represents the sequence $1,\ldots,T$. The use of O and Ω is standard, that is, given functions $f,g:\mathbb{N}\to[0,\infty)$, we say

$$f(x)=O(g(x)) \quad \text{if} \quad \limsup_{x\to\infty}\frac{f(x)}{g(x)}<\infty,$$

and

$$f(x)=\Omega(g(x)) \quad \text{if} \quad \liminf_{x\to\infty}\frac{f(x)}{g(x)}>0.$$

We use the notation $\mathbf{0}$ to represent a vector of 0's of appropriate dimension. For convenience of exposition, if $t>t'$, we let $\sum_{i=t}^{t'} x_i=0$ for any sequence $\{x_i\}$.

2. Model

In this section, we introduce the online dynamic contract design problem faced by a principal. We first describe the basic problem settings in Section 2.1. Then, we formulate the regret incurred by the online contract in Section 2.2 and discuss the inherent challenges in the problem. Finally, we present a recursive formulation in Section 2.3, which demonstrates curse of dimensionality, and explains why we choose to pursue the regret-minimization approach.

2.1. Basic Settings

Consider a principal hiring an agent over a discrete time horizon $t \in [T]$. The agent's effort can potentially generate an observable arrival worth R to the principal in each period. (There is no arrival without the agent's effort.) The agent's effort is not observable to the principal, and costs the agent b per period of time. The agent may be one of two types: the "capable" type that generates an arrival with probability $\overline{\lambda}$ in a period when exerting effort; the "incapable" type with probability $\underline{\lambda}$. We assume that

$$R\underline{\lambda} \leq b \leq R\overline{\lambda}\ , \tag{1}$$

that is, the capable type is worth hiring while the incapable one is not, from a societal perspective. Neither the principal nor the agent knows the agent's type, and they share a common prior probability P_0 in the beginning of the time horizon that the agent is capable. Therefore, the principal needs to motivate the agent to exert effort not only for the revenue, but also to learn its type.

In each period t, let $x_t \in \{0,1\}$ represent if there is an arrival ($x_t = 1$) or not ($x_t = 0$). Define $N_t := \sum_{s=1}^{t} x_t$ to be a counting process, which generates a filtration $\mathcal{N} := \{\mathcal{N}_t\}_{t\in[T]}$, where $\mathcal{N}_t = \sigma\{N_1, \ldots, N_t\}$. Denote an non-negative and $\mathcal{N}_t$-measurable random variable β_t to be the payment from the principal to the agent in period t, and $\boldsymbol{\beta} := \{\beta_t\}_{t\in[T]}$ the corresponding payment process, which is adapted to $\mathcal{N}$. The principal also decides an $\mathcal{N}$-stopping time τ to terminate the agent.

Define $\Gamma := (\boldsymbol{\beta}, \tau)$ to represent a contract of the principal. In response to the contract, the agent exerts effort according to an $\mathcal{N}$-predictable effort process $\boldsymbol{\nu} := \{\nu_t\}_{t\in[T]}$, in which $\nu_t = 1$ represents exerting effort, and $\nu_t = 0$ shirking, in period t. Without loss of generality, we let $\nu_t = 0$ for all $t > \tau$, indicating that an agent should not exert effort after being terminated. Denote $\mathbb{P}_{\boldsymbol{\nu}}(\cdot)$ as the probability measures induced by effort process $\boldsymbol{\nu}$.

Define the agent's total utility under contract Γ following effort process $\boldsymbol{\nu}$ as

$$W(\Gamma, \boldsymbol{\nu}) := \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} \beta_t - b\nu_t\right], \tag{2}$$

in which $\mathbb{E}_{\boldsymbol{\nu}}[\cdot]$ represents taking expectation over the arrival uncertainties induced by $\boldsymbol{\nu}$, as well as the uncertain agent's type.

In particular, for any contract Γ, define $\hat{\boldsymbol{\nu}}(\Gamma)$ to be the agent's *best response effort process*, such that

$$W(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma)) \geq W(\Gamma, \boldsymbol{\nu}),\ \forall \boldsymbol{\nu}. \tag{3}$$

Because the agent can always choose not to work at all and obtain a non-negative utility under any contract, we must have

$$W(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma)) \geq 0. \tag{4}$$

In other words, we do not need a separate explicit individual rationality constraint.

Further define the principal's utility under contract Γ and effort process $\boldsymbol{\nu}$ as

$$U(\Gamma,\boldsymbol{\nu}) \coloneqq \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} Rx_t - \beta_t\right]. \tag{5}$$

Overall, the principal tries to solve the following dynamic contract design problem with learning,

$$\mathcal{J} \coloneqq \max_{\Gamma} U(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma)). \tag{6}$$

It is worth discussing the learning component in our problem. Generally speaking, the contract needs to motivate the agent to exert effort to generate good events. Meanwhile, proper incentives to motivate effort depend on the agent's type, which is uncertain. Insufficient incentives create persistent divergence in the belief about the agent's type, which complicates the design of revenue maximizing contracts. For example, if the agent chooses to shirk without the principal's knowledge in a period, the lack of arrival in the period does not change the agent's belief about its capability, but distorts downward the principal's belief about the agent being capable. This potential divergence of belief probabilities between the players implies that in order to properly model incentive compatibility in a recursive form, the state space needs to capture all possible belief probabilities of the agent, beyond the one held by the principal. Equivalently, the state space needs to encode shirking that may have occurred in the past. This complexity makes it extremely hard to solve $\mathcal{J}$ directly using the traditional dynamic programming/optimal control approach. We explain this in more detail in Section 2.3.

Moreover, while both the principal and the agent learn about the agent's type on the fly, the principal constantly faces the well-known trade-off between exploration and exploitation. In our context, we can think of exploration as continuously motivating the agent to exert effort. This allows the principal to collect data towards more accurate estimation of the agent's type, even if the agent already appears not worth hiring under the currently available data. Exploitation can be thought of as terminating the agent to cut the loss for future periods. Exploration may offer payments that are too high, while exploitation may terminate the agent too late, compared with the optimal contract if the agent's type is known. Next, we define the regret minimization objective for online dynamic contract design.

2.2. The Regret and Challenges

In traditional single-decision maker online learning problems, "regret" is often defined as the difference between the expected objective values of the optimal control policy with perfect information and that of the proposed policy. In our online dynamic contract design problem, the "perfect information" benchmark corresponds to a setting in which the principal knows the agent's type, but the agent does not. Such a perfect information benchmark corresponds to a dynamic information design problem. How to solve it to optimality is unclear. Therefore, we consider an alternative benchmark, which corresponds to letting both the principal and the agent know about the agent's type. Generally speaking, even though the principal's knowledge improves its utility, the agent's knowing the type may decrease the principal's utility. Therefore, it is not *a priori* clear that such a construction indeed yields an upper bound of the true optimal value $\mathcal{J}$. Next, we formalize the benchmark and establish that it is an upper bound for $\mathcal{J}$.

To this end, define the societal utility under a contract Γ and effort process $\boldsymbol{\nu}$ as

$$V(\Gamma,\boldsymbol{\nu}) := U(\Gamma,\boldsymbol{\nu}) + W(\Gamma,\boldsymbol{\nu}), \tag{7}$$

which is the total utility between the principal and the agent.

Following (4), it is clear that

$$\max_{\Gamma} V\big(\Gamma,\hat{\boldsymbol{\nu}}(\Gamma)\big) \geq \mathcal{J}. \tag{8}$$

Furthermore, define the first-best societal value given the agent's type $\lambda \in \{\overline{\lambda},\underline{\lambda}\}$ as

$$\mathrm{OPT}(\lambda) := \max\{T(\lambda R - b), 0\}. \tag{9}$$

From assumption (1), it is clear that $\mathrm{OPT}(\overline{\lambda}) = T(\overline{\lambda}R - b)$ and $\mathrm{OPT}(\underline{\lambda}) = 0$. The following result presents an upper bound for the optimal value $\mathcal{J}$, which serves as the benchmark for us to define regret.

Proposition 1. *The principal's expected utility under the optimal contract for the dynamic contract design problem, $\mathcal{J}$, satisfies*

$$\mathcal{J} \leq \max_{\Gamma} V\big(\Gamma,\hat{\boldsymbol{\nu}}(\Gamma)\big) \leq P_0 \cdot OPT(\overline{\lambda}) + (1-P_0)\cdot OPT(\underline{\lambda}) = P_0 T\overline{\lambda}(R-\beta), \tag{10}$$

in which we define

$$\beta := \frac{b}{\overline{\lambda}}. \tag{11}$$

The value β is the lowest payment for each arrival to guarantee that the capable agent (with arrival probability $\overline{\lambda}$) exerts effort. Furthermore, in Section EC.1.2, we show that $\mathrm{OPT}(\bar{\lambda})$ and $\mathrm{OPT}(\underline{\lambda})$ equal to the principal's optimal utility when both parties know that the agent's type is $\bar{\lambda}$ and $\underline{\lambda}$, respectively.

For any contract Γ, define its regret as the upper bound defined in (10) minus the principal's utility,

$$\mathrm{Reg}(\Gamma;T) := P_0\overline{\lambda}(R-\beta)T - U(\Gamma,\hat{\boldsymbol{\nu}}(\Gamma)). \tag{12}$$

Later in the paper, we propose contracts that achieve a regret rates of $O(\ln T)$. Given that the upper bound in the regret calculation is the optimal societal value when both the principal and the agent know the agent's type, our results imply that as T increases to infinity, the principal's average profit per period gradually approaches the first-best societal value, while the agent's average rent per period diminishes to zero.

There are a number of challenges in designing our contract. First of all, describing a general history-dependent dynamic contract requires specifying the payment and stopping time under each of the 2^T trajectories of arrivals. Therefore, it is already non-trivial to even describe a dynamic contract. Furthermore, knowing the agent's effort process is critical for the principal to correctly update its belief about the agent's type. Consequently, naïvely extending an incentive compatible contract from the case where the agent's type is known (as in Sun and Tian 2018) may result in the agent strategically shirking, leading to a linear regret. In the following, we describe the information updating process assuming that the agent has been exerting effort, and then present an example which yields a regret linear in T.

For this purpose, define the full-effort process

$$\bar{\boldsymbol{\nu}} := \{\nu_1 = 1, \dots \nu_\tau = 1\}, \tag{13}$$

such that the agent always exerts effort under this process. For ease of exposition, we use $\mathbb{P}(\cdot)$ to represent $\mathbb{P}_{\bar{\boldsymbol{\nu}}}(\cdot)$ when the context is clear.

Equipped with these notations, we first present the following result that summarizes the two players' shared belief about the agent's type under the full-effort process.

Lemma 1. *After any time t with the full-effort process and history $\mathcal{N}_t$ such that the number of good arrivals is N_t, the posterior belief that the agent is of type $\overline{\lambda}$, denoted by $P_t(N_t)$, is*

$$P_t(N_t) := \mathbb{P}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_t\right) = \mathbb{P}\left(\lambda = \overline{\lambda} \mid N_t\right) = \left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t} \left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-N_t} \left(\frac{1}{P_0} - 1\right) + 1\right)^{-1}. \tag{14}$$

That is, the belief probability that agent is of type $\overline{\lambda}$ under the full-effort process $\bar{\nu}$ and history $\mathcal{N}_t$ only depends on the number of arrivals N_t up to period t, which follows the expression (14) for $P_t(N_t)$. In the remainder of the paper, when the context is clear, we use the notation P_t to abbreviate $P_t(N_t)$.

Next, we show an example in which we naïvely extend the contract from Sun and Tian (2018) by ignoring the possibility of the agent strategically shirking, and verify that the contract leads to a regret linear in T.

Example 1. Motivated by the optimal contract in Sun and Tian (2018) for a setting without learning, consider paying the agent a positive amount if there is an arrival in period t, and zero if there is not, such that the agent is indifferent between working and shirking in each period. That is, if the belief probability that the agent is of type $\overline{\lambda}$ is P at the beginning of period t, the payment for an arrival is

$$\beta(P) := \frac{b}{\overline{\lambda}P + \underline{\lambda}(1-P)}. \tag{15}$$

Should the agent estimate its likelihood of being $\overline{\lambda}$ as P, a payment $\beta(P)$ for an arrival leaves the agent indifferent between exerting effort or not (and hence does not mind exerting effort) in the period. Because the belief probability $P = P_t(N_t)$ itself depends on arrivals, this contract is history-*dependent*. Later in Section 4, we introduce and examine history-*independent* contracts, which, surprisingly, produce the optimal regret rate.

Furthermore, the contract terminates the agent when the belief probability of type $\overline{\lambda}$ is too low. Specifically, the termination time satisfies

$$\{\tau = t\} := \bigcap_{s=1}^{t-1}\{P_s \geq \bar{p}_s\} \bigcap \{P_t < \bar{p}_t\}. \tag{16}$$

for a sequence of thresholds $\{\bar{p}_s\}_{s\in[T]}$.

However, the agent's best-response effort process facing such a seemingly reasonable contract is to not always exert effort. Consequently, the expression P_t does not reflect the agent's belief in period t. In particular, P_t may underestimate the agent's belief, because it implicitly assumes that the agent has always exerted effort before period t. In order to illustrate the claims above, we consider an extreme case where $\underline{\lambda} = 0$, for which we can identify the agent's best-response effort process. We show that the corresponding regret grows linearly in T, as stated in the following result.

Proposition 2. *Consider $\underline{\lambda} = 0$ and define a contract $\check{\Gamma} = \left(\{\check{\beta}_t\}_{t\in[T]}, \check{\tau}\right)$ such that $\check{\beta}_t = \beta(P_t)$ if $x_t = 1$ and $\check{\beta}_t = 0$ if $x_t = 0$; and $\check{\tau}$ is defined according to (16) for any sequence $\{\bar{p}_s\}_{s\in[T]}$. We have*

$$\mathit{Reg}(\check{\Gamma}; T) = \Omega(T).$$

□

This example demonstrates that ignoring the agent's strategic response may yield a linear regret. Therefore, should the agent choose to shirk in a period, the principal would have to capture the divergence of belief probabilities to properly model incentive compatibility. In other words, if there exist optimal contracts that allow the agent to shirk from time to time while achieving satisfactory regret rate, such contracts need to take the agent's reaction following all possible belief probabilities into consideration, which appears a daunting task.

Given the complexity of the agent's best-response effort process in general, we focus on designing contracts that indeed create the incentive for the agent to always exert effort, and show that such a design yields desirable regret rate. For this purpose, we formally define a contract as being *incentive compatible* (IC) if it induces the agent to always exert effort before contract termination. That is,

$$W(\Gamma,\bar{\boldsymbol{\nu}}) \geq W(\Gamma,\boldsymbol{\nu}),\ \forall \boldsymbol{\nu}. \tag{IC}$$

Before closing this section, we provide a dynamic programming formulation to obtain the optimal incentive-compatible contract. Even if we focus on incentive compatible contracts, the aforementioned curse of dimensionality still occurs, which further justifies our regret-minimization approach.

2.3. Recursive Formulation

Define $\hat{J}(w)$ as the optimal expected profit for the principal who delivers the agent an expected utility of w by an incentive compatible contract. That is,

$$\begin{aligned} \hat{J}(w) := \max_{(\boldsymbol{\beta},\tau)} \mathbb{E}_{\bar{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau} Rx_t - \beta_t\right] \\ \text{s.t. } w = \mathbb{E}_{\bar{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau} \beta_t - b\right] \\ \mathbb{E}_{\bar{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau} \beta_t - b\right] \geq \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} \beta_t - b\nu_t\right],\ \forall \boldsymbol{\nu}. \end{aligned} \tag{17}$$

Note that $w \geq 0$ is implied by the second constraint and inequality (4). Clearly, we can fully characterize such an optimal incentive compatible contract by solving $\max_w \hat{J}(w)$. In order to express the optimization problem (17) in a recursive formulation, it is important to understand that if the agent has shirked, its belief about the type may be different from the principal's. In particular, after seeing n arrivals up to period t, the principal's belief is $P_t(n)$ as defined in (14), assuming that the agent has always exerted effort. For any $s \geq n$, the parameter

$$\lambda_{s,n} := \overline{\lambda} P_s(n) + \underline{\lambda}\big(1 - P_s(n)\big)$$

is the expected arrival rate after s working periods with n arrivals. The agent, knowing that it has shirked for k periods (among the $t-n$ periods with no arrival), has a different belief, $P_{t-k}(n)$. Consequently, the principal's optimization in each period needs to be contingent upon each possible shirking history of the agent. This is reflected in the following recursive form in the beginning of period t with $n \leq t-1$ arrivals up to the end of period $t-1$.

Proposition 3. *For any $t \in [T]$, $n \leq t-1$, and $(t-n)$-dimensional vector $\mathbf{w} := (w_0, w_1, \ldots, w_{t-1-n})^\intercal$, define the following dynamic programming recursion*

$$\begin{aligned}
J(t,n,\mathbf{w}) := \max_{\mathbf{w}^\pm,\beta^\pm,I} \; & I\left\{\lambda_{t,n}\left[(R-\beta^+) + J(t+1,n+1,\mathbf{w}^+)\right]\right. \\
& \left. +(1-\lambda_{t,n})\left[-\beta^- + J(t+1,n,\mathbf{w}^-)\right]\right\} \\
s.t. \qquad & w_0 = I(\lambda_{t,n}(\beta^+ + w_0^+) + (1-\lambda_{t,n})(\beta^- + w_0^-) - b) \\
& w_k = I\max\left\{\lambda_{t-k,n}(\beta^+ + w_k^+) + (1-\lambda_{t-k,n})(\beta^- + w_k^-) - b,\right. \\
& \qquad \left. \beta^- + w_{k+1}^-\right\}, \quad \forall t > 1, \; k = 0, \ldots, t-1-n \\
& \mathbf{w}^+ \in \mathbb{R}_+^{t-n}, \; \mathbf{w}^- \in \mathbb{R}_+^{t+1-n}, \beta^\pm \in \mathbb{R}_+, \; I \in \{0,1\},
\end{aligned} \tag{18}$$

with boundary conditions

$$J(T+1,n',\mathbf{0}) = 0, \; \textit{and } J(T+1,n',\mathbf{w}) = -\infty, \; \textit{if } \mathbf{w} \neq \mathbf{0}, \tag{19}$$

in which $n' \leq T$, and both $\mathbf{0}$ and $\mathbf{w}$ are $T+1-n'$-dimensional vectors.

We have $\hat{J}(w) = J(1,0,w)$.

In the dynamic program, the element w_0 of the state vector $\mathbf{w}$ represents the promised utility with 0 shirking period before, and w_k the "threat utility" (Fernandes and Phelan 2000) if the agent has shirked k times before. Decision vector $\mathbf{w}^+$ represents the next period's promised/threat utilities when there is an arrival in period t, and $\mathbf{w}^-$ the corresponding utilities if period t sees no arrival. Decisions β^+ and β^- represent the payment when there is and there is not a good arrival in period t, respectively. The subscript of these vectors still represents the number of shirking periods by the end of period t. Finally, the decision I indicates whether the agent is terminated in the beginning of period t. Clearly, the dimension of the state space grows with the number of time periods, making the dynamic programming formulation intractable.

Consequently, in the remainder of the paper, we focus on the regret-minimization approach. In the next section, we first present a lower bound on the rate of regret. After that, we provide two algorithms that achieve this regret lower bound.

3. Regret Lower Bound

In this section, we establish that any contract (incentive compatible or not) cannot yield a regret that grows slower than $O(\ln T)$.

In order to obtain a lower bound on the regret (12), we need to upper bound the principal's utility, $U(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma))$. Because of (4) and (7), we know that the societal utility $V(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma))$ is an upper bound of $U(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma))$. For clarity of exposition, we use $\hat{\boldsymbol{\nu}} = \{\hat{\nu}_t\}_{t\in[T]}$ in place of $\hat{\boldsymbol{\nu}}(\Gamma)$ to represent the best-response effort process with respect to contract Γ when there is no ambiguity. Note that $\hat{\boldsymbol{\nu}}$ is a stochastic process that is also parameterized by the true type of the agent. Following expressions (2), (4), (5) and (7), we have the following,

$$U(\Gamma,\hat{\boldsymbol{\nu}}(\Gamma)) \le V(\Gamma,\hat{\boldsymbol{\nu}}(\Gamma)) = P_0(\overline{\lambda}R-b)\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau}\hat{\nu}_t \mid \overline{\lambda}\right] - (1-P_0)(b-\underline{\lambda}R)\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau}\hat{\nu}_t \mid \underline{\lambda}\right].$$

Together with (12), we have

$$\mathrm{Reg}(\Gamma,T) \ge P_0\overline{\lambda}(R-\beta)\left(T-\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau}\hat{\nu}_t \mid \overline{\lambda}\right]\right) + (1-P_0)(b-\underline{\lambda}R)\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau}\hat{\nu}_t \mid \underline{\lambda}\right]. \tag{20}$$

We now consider two cases depending on whether an incapable agent exerts enough effort under contract Γ. First, suppose that the $\underline{\lambda}$-type agent exerts effort in sufficiently many periods. That is,

$$\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau}\hat{\nu}_t \mid \underline{\lambda}\right] \ge \frac{1}{2}C\ln T, \tag{21}$$

for a constant $C := (1-\overline{\lambda})/(1-\underline{\lambda}) \in [0,1]$. Dropping the first term on the right-hand side of inequality (20), which is non-negative, we obtain

$$\mathrm{Reg}(\Gamma,T) \ge \frac{1}{2}(1-P_0)(b-\underline{\lambda}R)C\ln T, \tag{22}$$

which readily verifies that the regret lower bound is in the order of $\ln T$ when condition (21) holds.

Next, suppose the opposite of (21) holds, that is,

$$\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau}\hat{\nu}_t \mid \underline{\lambda}\right] < \frac{1}{2}C\ln T. \tag{23}$$

For this case, we first introduce additional notations. Given a trajectory $\mathcal{N}$, let $\tau(\mathcal{N})$ denote the $\mathcal{N}$-stopping time in which the agent is terminated, and $\hat{\nu}_t(\mathcal{N}) \in \{0,1\}$ represent whether the agent exerts effort in period t according to the $\mathcal{N}$-predictable best-response effort process $\hat{\boldsymbol{\nu}}$. Furthermore, define $\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{N} \mid \lambda)$ as the probability of observing the trajectory $\mathcal{N}$ given the effort process $\boldsymbol{\nu}$ and the arrival rate $\lambda \in \{\overline{\lambda},\underline{\lambda}\}$. We have

$$\mathbb{P}_{\boldsymbol{\nu}}(\mathcal{N} \mid \lambda) := \lambda^{N_T}(1-\lambda)^{\sum_{t=1}^{T}\nu_t(\mathcal{N})-N_T}\prod_{t=1}^{T}\mathbb{1}\{\nu_t(\mathcal{N}) \ge x_t\},$$

which implies the following result.

Lemma 2. *Given any trajectory $\mathcal{N}$ and effort process $\boldsymbol{\nu}$, we have*

$$\mathbb{P}_{\boldsymbol{\nu}}\left(\mathcal{N} \mid \overline{\lambda}\right) \ge C^{\sum_{t=1}^{T}\nu_t(\mathcal{N})}\mathbb{P}_{\boldsymbol{\nu}}\left(\mathcal{N} \mid \underline{\lambda}\right). \tag{24}$$

Recall that we let $\hat{\nu}_t(\mathcal{N}) = 0$ for $t > \tau(\mathcal{N})$. Then, in case that condition (23) holds, there must exist a set of trajectories $\mathcal{A}$ such that the following two conditions are simultaneously satisfied:

$$\sum_{\mathcal{N}\in\mathcal{A}} \mathbb{P}_{\hat{\boldsymbol{\nu}}}\left(\mathcal{N} \mid \underline{\lambda}\right) \geq \frac{1}{2}, \text{ and} \tag{25}$$

$$\sum_{t=1}^{T} \hat{\nu}_t(\mathcal{N}) < C\ln T, \ \forall \mathcal{N}\in\mathcal{A}. \tag{26}$$

Inequalities (24), (25) and (26), together with the fact that $x\ln x \geq -1/e$, imply the following result.

Lemma 3. *For any contract and its best-response effort process that satisfies* (23)*, we have*

$$\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{T}(1-\hat{\nu}_t) \mid \bar{\lambda}\right] \geq \frac{1}{2}T^{-1/e}(T - C\ln T). \tag{27}$$

Finally, dropping the second term on the right-hand side of inequality (20), which is non-negative, we obtain

$$\begin{aligned}
\mathrm{Reg}(\Gamma, T) &\geq P_0(\bar{\lambda}R - b)\left(T - \mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{\tau}\hat{\nu}_t \mid \bar{\lambda}\right]\right)\\
&= P_0(\bar{\lambda}R - b)\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\sum_{t=1}^{T}(1-\hat{\nu}_t) \mid \bar{\lambda}\right]\\
&\geq \frac{1}{2}P_0(\bar{\lambda}R - b)T^{-1/e}\left(T - C\ln T\right),
\end{aligned}$$

where the last inequality follows from (27).

In conclusion, we have the following result on the lower bound of the regret.

Theorem 1. *For any contract* Γ*, we have*

$$\mathit{Reg}(\Gamma, T) \geq \min\left\{\frac{1}{2}(1-P_0)(b - \underline{\lambda}R)C\ln T, \ \frac{1}{2}P_0(\bar{\lambda}R - b)T^{-1/e}\left(T - C\ln T\right)\right\} = \Omega(\ln T).$$

In the next section, we present two contracts that achieve $O(\ln T)$ regret rate, which implies that the lower bound rate of Theorem 1 is tight.

4. Contract Design

As discussed in Section 2, even describing a history-*dependent* dynamic contract can be non-trivial, because it may require specifying payment and stopping time under each of the 2^T trajectories of arrivals. Furthermore, the agent may manipulate the principal's belief, which requires the principal to anticipate the agent's effort process and belief. In contrast, focusing on history-*independent* contracts makes the dynamic contract design problem much simpler. We refer to a contract as being history-*independent* if the payment upon an arrival in a period t only depends on t, and not on history $\mathcal{N}_t$; furthermore, whether the agent is terminated in period t only depends on N_t and not the entire history $\mathcal{N}_t$. Interestingly, restricting to history-*independent* contracts does not hurt the regret rate. In this section, we propose two history-*independent* contracts that are both incentive compatible and easy to describe, and both can meet the optimal regret rate stated in Section 3.

4.1. Explore-and-Exploit (EE) Contract with Dynamically Adjusted Payment

We first propose an online dynamic contract following the *optimism in the face of uncertainty* principle. The widely-known Upper Confidence Bound algorithm for MAB problems exemplifies this idea. In that problem, an arm is chosen if either the estimated expected reward is large, or the variance of the reward is large. The general idea behind our proposed contract follows the same vein. That is, we keep paying for the agent's effort if either the arrival record appears strong, or the probability that the agent has been unlucky is still relatively high. In particular, the contract terminates the agent if the empirical arrival rate N_t/t is significantly lower than $\overline{\lambda}$. Specifically, define the following sequence for $t \in [T]$,

$$\epsilon_t := \sqrt{\ln(T-t+1)/t}, \tag{28}$$

and terminate the agent as soon as $N_t < t(\overline{\lambda}-\epsilon_t)$, where $t(\overline{\lambda}-\epsilon_t)$ increases in t. That is, define the following stopping threshold for N_t

$$a_t := t\max\{0, \overline{\lambda}-\epsilon_t\}. \tag{29}$$

Similar to (16), we define the stopping time τ to be

$$\{\tau = t\} := \bigcap_{s=1}^{t-1}\{N_s \geq a_s\} \bigcap \{N_t < a_t\}, \ \forall t = 1, \dots, T-1. \tag{30}$$

The contract also involves paying the agent zero when there is no arrival, and a positive amount $\beta^{(t)}$ when there is an arrival, which only depends on the time period t, regardless of past arrivals. For this purpose, we first define $P^{(t)}$, the belief probability that the agent is of type $\overline{\lambda}$ in period t having received a_t arrivals, as follows

$$P^{(t)} := P_t(a_t), \tag{31}$$

in which function $P_t(\cdot)$ is defined in (14). Condition on the agent not yet terminated by period t (that is, $\tau > t$), this probability serves as a lower bound on the belief probability that the agent is of type $\overline{\lambda}$, due to the fact that $N_t \geq a_t$ when the agent has not been terminated. Then, define the payment $\beta^{(t)}$ as

$$\beta^{(t)} := \beta(P^{(t-1)}) \geq \beta \quad \forall 1 \leq t < \tau, \tag{32}$$

```
Algorithm 1: “Explore-and-exploit” contract
Input: $T$, $P_0$, $\overline{\lambda}$, $\underline{\lambda}$
Initialization: $N_0 = 0$
for $t \leftarrow 1$ to $T$ do
    Set $\epsilon_t$, $a_t$, $P^{(t)}$, and $\beta^{(t)}$ according to Equations (28), (29), (31), and (32)
end
for $t \leftarrow 1$ to $T$ do
    if $x_t = 1$ then
        Pay the agent $\beta^{(t)}$
    end
    Update $N_t = N_{t-1} + x_t$
    if $N_t < a_t$ then
        Terminate the agent
        Break
    end
end
```

where the function $\beta(\cdot)$ is defined in (15). The payment $\beta^{(t)}$ captures the minimum payment to ensure effort when the belief probability of type $\overline{\lambda}$ is $P^{(t-1)}$ in the beginning of period t. Putting all together, Algorithm 1 summarizes our contract dynamics.

Intuitively, the principal specifies a_t as the minimum “acceptable” number of arrivals that the agent must generate up to period t in order to continue. Furthermore, the payment $\beta^{(t)}$ provides sufficient incentive for any myopic agent who has generated at least a_t arrivals up to now to continue exerting effort in period t. The threshold a_t is set to be low enough, which gives an unlucky type $\overline{\lambda}$ agent a chance to continue and prove itself. (An $\overline{\lambda}$ agent lucky enough to have received more than a_t arrivals is over-paid.) On the other hand, the design also needs to ensure that a_t is not too low such that the type $\underline{\lambda}$ agent can be screened out.

Figure 1 illustrates the contract dynamics with two sample trajectories. The trajectory of the total number of arrivals N_t for the capable agent (type $\overline{\lambda}$) is always above the threshold a_t until the end of the time horizon. In comparison, the trajectory of the incapable agent (type $\underline{\lambda}$) hits the threshold a_t before period 20, and is terminated at that point.

Proposition 4 next asserts that the proposed contract is incentive compatible.

Proposition 4. *Let Γ be the contract generated according to Algorithm 1. This contract Γ satisfies* (IC).

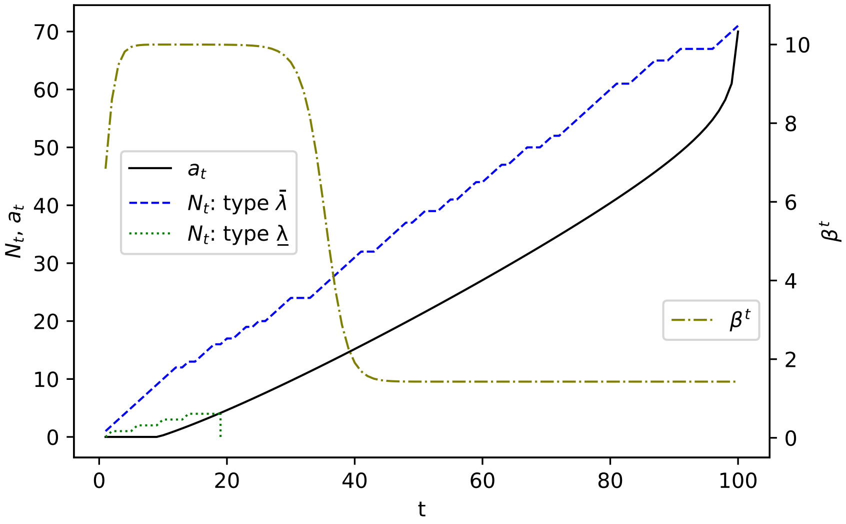

Figure 1 Illustration of the "explore-and-exploit" contract with parameters $\overline{\lambda} = 0.7$, $\underline{\lambda} = 0.1$, $P_0 = 0.2$, $T = 100$, $b = 1$, and $R = 3$. The figure also plots two sample trajectories, one for a capable agent and the other for an incapable agent.

Based on Proposition 4, we know that the agent always exerts effort before termination when facing contract Γ, and both players' belief probabilities that the agent is of type $\overline{\lambda}$ follow (14). The main result of this section is the following theorem.

Theorem 2. *For any instance of the online dynamic contract design problem with T periods, let Γ be a contract as described in Algorithm 1. The total expected regret of implementing contract Γ satisfies*

$$Reg(\Gamma, T) = O(\ln T). \tag{33}$$

Comparing Theorem 2 with Theorem 1, we observe that the proposed contract achieves the optimal regret rate. The complete proof is in the Online Appendix Section EC.3. Here, we outline the key steps of the proof.

Proof outline. First, derived from definitions (5) and (12), we can upper bound the regret by a summation of three separate parts as follows.

$$\mathrm{Reg}(\hat{\Gamma}, T) \le P_0 \overline{\lambda} \sum_{t=1}^{T} (\beta^{(t)} - \beta) + P_0 \left(T - \mathbb{E}\left[\tau \mid \overline{\lambda}\right]\right) \overline{\lambda}(R - \beta) + (1 - P_0)\mathbb{E}\left[\tau \mid \underline{\lambda}\right] \underline{\lambda}(\bar{\beta} - R), \tag{34}$$

in which

$$\bar{\beta} := b/\underline{\lambda}. \tag{35}$$

This result is summarized as Lemma EC.2 and proved in Appendix Section EC.3.2. The first part, $P_0\overline{\lambda}\sum_{t=1}^{T}(\beta^{(t)}-\beta)$, encapsulates the expected "overpayment" if the agent is of type $\overline{\lambda}$; the second part, $P_0\left(T-\mathbb{E}\left[\tau\mid\overline{\lambda}\right]\right)\overline{\lambda}(R-\beta)$, captures the loss from mistakenly terminating the type $\overline{\lambda}$ agent; and the third part, $(1-P_0)\mathbb{E}\left[\tau\mid\underline{\lambda}\right]\underline{\lambda}(\bar{\beta}-R)$, upper bounds the total loss due to hiring the type $\underline{\lambda}$ agent. In order to establish that the regret is indeed $O(\ln T)$, we need to demonstrate that

(1) the payment $\beta^{(t)}$ converges to β fast enough, so that the capable agent is not over-paid by too much;

(2) the type $\overline{\lambda}$ agent is either not mistakenly terminated within T, or kept for long enough before being terminated. Equivalently, the expected number of periods the principal hires is sufficiently large; and

(3) the type $\underline{\lambda}$ agent is terminated fast enough.

To facilitate our discussion, we first define the following constants,

$$\begin{aligned}&\alpha:=\ln\frac{\overline{\lambda}(1-\underline{\lambda})}{\underline{\lambda}(1-\overline{\lambda})}>0,\quad \alpha':=\ln\left[\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{(1-\overline{\lambda})}\right]<0\ ,\\ &c_1:=\left(\frac{\alpha}{\alpha'}\right)^2>0,\quad c_2:=-\frac{\alpha^2}{\alpha'}>0,\quad \text{and } c_3:=\frac{b\left(\overline{\lambda}-\underline{\lambda}\right)}{\overline{\lambda}^2}\left(\frac{1}{P_0}-1\right)\geq 0.\end{aligned} \tag{36}$$

It is noteworthy that these constants are independent of T. In addition, the fact that $\alpha'<0$ is formally verified in Lemma EC.3.

In the next lemma, we establish an upper bound on the overpayment $\beta^{(t)}-\beta$ for any $t\in[T]$. In particular, we can divide the time periods into consecutive and disjoint groups, indexed by k, and bound the per period overpayment for every period within each of the groups.

Lemma 4. *let Γ be a corresponding contract as illustrated in Algorithm 1. For any $k\geq 0$, and $t\in[T]$ such that*

$$t\geq\max\left\{c_1(1+3k),1/\overline{\lambda}^2\right\}\ln T,$$

we have

$$\beta^{(t)}-\beta\leq c_3T^{-c_2k}. \tag{37}$$

The bound provided by Lemma 4 suggests that the upper bounds of overpayments decrease in time. Furthermore, as the length of the planning horizon T increases, the number of periods within each group increases; however, the upper bound of overpayment for a fixed group k decreases in T. Based on Lemma 4, the next Lemma 5 verifies point (1) by showing that the over-payment term towards a type $\overline{\lambda}$ agent, $\sum_{t=1}^{T}(\beta^{(t)}-\beta)$, is indeed upper bounded by $O(\ln T)$.

Lemma 5. *Under contract Γ, for any $T\geq 2$, we have*

$$\sum_{t=1}^{T}(\beta^{(t)}-\beta)\leq\left(c_1+\frac{1}{\overline{\lambda}^2}+\frac{3c_1}{1-T^{-c_2}}\right)c_3\ln T\leq\left(c_1+\frac{1}{\overline{\lambda}^2}+\frac{3c_1}{1-2^{-c_2}}\right)c_3\ln T. \tag{38}$$

Next, both points (2) and (3) follow from concentration inequalities. In particular, conditioning on the type of the agent, the Hoeffding's inequality directly implies the following expressions.

Lemma 6. *For $\epsilon_t \le \overline{\lambda} - \underline{\lambda}$, we have*

$$\begin{aligned} &\mathbb{P}\left(N_t < a_t \mid \overline{\lambda}\right) \le e^{-2t\epsilon_t^2} = \frac{1}{(T-t+1)^2}, \ \textit{and} \\ &\mathbb{P}\left(N_t \ge a_t \mid \underline{\lambda}\right) \le e^{-2t(\overline{\lambda}-\underline{\lambda}-\epsilon_t)^2}. \end{aligned} \tag{39}$$

Because ϵ_t, defined in (28), decreases in t, Lemma 6 suggests that as t increases, the probability of keeping an incapable agent decreases exponentially. Furthermore, Lemma 6 also provides an upper bound for the probability of terminating a capable agent. Following Lemma 6, we have the next lemma, confirming the aforementioned points (2) and (3).

Lemma 7. *For the termination time τ of our contract, we have*

$$\mathbb{E}\left[\tau \mid \overline{\lambda}\right] \ge T - \ln T - 1, \ \textit{and}\ \mathbb{E}\left[\tau \mid \underline{\lambda}\right] \le \frac{4\ln T + 2e/T^2}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}. \tag{40}$$

At last, putting together Lemma 5, Lemma 7, and inequality (34), we can obtain the upper bound on the regret as stated in Theorem 2.

4.2. Explore-then-Commit (ETC) Contract

Now we propose another contract, following a spirit of "explore-then-commit." Specifically, the contract selects a switching time period $\bar{t}$. Before and up to $\bar{t}$, the principal pays a constant $\bar{\beta}$ (defined in (35)) for each arrival to guarantee the agent's effort. We may consider periods 1 to $\bar{t}$ as the "exploration" periods. At period $\bar{t}$, if the belief probability $P_{\bar{t}}$ that the agent is of type $\overline{\lambda}$ falls below a threshold P_a, the agent is terminated. Otherwise, the principal continues with the agent. A crucial difference between the proposed contract and the majority of existing explore-then-commit algorithms in the literature is that, after period $\bar{t}$, whenever the likelihood of the agent being type $\overline{\lambda}$ falls below a threshold P_a ($P_t < P_a$ for any $t > \bar{t}$), the agent is also terminated. Despite this crucial difference, for ease of exposition, we still refer to periods $\bar{t}+1$ to T as the "commitment" periods, and the proposed contract as the "explore-then-commit" contract.

In particular, we define the switching time $\bar{t}$, a threshold a on the number of arrivals up to time $\bar{t}$, the corresponding threshold P_a on the belief probability that the agent is of type $\overline{\lambda}$, and the payment-upon-arrival β_a after $\bar{t}$ as

$$\bar{t} := \left\lceil \max\left\{c_1(1+3/c_2),\ 1/\overline{\lambda}^2\right\}\ln T\right\rceil, \quad a := \left\lceil \overline{\lambda}\bar{t} - \sqrt{\bar{t}\ln T}\right\rceil, \quad P_a := P_{\bar{t}}(a), \quad \text{and } \beta_a := \beta(P_a), \tag{41}$$

respectively, in which $P_{\bar{t}}(a)$ is defined in (14) and $\beta(P_a)$ in (15).

Algorithm 2: "Explore-then-commit" contract

```
Input: $T$, $P_0$, $\overline{\lambda}$, $\underline{\lambda}$
Initialization: Set $N_0 = 0$
Set $\bar{\beta}$, $\bar{t}$, $a$, $P_a$, and $\beta_a$ according to (35) and (41)
for $t \leftarrow 1$ to $\bar{t}$ do
    if $x_t = 1$ then
        Pay the agent $\bar{\beta}$
    end
    Update $N_t = N_{t-1} + x_t$
    Calculate $P_t$ according to (14)
end
if $P_{\bar{t}} < P_a$ then
    Terminate the agent
    Break
end
for $t \leftarrow \bar{t} + 1$ to $T$ do
    if $x_t = 1$ then
        Pay the agent $\beta_a$
    end
    Update $N_t = N_{t-1} + x_t$
    Calculate $P_t$ according to (14)
    if $P_t < P_a$ then
        Terminate the agent
        Break
    end
end
```

Algorithm 2 summarizes the contract dynamics. Figure 2 illustrates the dynamics of the contract and plots two sample trajectories, similar to Figure 1.

We first have the following incentive compatibility result on the proposed "explore-then-commit" contract.

PROPOSITION 5. *The proposed "explore-then-commit" contract satisfies* (IC).

Consequently, in the following analysis, we have that the agent always exerts effort before termination, and both players' belief probabilities that the agent is of type $\overline{\lambda}$ follow (14). We can derive the following main result on the upper bound of the expected total regret.

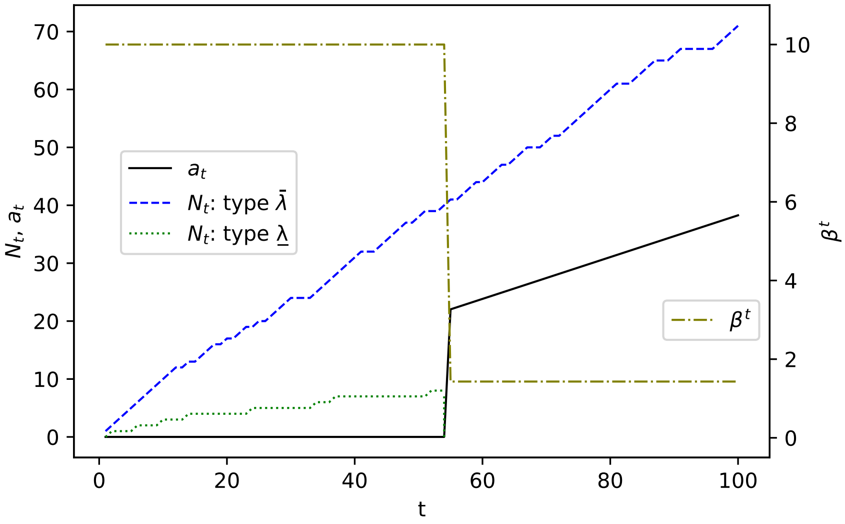

Figure 2 Illustration of the "explore-then-commit" contract with parameters $\overline{\lambda} = 0.7$, $\underline{\lambda} = 0.1$, $P_0 = 0.2$, $T = 100$, $b = 1$, and $R = 3$. The figure also plots two sample trajectories, one for a capable agent and the other for an incapable agent.

Theorem 3. *For any instance of the online dynamic contract design problem with T periods, let Γ be an "explore-then-commit" contract as defined in Algorithm 2. The total expected regret can be upper bounded as*

$$Reg(\Gamma, T) = O(\ln T). \tag{42}$$

Theorem 3 suggests that the proposed "explore-then-commit" contract also achieves the optimal regret rate. Next, we outline the proof of Theorem 3. Similar to that of Theorem 2, we can decompose the total expected regret and bound it from above by three parts. Specifically,

$$\text{Reg}(\Gamma; T) \le P_0\overline{\lambda}\left[(\bar{\beta} - \beta)\bar{t} + (\beta_a - \beta)(T - \bar{t})\right] + P_0 T\overline{\lambda}(R - \beta)\mathbb{P}\left(\tau < T \mid \overline{\lambda}\right) + (1 - P_0)\underline{\lambda}(\bar{\beta} - R)\bar{t}. \tag{43}$$

where the first part encapsulates the "overpayment" if the agent is of type $\overline{\lambda}$, and the overpayment consists of those incurred during and after the exploration periods; the second part captures the loss from mistakenly terminating the type $\overline{\lambda}$ agent before T; the third part upper bounds the total lost of hiring a type $\underline{\lambda}$ agent. We observe that the third term of (43), corresponding to the loss of hiring a type $\underline{\lambda}$ agent, is linear in the length of the exploration period, which is logarithmic in T. Therefore, this part of the regret is readily in $O(\ln T)$. In order to establish that the total regret is indeed logarithmic in T, we need to demonstrate that (1) the "overpayment" is sufficiently small; and (2) the probability of type $\overline{\lambda}$ agent being terminated before T is sufficiently small. For (1), the setup of the "explore-then-commit" contract implies that $(\bar{\beta}-\beta)\bar{t}$ is at most logarithmic in T. Therefore, we only need to show that $\beta_a-\beta$ is sufficiently small. For (2), we need to show that the probability of terminating a type $\overline{\lambda}$ agent prematurely is sufficiently low.

We first bound the overpayment. The following Lemma 8 suggests that the overpayment $\beta_a-\beta$ is bounded from above by a multiple of the inverse of T.

Lemma 8. *We have*

$$\beta_a-\beta \le \frac{c_3}{T}. \tag{44}$$

Therefore, $(\beta_a-\beta)(T-\bar{t})$ in the second term of (43) is in $O(1)$. Next, for point (2), we invoke the Hoeffding's inequality to obtain the following result.

Lemma 9. *The probability of terminating a capable agent before T satisfies*

$$\mathbb{P}\left(\tau < T \,|\, \overline{\lambda}\right) \le \frac{1}{T}. \tag{45}$$

Combining (43), (44) and (45), we have

$$\mathrm{Reg}(\Gamma;T) \le P_0\overline{\lambda}\left[(\bar{\beta}-\beta)\bar{t}+c_3(1-\bar{t}/T)\right]+P_0\overline{\lambda}(R-\beta)+(1-P_0)\underline{\lambda}(\bar{\beta}-R)\bar{t}=O(\ln T),$$

where the last equality follows because $\bar{t}=O(\ln T)$.

We conclude this section by comparing the two proposed contracts. Theoretically, both regret upper bounds of the "explore-and-exploit" contract described in Section 4.1 and the "explore-then-commit" contract in Section 4.2 are in the same order, and differ only in the constant terms. From a practical perspective, the "explore-then-commit" contract appears easier to implement, because it does not require the principal to frequently vary payments towards arrivals. We numerically compare the regrets from the "explore-and-exploit" and the "explore-then-commit" contracts. Figure 3 illustrates the result. Here, we set the model parameters $b=1$, $R=3$, $P_0=0.5$, $\overline{\lambda}=0.8$, and $\underline{\lambda}=0.3$, and vary the time horizon T. For each choice of T, we generate $10,000$ samples. Within each sample, we first randomly generate the agent's type $\lambda\in\{\overline{\lambda},\underline{\lambda}\}$ according to the common prior distribution. Then in each period $t\in[T]$, the simulation generates a sample trajectory of x_t's according to Bernoulli(λ). Figure 3 clearly demonstrates that the "explore-and-exploit" contract achieves a lower regret.

5. Continuous-Time Setting

In this section, we extend the previous analysis to a continuous-time setting, which connects our work with prior literature on continuous-time dynamic contracting with Poisson arrivals (see, for example, Biais et al. 2010, Sun and Tian 2018, Cao et al. 2023).

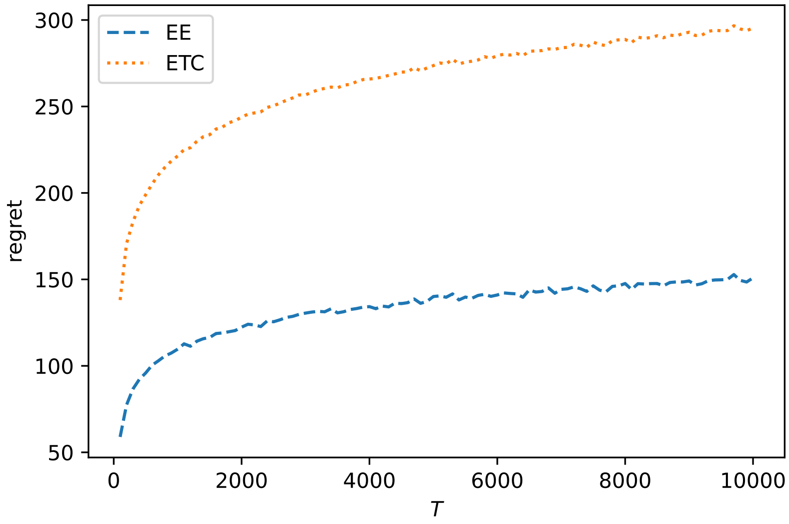

Figure 3 Average regrets of the "explore-and-exploit" and the "explore-then-commit" contracts over different time horizons T under parameter $\overline{\lambda} = 0.8$, $\underline{\lambda} = 0.3$. The horizon T takes values of $100, 200, \ldots, 10,000$, and for each T we ran $10,000$ samples.

Consider a time horizon of length $T \in \mathbb{R}_+$ with N arrivals. Define the corresponding sequence of arrival times as $\{T_n\}_{n\in\{0,1,\ldots,N\}} \in [0,T)$ such that $T_0 = 0 < T_1 < T_2 < \ldots < T_N < T_{N+1} := T$. Define a counting process $\{N_t\}$ such that

$$N_t = n \quad \text{for } t \in [T_n, T_{n+1}),\ n \in \{0, \ldots, N\},$$

which represents the total number of arrivals before time t. The counting process generates a filtration $\mathcal{N} := \{\mathcal{N}_t\}_{t\in[0,T)}$. The arrivals are generated according to a Poisson process with rate being either $\overline{\lambda}$ or $\underline{\lambda}$, depending on the agent's type, and if the agent exerts effort. The agent's effort is not observable by the principal, and the instantaneous arrival rate is 0 when the agent does not exert effort. The ex-ante probability of the agent being of type $\overline{\lambda}$ is P_0. Therefore, given a time horizon T, the number of arrivals N is random.

Now, we define a contract consisting of payments and a termination time. Let $L := \{L_t\}_{t\in[0,T)}$ be an $\mathcal{N}$-adapted process tracking the principal's cumulative payment to the agent, with $L_0 = 0$ and

$$L_t - L_{t'} \geq 0,\ \forall t, t' \in [0,T) \text{ and } t \geq t',$$

which corresponds to $\beta_t \geq 0$ in the discrete-time case and captures the agent's limited liability. Define the termination time τ as an $\mathcal{N}$-stopping time. In response to a contract $\Gamma := (L, \tau)$, the agent exerts effort according to an $\mathcal{N}$-predictable effort process $\boldsymbol{\nu} := \{\nu_t\}_{t\in[0,T)}$, with $\nu_t \in \{0,1\}$. Given a contract Γ and an effort process $\boldsymbol{\nu}$, the expected utility of the agent is

$$W(\Gamma,\boldsymbol{\nu}) := \mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^{\tau} \mathrm{d}L_t - b\nu_t \mathrm{d}t\right]. \tag{46}$$

The corresponding principal's utility is defined as

$$U(\Gamma,\boldsymbol{\nu}) := \mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^{\tau} R\mathrm{d}N_t - \mathrm{d}L_t\right]. \tag{47}$$

For any contract Γ, still let $\hat{\boldsymbol{\nu}}(\Gamma)$ be the agent's best response effort process, such that (3) is satisfied. The principal's dynamic contract design problem is still (6) with $W(\Gamma,\boldsymbol{\nu})$ and $U(\Gamma,\boldsymbol{\nu})$ defined in (46) and (47), respectively. The upper bound expression in Proposition 1 still holds, and the regret of a contract Γ is as defined in (12).

If the agent of type λ always exerts effort, then the number N_t of good arrivals up to time t follows a Poisson distribution with parameter λt. The following result corresponds to Lemma 1, which summarizes the two players' share belief about the agent's type if the agent always exerts effort.

Lemma 10. *After any time t with the full-effort process and history $\mathcal{N}_t$ such that the number of good arrivals is N_t, the posterior belief that the agent is of type $\overline{\lambda}$, denoted by $P_t(N_t)$, is*

$$P_t(N_t) := \mathbb{P}\left(\lambda = \overline{\lambda} \,|\, \mathcal{N}_t\right) = \mathbb{P}\left(\lambda = \overline{\lambda} \,|\, N_t\right) = \left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t} \cdot e^{(\overline{\lambda}-\underline{\lambda})t} \cdot \left(\frac{1}{P_0} - 1\right) + 1\right)^{-1}. \tag{48}$$

5.1. Regret Lower Bound in Continuous-Time Setting

Similar to inequality (20) of Section 3, for any contract Γ and the corresponding best-response effort process $\hat{\boldsymbol{\nu}}$, we have

$$\operatorname{Reg}(\Gamma,T) \geq P_0\overline{\lambda}(R-\beta)\left(T - \mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=0}^{\tau} \hat{\nu}_t \mathrm{d}t \,|\, \overline{\lambda}\right]\right) + (1-P_0)(b-\underline{\lambda}R)\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=0}^{\tau} \hat{\nu}_t \mathrm{d}t \,|\, \underline{\lambda}\right]. \tag{49}$$

We consider two cases depending on whether a type $\underline{\lambda}$ agent exerts enough effort under contract Γ. First, suppose that the $\underline{\lambda}$-type agent exerts effort in a sufficiently long time. That is,

$$\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=0}^{\tau} \hat{\nu}_t \mathrm{d}t \,|\, \underline{\lambda}\right] \geq \frac{1}{2} C \ln T, \tag{50}$$

for a constant $C := 1/[2(\overline{\lambda}-\underline{\lambda})]$. Dropping the first term on the right-hand side of inequality (49), which is non-negative, we obtain

$$\operatorname{Reg}(\Gamma,T) \geq \frac{1}{2}(1-P_0)(b-\underline{\lambda}R)C\ln T. \tag{51}$$

Now suppose that the opposite of (50) holds. That is,

$$\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=0}^{\tau} \hat{\nu}_t \mathrm{d}t \mid \underline{\lambda}\right] < \frac{1}{2} C \ln T. \tag{52}$$

We will prove that when (52) holds, the first term on the right-hand side of (49) is at least in the order of $\ln T$.

We start with introducing additional notations. Note that for a generic Poisson process with rate λ, the joint distribution of observing n arrivals over a time interval $[0, T)$ with arrival time epochs $\{t_i\}_{i=1}^n$ such that $0 \le t_1 < t_2 < \ldots, < t_n < T$ is

$$f(n, t_1, t_2, \ldots, t_n; \lambda) = \lambda^n e^{-\lambda T}. \tag{53}$$

Online Appendix Section EC.4.2 provides a detailed derivation for (53). Based on this expression, we consider our setting with an effort process $\boldsymbol{\nu}$ and a trajectory $\mathcal{N}_t$ over the time horizon $[0, T)$. Let n be the total number of arrivals with corresponding arrival time epochs $\{t_i\}_{i=1}^n$, such that $0 \le t_1 < \ldots < t_n < t$ and $\nu_{t_i}(\mathcal{N}_t) = 1$. Further, define the corresponding "effective total effort time" as $\mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_t) := \int_{s=0}^{t} \nu_s(\mathcal{N}_s)\mathrm{d}s$. Following from (53), its mixed density function is

$$f_{\boldsymbol{\nu}}(n, t_1, \ldots, t_n; \lambda, t) = \begin{cases} \lambda^n e^{-\lambda \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_t)}, & \text{if } \nu_{t_i}(\mathcal{N}_t) = 1, \forall i \in [n], \\ 0, & \text{otherwise.} \end{cases} \tag{54}$$

Because the joint distributions do not depend on specific values of $t_1, \ldots, t_n$, for notation clarity, we define

$$f_{\boldsymbol{\nu}}(\mathcal{N}_t \mid \lambda) := f_{\boldsymbol{\nu}}(n, t_1, \ldots, t_n; \lambda, t). \tag{55}$$

Besides, for any real value t, integer n and a sequence $t_1, \ldots, t_n$, we use expression $t_1, \ldots, t_n \in [0..t)$ to represent the condition that $0 \le t_1 \le \ldots \le t_n < t$. With a slight abuse of notation, for any function $g(n, t_1, \ldots, t_n)$, we define the conditional expectation of function $g(\cdot)$ by the following simplified expression

$$\int g(\mathcal{N}_t) f_{\boldsymbol{\nu}}(\mathcal{N}_t \mid \lambda) \mathrm{d}\mathcal{N}_t := \sum_{n=0}^{\infty} \int \cdots \int_{t_1, \ldots, t_n \in [0..t)} g(n, t_1, \ldots, t_n) f_{\boldsymbol{\nu}}(n, t_1, \ldots, t_n; \lambda, t) \mathrm{d}t_1 \ldots \mathrm{d}t_n. \tag{56}$$

Similar to Lemma 2, we obtain the next result following (53).

Lemma 11. *Given any trajectory $\mathcal{N}$ and effort process $\boldsymbol{\nu}$, we have*

$$f_{\boldsymbol{\nu}}(\mathcal{N} \mid \overline{\lambda}) \ge e^{-(\overline{\lambda} - \underline{\lambda}) \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N})} f_{\boldsymbol{\nu}}(\mathcal{N} \mid \underline{\lambda}). \tag{57}$$

In the same spirit of (25) and (26), condition (52) implies that there must exist an event $\mathcal{A}$ such that the following two conditions are simultaneously satisfied:

$$\int_{\mathcal{N} \in \mathcal{A}} f_{\hat{\boldsymbol{\nu}}}(\mathcal{N} \mid \underline{\lambda}) \mathrm{d}\mathcal{N} \ge \frac{1}{2}, \text{ and } \mathcal{T}_{\hat{\boldsymbol{\nu}}}(\mathcal{N}) < C \ln T, \text{ a.e. in } \mathcal{A}. \tag{58}$$

Inequalities (57) and (58) imply the following result.

Lemma 12. *For any contract and its best-response effort process that satisfies* (52), *we have*

$$\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=0}^{T}(1-\hat{\nu}_t)\mathrm{d}t \mid \overline{\lambda}\right] \geq \frac{1}{2}T^{-1/2}(T - C\ln T). \tag{59}$$

Finally, dropping the second term on the right-hand side of inequality (49), which is non-negative, we obtain

$$\begin{aligned}\mathrm{Reg}(\Gamma,T) &\geq P_0\overline{\lambda}(R-\beta)\left(T - \mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=0}^{\tau}\hat{\nu}_t\mathrm{d}t \mid \overline{\lambda}\right]\right)\\ &= P_0(\overline{\lambda}R-b)\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=1}^{T}(1-\hat{\nu}_t)\mathrm{d}t \mid \overline{\lambda}\right]\\ &\geq \frac{1}{2}P_0(\overline{\lambda}R-b)T^{-1/2}\left(T - C\ln T\right),\end{aligned}$$

where the last inequality follows (59).

In conclusion, we have the following result, which lower bounds the regret.

Theorem 4. *For any contract* Γ, *we have*

$$\mathit{Reg}(\Gamma,T) \geq \min\left\{\frac{1}{2}(1-P_0)(b-\underline{\lambda}R)C\ln T,\ \frac{1}{2}P_0(\overline{\lambda}R-b)T^{-1/2}\left(T-C\ln T\right)\right\} = \Omega(\ln T).$$

The result demonstrates that extending the online dynamic contract design problem into a continuous-time setting does not affect the order of the lower regret bound in the length of the planning horizon, T.

5.2. Contract Design in Continuous-Time Setting

Here we provide a continuous-time contract design, in analogous to the discrete-time "explore-and-exploit" contract of Section 4.1. Following the same spirit of (28), define[1]

$$\varepsilon_t := (e-1)\sqrt{\frac{2\overline{\lambda}\ln(T)}{t}}.$$

With a slight abuse of notations, define a_t in the same way as in (29), i.e., $a_t := t\max\{0, \overline{\lambda} - \varepsilon_t\}$. The agent is terminated as soon as $N_t < a_t$. Similar to (30) we define the termination time to be

$$\{\tau = t\} := \cap_{s\in[0,t)}\{N_s \geq a_s\} \cap \{N_t < a_t\},\ \forall t \in [0,T). \tag{60}$$

Because N_t's are integers and increasing in t, whereas a_t is non-decreasing deterministic function of t, the actual stopping time can only take values from a finite set. In particular, for any integer $n \in [\lfloor a_T \rfloor]$, define time epoch s_n such that

$$a_{s_n} = n.$$

[1] The reason that we have a constant factor $e-1$ in ε_t is due to the concentration inequality that we use here for the continuous-time case. Inside the square-root term, we use $\ln T$ instead of $\ln(T-t+1)$ because the expression in (28) creates issues when time epoch t gets too close to T.

It may be easier to understand the termination time in terms of threshold times to receive certain numbers of arrivals. That is, the agent can only be terminated at some s_n, if the n-th arrival does not occur by time s_n, or,

$$\{\tau > s_n\} := \cap_{j\in[n]}\{T_j \leq s_j\}, \ \forall n \in [\lfloor a_T \rfloor]\,. \tag{61}$$

The continuous-time contract also pays the agent zero when there is no arrival, and a positive amount $\beta^{(t)}$ when there is an arrival. The payment only depends on the time period t, regardless of past arrivals. For this purpose, similar to (31), define probability

$$P^{(t)} := \left(\left(\frac{\underline{\lambda}}{\overline{\lambda}} \right)^{a_t} \cdot e^{(\overline{\lambda}-\underline{\lambda})t} \cdot \left(\frac{1}{P_0} - 1 \right) + 1 \right)^{-1}, \tag{62}$$

and the corresponding payment

$$\beta^{(t)} := \beta(P^{(t)}), \tag{63}$$

where the function $\beta(\cdot)$ is defined in (15). Figure 4 illustrates the dynamics of the continuous-time contract, similar to Figure 1.

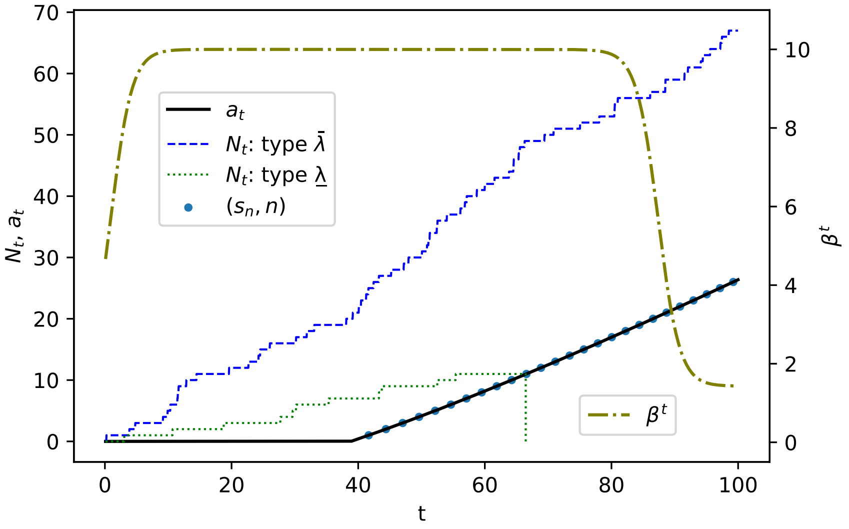

Figure 4 Illustration of our algorithm under the continuous time version with parameters $\overline{\lambda} = 0.7$, $\underline{\lambda} = 0.1$, $P_0 = 0.1$, $T = 100$, $b = 1$, and $R = 3$. The marked points (s_n, n) represent all the epochs that the agent is possible to be terminated and the corresponding $a_{s_n} = n$.

We have the following Proposition 6, which establishes that the proposed contract is incentive compatible.

PROPOSITION 6. *Consider a contract Γ that terminates according to τ defined in (60), pays the agent $\beta^{(t)}$ if there is an arrival at time t, and zero if not. That is, $\mathrm{d}L_t = \beta^{(t)}\mathrm{d}N_t$. This contract Γ satisfies (IC), in which the full-effort process in the continuous time is defined as*

$$\bar{\boldsymbol{\nu}} := \{\nu_t = 1\}_{t\in[0,T)}.$$

Therefore, the agent's best-response effort process is $\bar{\boldsymbol{\nu}}$, and both players' belief probabilities that the agent is of type $\overline{\lambda}$ follow (48).

Similar to (34), we can divide the regret into three parts, that is

$$\begin{aligned}\mathrm{Reg}(\Gamma,T) = {} & P_0\left(\overline{\lambda}(R-\beta)T - \mathbb{E}_{\hat{\boldsymbol{\nu}}(\Gamma)}\left[\int_0^{\tau}(R\mathrm{d}N_t - \mathrm{d}L_t)\mid\overline{\lambda}\right]\right) - (1-P_0)\mathbb{E}_{\hat{\boldsymbol{\nu}}(\Gamma)}\left[\int_0^{\tau}(R\mathrm{d}N_t - \mathrm{d}L_t)\mid\underline{\lambda}\right] \\ & \le P_0\overline{\lambda}\int_0^T(\beta^{(t)}-\beta)\mathrm{d}t + P_0\left(T-\mathbb{E}\left[\tau\mid\overline{\lambda}\right]\right)\overline{\lambda}(R-\beta) + (1-P_0)\mathbb{E}\left[\tau\mid\underline{\lambda}\right]\underline{\lambda}(\bar{\beta}-R). \qquad (64)\end{aligned}$$

Again, the three parts correspond to the overpayment to a type $\overline{\lambda}$ agent, the loss from mistakenly terminating a type $\overline{\lambda}$ agent, and the loss due to hiring the type $\underline{\lambda}$ agent for too long. In order to establish an upper bound on the total regret, we need to bound each of these three parts. First, because Poisson distribution is a sub-exponential distribution, we have the following results corresponding to Lemmas 6 and 7.

LEMMA 13. *For any $t \in [0,T)$, we have*

$$\begin{aligned}&\mathbb{P}\left(N_t < a_t \mid \overline{\lambda}\right) \le e^{-t\varepsilon_t^2/((e-1)^2\overline{\lambda})} = \frac{1}{T^2},\ \textit{for } \varepsilon_t \le \overline{\lambda},\ \textit{and} \\ &\mathbb{P}\left(N_t \ge a_t \mid \underline{\lambda}\right) \le e^{-t(\overline{\lambda}-\underline{\lambda}-\varepsilon_t)^2/((e-1)^2\underline{\lambda})},\ \textit{for } \varepsilon_t \le \min\{\overline{\lambda}-\underline{\lambda},(e-1)\underline{\lambda}\}.\end{aligned} \qquad (65)$$

The insights revealed by Lemma 13 are similar to those discussed following Lemma 6. Lemma 13 further implies the following Lemma 14, confirming that the second and third terms on the right-hand side of inequality (64) are at most in the order of $\ln T$.

LEMMA 14. *We have*

$$\mathbb{E}\left[\tau\mid\overline{\lambda}\right] \ge T-\overline{\lambda},\ \textit{and}\ \mathbb{E}\left[\tau\mid\underline{\lambda}\right] \le \alpha''\ln T + 1/T, \qquad (66)$$

where $\alpha'' := \max\{8(e-1)^2/[\overline{\lambda}(\overline{\lambda}-\underline{\lambda})^2],\ 2/[\overline{\lambda}\underline{\lambda}^2]\}$.

Similar to Lemma 4 and Lemma 5, we have the following two lemmas for bounding the first term in (64).

LEMMA 15. *For any $k = 1,2\ldots$, if $t \ge (c_2'/c_1')^2(1+3k)\ln T$, we have*

$$\beta^{(t)} - \beta \le c_3 e^{c_3' k \ln T}, \qquad (67)$$

where c_3 is defined in (36), and c_1', c_2', c_3' are negative constants defined as follows:

$$c_1' := \ln\left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}} e^{(\overline{\lambda}-\underline{\lambda})}\right),\ c_2' := 2(e-1)\sqrt{\overline{\lambda}}\ln\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right),\ c_3' := \frac{c_2'^2}{c_1'}.$$

Then, following from Lemma 15, we obtain the next result verifying that the total overpayment is in the order of $\ln T$.

Lemma 16. *For any $T \geq e$, we have*

$$\int_{t=0}^{T} (\beta^{(t)} - \beta)\mathrm{d}t \leq 4c_3 \left(\frac{c_2'}{c_1'}\right)^2 \ln T + \frac{3c_3}{-c_3'}. \tag{68}$$

Putting all of the above results together, we obtain the main result of this section, corresponding to Theorem 2, as the following theorem.

Theorem 5. *For any instance of the online dynamic contract design problem in continuous time with a time horizon of length T, let Γ be the contract designed for the continuous time situation, that terminates according to τ defined in (60), and pays the agent $\beta^{(t)}$ according to (63) if and only if there is an arrival at time t. The total expected regret of implementing contract Γ satisfies*

$$Reg(\Gamma, T) = O(\ln T).$$

Comparing Theorem 5 with Theorem 4, we observe that the proposed continuous-time contract achieves the optimal regret rate. This regret rate mirrors the one identified in the discrete-time setting. In the continuous-time setting, we can interpret the termination criterion as a sequence of "milestone times" $\{s_n\}_{n\geq 1}$. An agent is allowed to continue only if the n-th arrival has occurred by time s_n. In certain practical settings, these milestone times may be easier to articulate and understood than the threshold a_t.

6. Conclusion

In this paper, we study a dynamic moral hazard problem in which a principal hires an agent over a finite time horizon with length T. In each period the agent can choose to either shirk or exert costly effort, which is not observable by the principal. The agent's effort generates an uncertain arrival in any period that is beneficial to the principal. The probability of arrival in a period when the agent exerts effort can be one of two values. The high arrival probability corresponds to an agent who is worth hiring, while a low arrival probability agent is not worth hiring. The agent's type is unknown to both the principal and the agent beyond a common prior, and can be learned over time. The profit-maximizing principal can commit to a long-term contract consisting of payments and contract termination.

Due to its complexity, we resort to a regret-minimization approach. In particular, we propose two online dynamic contracts that motivate the agent to exert effort before termination. Moreover, we show that the regret of both contracts is at most in the order of $\ln T$, matching the rate of a regret lower bound. That is, our contracts achieve the best possible regret rate. Our contracts are simple to describe and easy to implement, because the corresponding payment upon arrival in a period only depends on the period index, not on any other information regarding past arrivals. Furthermore, we obtained similar results for a continuous-time setting.

We conclude this paper with some thoughts on potential future research directions. First, in this paper, we assume that the agent is either worth hiring or not so. In general, both types may be worth hiring. However, such a setting appears difficult to analyze, and naïvely extending the proposed contracts of this paper does not seem to work. If solvable, its solution may further allow us to consider the multi-type agent case. Second, our paper assumes that when the agent shirks, the arrival probability is zero. A natural extension is to include a "background arrival rate" even without the agent's effort. In such a setting, if we assume that both players know the agent's type, the corresponding principal's utility may not be an upper bound of the original problem. This is because the agent's information rent may hurt the principal in general. (In our paper with no background arrivals, this information rent happens to be zero.) On the other hand, the societal utility, when assuming that both players know the agent's type, is indeed an upper bound of the principal's utility in the original setting. However, even the regret-rate lower bound against such a benchmark appears linear in the number of periods. Therefore, it is unclear whether and how one can define a proper upper bound against which we can obtain a sub-linear regret rate. Third, it may be interesting to study a multi-agent case with a budget constraint. One can perceive the budget as the amount of resources/jobs to be assigned to agents in each period. The principal can use payment schedules, termination criteria, and budget allocations to incentivize the agent to exert effort and hedge against the risk of facing an incapable agent. The principal may be able to leverage the competition among multiple agents to reduce rent payment.

References

Amin K, Rostamizadeh A, Syed U (2013) Learning prices for repeated auctions with strategic buyers. *Advances in Neural Information Processing Systems*, volume 26 (Curran Associates, Inc.).

Amin K, Rostamizadeh A, Syed U (2014) Repeated contextual auctions with strategic buyers. *Advances in Neural Information Processing Systems*, volume 27 (Curran Associates, Inc.).

Bennett G (1962) Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association* 57(297):33–45.

Biais B, Mariotti T, Rochet JC, Villeneuve S (2010) Large risks, limited liability, and dynamic moral hazard. *Econometrica* 78(1):73–118.

Bubeck S (2011) Introduction to online optimization. *Lecture notes* 2:1–86.

Cao P, Sun P, Tian F (2023) Punish underperformance with suspension: Optimal dynamic contracts in the presence of switching cost. *Management Science* .

Dawande M, Janakiraman G, Qi A, Wu Q (2019) Optimal incentive contracts in project management. *Production and Operations Management* 28(6):1431–1445.

Demarzo PM, Sannikov Y (2017) Learning, termination, and payout policy in dynamic incentive contracts. *The Review of Economic Studies* 84(1 (298)):182–236, ISSN 00346527, 1467937X.

Fernandes A, Phelan C (2000) A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory* 91(2):223–247.

Gibbons R, Murphy KJ (1992) Optimal incentive contracts in the presence of career concerns: theory and evidence. *Journal of Political Economy* 100(3):468–505.

Grossman SJ, Hart OD (1983) An analysis of the principal-agent problem. *Econometrica* 51(1):7–45.

Gupta S, Chen W, Dawande M, Janakiraman G (2023) Three years, two papers, one course off: optimal nonmonetary reward policies. *Management Science* 69(5):2852–2869.

Hazan E (2016) Introduction to online convex optimization. *Foundations and Trends® in Optimization* 2(3-4):157–325, ISSN 2167-3888.

He Z, Wei B, Yu J, Gao F (2017) Optimal long-term contracting with learning. *The Review of Financial Studies* 30(6):2006–2065, ISSN 0893-9454, 1465-7368.

Holmström B (1979) Moral hazard and observability. *The Bell Journal of Economics* 10(1):74–91, ISSN 0361-915X.

Holmström B (1999) Managerial incentive problems: a dynamic perspective. *The Review of Economic Studies* 66(1):169–182.

Lattimore T, Szepesvári C (2020) *Bandit Algorithms* (Cambridge University Press), 1 edition, ISBN 978-1-108-57140-1 978-1-108-48682-8.

Prat J, Jovanovic B (2014) Dynamic contracts when agent's quality is unknown. *Theoretical Economics* 9:865–914.

Rogerson WP (1985) Repeated moral hazard. *Econometrica* 53(1):69–76.

Sannikov Y (2008) A Continuous-Time Version of the Principal: Agent Problem. *The Review of Economic Studies* 75(3):957–984, publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Shalev-Shwartz S (2012) Online learning and online convex optimization. *Foundations and Trends® in Machine Learning* 4(2):107–194, ISSN 1935-8237, 1935-8245, publisher: Now Publishers, Inc.

Slivkins A, et al. (2019) Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning* 12(1-2):1–286.

Spear SE, Srivastava S (1987) On repeated moral hazard with discounting. *The Review of Economic Studies* 54(4):599, ISSN 00346527.

Sun P, Tian F (2018) Optimal contract to induce continued effort. *Management Science* 64(9):4193–4217.

Zhao X, Zhu R, Haskell WB (2022) Learning to price supply chain contracts against a learning retailer. *SSRN Electronic Journal* .

Zhu B, Bates S, Yang Z, Wang Y, Jiao J, Jordan MI (2023) The sample complexity of online contract design. *Proceedings of the 24th ACM Conference on Economics and Computation*, 1188–1188 (London United Kingdom: ACM), ISBN 9798400701047.

Zorc S, Tsetlin I, Hasija S, Chick SE (2023) The when and how of delegated search. *Operations Research* .

Electronic Companions

EC.1. Proofs and supplemental materials of Section 2

EC.1.1. Proof of Proposition 1

Proposition 1. *The principal's expected utility under the optimal contract for the dynamic contract design problem, $\mathcal{J}$, satisfies*

$$\mathcal{J} \le \max_{\Gamma} V\big(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma)\big) \le P_0 \cdot OPT(\overline{\lambda}) + (1-P_0)\cdot OPT(\underline{\lambda}) = P_0 T \overline{\lambda}(R-\beta), \tag{10}$$

in which we define

$$\beta := \frac{b}{\overline{\lambda}}. \tag{11}$$

Proof: The first inequality follows directly from (8). Following (2), (5), and (7) we have that for any Γ and $\boldsymbol{\nu}$,

$$\begin{align}
V(\Gamma,\boldsymbol{\nu}) &= \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} \beta_t - b\nu_t\right] + \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} Rx_t - \beta_t\right] \nonumber\\
&= \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} Rx_t - b\nu_t\right] \nonumber\\
&= P_0\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} Rx_t - b\nu_t \mid \overline{\lambda}\right] + (1-P_0)\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} Rx_t - b\nu_t \mid \underline{\lambda}\right] \nonumber\\
&= P_0\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} \mathbb{E}\left[Rx_t - b\nu_t \mid \nu_t = 1, \overline{\lambda}\right] \mathbb{1}\left\{\nu_t = 1\right\} \mid \overline{\lambda}\right] \nonumber\\
&\quad + (1-P_0)\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} \mathbb{E}\left[Rx_t - b\nu_t \mid \nu_t = 1, \underline{\lambda}\right] \mathbb{1}\left\{\nu_t = 1\right\} \mid \underline{\lambda}\right] \tag{EC.1.1}\\
&= P_0\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} (R\overline{\lambda} - b)\nu_t \mid \overline{\lambda}\right] + (1-P_0)\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{t=1}^{\tau} (R\underline{\lambda} - b)\nu_t \mid \underline{\lambda}\right] \tag{EC.1.2}\\
&\le P_0 T(R\overline{\lambda} - b) + 0, \tag{EC.1.3}
\end{align}$$

where (EC.1.1) uses the fact that $x_t = 0$ conditioning on $\nu_t = 0$; (EC.1.2) follows from $\mathbb{1}\left\{\nu_1 = 1\right\} = \nu_t$ and $\mathbb{E}\left[x_t \mid \nu_t = 1, \lambda\right] = \lambda$; and (EC.1.3) follow from assumption (1), which implies that $R\overline{\lambda} - b > 0$ and $R\underline{\lambda} - b < 0$. Thus we have

$$\max_{\Gamma} V(\Gamma, \hat{\boldsymbol{\nu}}(\Gamma)) \le \max_{\Gamma,\boldsymbol{\nu}} V(\Gamma,\boldsymbol{\nu}) \le \ P_0 T(R\overline{\lambda} - b).$$

Then by taking into the definition of $\mathrm{OPT}(\lambda)$ and β, we complete the proof. $\square$

EC.1.2. The clairvoyant problems and Proposition EC.1

In this section, we define the clairvoyant problem where both the agent and the principal know the exact type of the agent, then we show that the optimal expected profit of the clairvoyant problem achieves the benchmark of the expected profit in the online version specified in Proposition 1.

When both the agent and the principal know exactly the type λ of the agent, following the same spirit of (2)-(6), define the type λ agent's total utility under contract Γ following effort process ν as

$$W^{\lambda}(\Gamma,\nu) := \mathbb{E}_{\nu}\left[\sum_{t=1}^{\tau}\beta_t - b\nu_t \mid \lambda\right];$$

For any contract Γ, define $\hat{\nu}^{\lambda}(\Gamma)$ to be the agent's *best response effort process* with type λ, such that

$$W^{\lambda}(\Gamma,\hat{\nu}^{\lambda}(\Gamma)) \geq W^{\lambda}(\Gamma,\nu),\ \forall \nu, \tag{EC.1.4}$$

and we must also have

$$W^{\lambda}(\Gamma,\hat{\nu}^{\lambda}(\Gamma)) \geq 0; \tag{EC.1.5}$$

Further define the principal's utility given type λ under contract Γ and effort process ν as

$$U^{\lambda}(\Gamma,\nu) := \mathbb{E}_{\nu}\left[\sum_{t=1}^{\tau}Rx_t - \beta_t \mid \lambda\right]; \tag{EC.1.6}$$

The principal tries to solve the following dynamic contract design problem for $\lambda \in \{\overline{\lambda},\underline{\lambda}\}$:

$$\mathcal{J}^{\lambda} := \max_{\Gamma} U^{\lambda}(\Gamma,\hat{\nu}^{\lambda}(\Gamma)). \tag{EC.1.7}$$

Next, we show that $\mathrm{OPT}(\bar{\lambda})$ and $\mathrm{OPT}(\underline{\lambda})$ equal to the principal's optimal utility when both parties know that the agent's type is $\bar{\lambda}$ and $\underline{\lambda}$, correspondingly..

Proposition EC.1. *We have*

$$\mathcal{J}^{\lambda} = OPT(\lambda),$$

where $OPT(\lambda)$ is defined in (9) *.*

Proof: Define the societal utility of a contract Γ given type λ as

$$\begin{aligned}
V^{\lambda}(\Gamma) &:= U^{\lambda}(\Gamma,\hat{\nu}^{\lambda}(\Gamma)) + W^{\lambda}(\Gamma,\hat{\nu}^{\lambda}(\Gamma)) \\
&= \mathbb{E}_{\hat{\nu}^{\lambda}(\Gamma)}\left[\sum_{t=1}^{\tau}Rx_t - b\hat{\nu}_t^{\lambda} \mid \lambda\right] \\
&= \mathbb{E}_{\hat{\nu}^{\lambda}(\Gamma)}\left[\sum_{t=1}^{\tau}(R\lambda - b)\mathbb{1}\left\{\hat{\nu}_t^{\lambda} = 1\right\} \mid \lambda\right] \\
&\leq \max\{T(\lambda R - b), 0\}
\end{aligned}$$

Following (EC.1.5) we have

$$\mathcal{J}^{\lambda} \leq \max_{\Gamma} V^{\lambda}(\Gamma) \leq \max\{T(\lambda R - b), 0\} = \mathrm{OPT}(\lambda). \tag{EC.1.8}$$

Then we design a contract under λ, such that the asserted value can be achieved. Specifically, we define a contract $\Gamma^{\lambda} = (\boldsymbol{\beta}, \tau)$ as in the following. The termination time $\tau = 0$ if $\lambda \leq b/R$, and $\tau = T$ otherwise. The payment at period t, $\beta_t = 0$ if $x_t = 0$ and $\beta_t = b/\lambda$ if $x_t = 1$. Then we show that $\hat{\nu}^{\lambda}(\Gamma^{\lambda}) = \bar{\nu}$. Specifically, if $\lambda \leq b/R$, then $\tau = 0$, the result is trivial. If $\lambda \geq b/R$, then $\tau = T$, thus the agent's expected utility of always exerting effort is

$$\mathbb{E}_{\bar{\nu}}\left[\sum_{t=1}^{\tau} \beta_t - b \mid \lambda\right] = \mathbb{E}_{\bar{\nu}}\left[\sum_{t=1}^{T} \frac{b}{\lambda} x_t \mid \lambda\right] - Tb = T\frac{b}{\lambda}\lambda - Tb = 0$$

For any effort strategy $\tilde{\nu} \neq \bar{\nu}$, we have

$$\mathbb{E}_{\tilde{\nu}}\left[\sum_{t=1}^{\tau} \beta_t - b\tilde{\nu}_t \mid \lambda\right] = \sum_{t=1}^{T} \mathbb{E}_{\tilde{\nu}}\left[\mathbb{E}\left[\frac{b}{\lambda} x_t - b \mid \lambda, \tilde{\nu}_t = 1\right]\right] = 0.$$

Combining the above two equations, we conclude that contract Γ^{λ} is incentive compatible. Therefore, we have

$$U^{\lambda}(\Gamma^{\lambda}, \hat{\nu}^{\lambda}(\Gamma^{\lambda})) = U^{\lambda}(\Gamma, \bar{\nu}) = \mathbb{E}_{\bar{\nu}}\left[\sum_{t=1}^{\tau} R x_t - \beta_t \mid \lambda\right] = \begin{cases} 0, & \text{if } \lambda \leq b/R \\ T(\lambda R - b), & \text{otherwise,} \end{cases}$$

which together with (EC.1.8) completes the proof. □

EC.1.3. Proof of Lemma 1

Lemma 1. *After any time t with the full-effort process and history $\mathcal{N}_t$ such that the number of good arrivals is N_t, the posterior belief that the agent is of type $\overline{\lambda}$, denoted by $P_t(N_t)$, is*

$$P_t(N_t) := \mathbb{P}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_t\right) = \mathbb{P}\left(\lambda = \overline{\lambda} \mid N_t\right) = \left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t} \left(\frac{1 - \underline{\lambda}}{1 - \overline{\lambda}}\right)^{t - N_t} \left(\frac{1}{P_0} - 1\right) + 1\right)^{-1}. \tag{14}$$

Proof: Using the conditional probability formula, we have

$$\begin{aligned} \mathbb{P}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_t\right) &= \frac{\mathbb{P}\left(\mathcal{N}_t, \lambda = \overline{\lambda}\right)}{\mathbb{P}\left(\mathcal{N}_t\right)} \\ &= \frac{\mathbb{P}\left(\lambda = \overline{\lambda}\right) \cdot \mathbb{P}\left(\mathcal{N}_t \mid \lambda = \overline{\lambda}\right)}{\mathbb{P}\left(\lambda = \overline{\lambda}\right) \cdot \mathbb{P}\left(\mathcal{N}_t \mid \lambda = \overline{\lambda}\right) + \mathbb{P}\left(\lambda = \underline{\lambda}\right) \cdot \mathbb{P}\left(\mathcal{N}_t \mid \lambda = \underline{\lambda}\right)} \\ &= \frac{P_0 \overline{\lambda}^{N_t} (1 - \overline{\lambda})^{t - N_t}}{P_0 \overline{\lambda}^{N_t} (1 - \overline{\lambda})^{t - N_t} + (1 - P_0) \underline{\lambda}^{N_t} (1 - \underline{\lambda})^{t - N_t}} \\ &= \left(1 + \frac{(1 - P_0) \underline{\lambda}^{N_t} (1 - \underline{\lambda})^{t - N_t}}{P_0 \overline{\lambda}^{N_t} (1 - \overline{\lambda})^{t - N_t}}\right)^{-1} \\ &= \left(1 + \left(\frac{1}{P_0} - 1\right) \left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t} \left(\frac{1 - \underline{\lambda}}{1 - \overline{\lambda}}\right)^{t - N_t}\right)^{-1}, \end{aligned} \tag{EC.1.9}$$

which only depends on N_t.

$$\begin{aligned}\mathbb{P}\left(\lambda=\overline{\lambda}\,|\,N_t\right) &= \frac{\mathbb{P}\left(N_t,\lambda=\overline{\lambda}\right)}{\mathbb{P}\left(N_t\right)}\\ &= \frac{\mathbb{P}\left(N_t\right)\sum_{\hat{\mathcal{N}}_t}\left[\mathbb{P}\left(\hat{\mathcal{N}}_t\,|\,N_t\right)\mathbb{P}\left(\lambda=\overline{\lambda}\,|\,\mathcal{N}_t\right)\right]}{\mathbb{P}\left(N_t\right)}\\ &= \sum_{\hat{\mathcal{N}}_t}\left[\mathbb{P}\left(\hat{\mathcal{N}}_t\,|\,N_t\right)\mathbb{P}\left(\lambda=\overline{\lambda}\,|\,\mathcal{N}_t\right)\right].\end{aligned}$$

From (EC.1.9), for any $\hat{\mathcal{N}}_t$ such that $\mathbb{P}\left(\hat{\mathcal{N}}_t\right)\neq 0$, *i.e.*, the number of good arrivals in $\hat{\mathcal{N}}_t$ is N_t, we have

$$\mathbb{P}\left(\lambda=\overline{\lambda}\,|\,\mathcal{N}_t\right)=\left(1+\left(\frac{1}{P_0}-1\right)\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-N_t}\right)^{-1}.$$

Combining the above two inequalities we have

$$\begin{aligned}\mathbb{P}\left(\lambda=\overline{\lambda}\,|\,N_t\right) &= \sum_{\hat{\mathcal{N}}_t}\left[\mathbb{P}\left(\hat{\mathcal{N}}_t\,|\,N_t\right)\right]\left(1+\left(\frac{1}{P_0}-1\right)\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-N_t}\right)^{-1}\\ &= \left(1+\left(\frac{1}{P_0}-1\right)\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-N_t}\right)^{-1}.\end{aligned}$$

This completes the proof. □

EC.1.4. Technical Lemma EC.1

To prove Proposition 2, we provide the following technical lemma.

LEMMA EC.1.

$$-\sum_{k=0}^{c-1}\left((1-\overline{\lambda})^k(k+1)\right)=\frac{(c\overline{\lambda}+1)(1-\overline{\lambda})^c-1}{\overline{\lambda}^2}.$$

Proof: Let $f(\overline{\lambda}) := \sum_{k=0}^{c-1}(1-\overline{\lambda})^{k+1}$. Then take the derivative of $f(\overline{\lambda})$ we have

$$f'(\overline{\lambda})=-\sum_{k=0}^{c-1}\left((1-\overline{\lambda})^k(k+1)\right).$$

Since

$$f(\overline{\lambda})=\frac{(1-\overline{\lambda})(1-(1-\overline{\lambda})^c)}{\overline{\lambda}}=\frac{1-\overline{\lambda}}{\overline{\lambda}}-\frac{(1-\overline{\lambda})^{c+1}}{\overline{\lambda}},$$

we have

$$f'(\overline{\lambda})=-\frac{1}{\overline{\lambda}^2}-\frac{-(c+1)(1-\overline{\lambda})^c\overline{\lambda}-(1-\overline{\lambda})^{c+1}}{\overline{\lambda}^2}=\frac{(c\overline{\lambda}+1)(1-\overline{\lambda})^c-1}{\overline{\lambda}^2}.$$

This completes the proof. □

EC.1.5. Proof of Proposition 2

Proposition 2. *Consider $\underline{\lambda}=0$ and define a contract $\check{\Gamma}=\left(\{\check{\beta}_t\}_{t\in[T]},\check{\tau}\right)$ such that $\check{\beta}_t=\beta(P_t)$ if $x_t=1$ and $\check{\beta}_t=0$ if $x_t=0$; and $\check{\tau}$ is defined according to (16) for any sequence $\{\bar{p}_s\}_{s\in[T]}$. We have*

$$Reg(\check{\Gamma};T)=\Omega(T).$$

Proof: Consider a case when $\underline{\lambda}=0$, where the bad agent can never make an arrival. The posterior belief after t periods without any good arrival is

$$P_t=\left(\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^t\left(\frac{1}{P_0}-1\right)+1\right)^{-1}=\left(\left(\frac{1}{1-\overline{\lambda}}\right)^t\left(\frac{1}{P_0}-1\right)+1\right)^{-1},$$

and if there is an arrival at period $t+1$, the agent will get payment

$$\check{\beta}_{t+1}=\frac{b}{P_t\overline{\lambda}+(1-P_t)\underline{\lambda}}=\frac{b}{P_t\overline{\lambda}}=\frac{b}{\overline{\lambda}}\left(\left(\frac{1}{1-\overline{\lambda}}\right)^t\left(\frac{1}{P_0}-1\right)+1\right),\tag{EC.1.10}$$

then the posterior belief will be $P_s=1,\forall s\geq t+1$. Besides, there is a threshold $\check{t}$ such that if there is no arrival up to period $\check{t}$, the contract will be terminated. The agent's optimal strategy, denoted by $\check{\nu}$, is to shirk in the first $s\leq\check{t}$ period, then always exerting efforts in the following $c:=\check{t}-s$ periods. The expected utility of the agent is

$$\begin{aligned}
g(c)&:=\sum_{k=0}^{c-1}P_0(1-\overline{\lambda})^k\overline{\lambda}\left(\check{\beta}_{s+k+1}-(k+1)b\right)-P_0(1-\overline{\lambda})^c cb-(1-P_0)cb\\
&=\sum_{k=0}^{c-1}P_0(1-\overline{\lambda})^k\overline{\lambda}\left(\frac{b}{\overline{\lambda}}\left(\left(\frac{1}{1-\overline{\lambda}}\right)^{s+k}\left(\frac{1}{P_0}-1\right)+1\right)-(k+1)b\right)-P_0(1-\overline{\lambda})^c cb-(1-P_0)cb\\
&=P_0cb\left(\frac{1}{1-\overline{\lambda}}\right)^s\left(\frac{1}{P_0}-1\right)+P_0b\sum_{k=0}^{c-1}(1-\overline{\lambda})^k-P_0b\overline{\lambda}\sum_{k=0}^{c-1}\left((1-\overline{\lambda})^k(k+1)\right)-P_0(1-\overline{\lambda})^c cb-(1-P_0)cb\\
&\overset{(*)}{=}(1-P_0)cb(1-\overline{\lambda})^{c-\check{t}}+P_0b\frac{1-(1-\overline{\lambda})^c}{\overline{\lambda}}+P_0b\overline{\lambda}\frac{(c\overline{\lambda}+1)(1-\overline{\lambda})^c-1}{\overline{\lambda}^2}-P_0(1-\overline{\lambda})^c cb-(1-P_0)cb\\
&=(1-P_0)cb(1-\overline{\lambda})^{c-\check{t}}+P_0b\frac{c\overline{\lambda}(1-\overline{\lambda})^c}{\overline{\lambda}}-P_0(1-\overline{\lambda})^c cb-(1-P_0)cb\\
&=(1-P_0)cb(1-\overline{\lambda})^{c-\check{t}}+P_0bc(1-\overline{\lambda})^c-P_0(1-\overline{\lambda})^c cb-(1-P_0)cb\\
&=(1-P_0)cb(1-\overline{\lambda})^{c-\check{t}}-(1-P_0)cb.
\end{aligned}$$

Note that $\overset{(*)}{=}$ follows from Lemma EC.1.

The agent wants to choose the best c, which is

$$c^*:=\underset{c}{\operatorname{argmax}}\left\{(1-P_0)cb(1-\overline{\lambda})^{c-\check{t}}-(1-P_0)cb\right\}$$

Take the derivative of $g(c)$ we have

$$\begin{aligned} g'(c) &= (1-P_0)b(1-\overline{\lambda})^{c-\check{t}} + (1-P_0)cb(1-\overline{\lambda})^{c-\check{t}}\ln(1-\overline{\lambda}) - (1-P_0)b \\ &= (1-P_0)b(1-\overline{\lambda})^{c-\check{t}}(1+c\ln(1-\overline{\lambda})) - (1-P_0)b. \end{aligned}$$

Since $\ln(1-\overline{\lambda})<0$, the optimal c should be

$$c^* < -\frac{1}{\ln(1-\overline{\lambda})} + 1.$$

Therefore, the probability of a good agent being fired at $\check{t}$ is

$$\mathbb{P}_{\check{\nu}}\left(\check{\tau}=\check{t}\,|\,\overline{\lambda}\right) = (1-\overline{\lambda})^{c^*} \geq (1-\overline{\lambda})^{-1/\ln(1-\overline{\lambda})+1} = \frac{1-\overline{\lambda}}{e},$$

where the equality uses the fact that for any $x>0$, we have $x^{-1/\ln x)} = e^{\ln x/(-\ln x)} = 1/e$. Then following (5) and (12), we can rewrite the regret as

$$\begin{aligned} \operatorname{Reg}(\check{\Gamma},T) = &\; P_0\left(\overline{\lambda}(R-\beta)T - \mathbb{E}_{\check{\nu}}\left[\sum_{t=1}^{\check{\tau}} Rx_t - \check{\beta}_t\,|\,\overline{\lambda}\right]\right) - (1-P_0)\mathbb{E}_{\check{\nu}(\Gamma)}\left[\sum_{t=1}^{\check{\tau}} Rx_t - \check{\beta}_t\,|\,\underline{\lambda}\right] && \text{(EC.1.11)} \\ = &\; P_0\left(\overline{\lambda}(R-\beta)T - \mathbb{E}_{\check{\nu}}\left[\sum_{t=1}^{\check{\tau}} Rx_t - \check{\beta}_t\,|\,\overline{\lambda}\right]\right) && \text{(EC.1.12)} \end{aligned}$$

because $\underline{\lambda}=0$. Then we have

$$\begin{aligned} &\overline{\lambda}(R-\beta)T - \mathbb{E}_{\check{\nu}}\left[\sum_{t=1}^{\check{\tau}} Rx_t - \check{\beta}_t\,|\,\overline{\lambda}\right] \\ &= \overline{\lambda}(R-\beta)(\check{t}-c^*) + \left(\overline{\lambda}(R-\beta)c^* - \mathbb{E}_{\check{\nu}}\left[\sum_{t=\check{t}-c^*+1}^{\check{t}} Rx_t - \check{\beta}_t\,|\,\overline{\lambda}\right]\right) + \overline{\lambda}(R-\beta)(T-\check{t})\left(1-\left(1-\mathbb{P}_{\check{\nu}}\left(\check{\tau}=\check{t}\,|\,\overline{\lambda}\right)\right)\right) \\ &\geq \overline{\lambda}(R-\beta)(\check{t}-c^*) + \overline{\lambda}(R-\beta)(T-\check{t})\frac{1-\overline{\lambda}}{e} = \Omega(T), \end{aligned}$$

no matter which value $\check{t}$ takes, where the inequality is because $\check{\beta}_t \geq \beta$. This completes the proof.

□

EC.1.6. Proof of Proposition 3

Proposition 3. *For any $t\in[T]$, $n\leq t-1$, and $(t-n)$-dimensional vector* $\mathbf{w} := (w_0, w_1, \ldots, w_{t-1-n})^\intercal$, *define the following dynamic programming recursion*

$$\begin{aligned} J(t,n,\mathbf{w}) := \max_{\mathbf{w}^{\pm},\beta^{\pm},I}\; & I\left\{\lambda_{t,n}\left[(R-\beta^+) + J(t+1,n+1,\mathbf{w}^+)\right]\right. \\ & \left.+(1-\lambda_{t,n})\left[-\beta^- + J(t+1,n,\mathbf{w}^-)\right]\right\} \\ \text{s.t.}\quad & w_0 = I(\lambda_{t,n}(\beta^+ + w_0^+) + (1-\lambda_{t,n})(\beta^- + w_0^-) - b) \\ & w_k = I\max\left\{\lambda_{t-k,n}(\beta^+ + w_k^+) + (1-\lambda_{t-k,n})(\beta^- + w_k^-) - b,\right. \\ & \qquad\left.\beta^- + w_{k+1}^-\right\},\quad \forall t>1,\; k=0,\ldots,t-1-n \\ & \mathbf{w}^+ \in \mathbb{R}_+^{t-n},\; \mathbf{w}^- \in \mathbb{R}_+^{t+1-n}, \beta^{\pm}\in\mathbb{R}_+,\; I\in\{0,1\}, \end{aligned} \tag{18}$$

with boundary conditions

$$J(T+1,n',\mathbf{0})=0,\ \textit{and}\ J(T+1,n',\mathbf{w})=-\infty,\ \textit{if}\ \mathbf{w}\neq\mathbf{0}, \tag{19}$$

in which $n'\leq T$, and both $\mathbf{0}$ and $\mathbf{w}$ are $T+1-n'$-dimensional vectors.

We have $\hat{J}(w)=J(1,0,w)$.

Proof:

For the clarity of exposition, we omit the constraints $\mathbf{w}^{+}\in\mathbb{R}_{+}^{t-n}$, $\mathbf{w}^{-}\in\mathbb{R}_{+}^{t+1-n}, \beta^{\pm}\in\mathbb{R}_{+}$ in this proof when the context is clear. According to the boundary condition, we have

$$\begin{aligned}
J(T,N_{T-1},\mathbf{w}) = \max_{\beta^{\pm},I}\ & I\left\{\lambda_{T,N_{T-1}}(R-\beta^{+})+(1-\lambda_{T,N_{T-1}})\left(-\beta^{-}\right)\right\}\\
\text{s.t. }\ & w_0 = I(\lambda_{T,N_{T-1}}\beta^{+}+(1-\lambda_{T,N_{T-1}})\beta^{-}-b)\\
& w_k = I\max\left\{\lambda_{T-k,N_{T-1}}\beta^{+}+(1-\lambda_{T-k,N_{T-1}})\beta^{-}-b,\beta^{-}\right\},\\
& \quad \forall k=0,\ldots,T-1-N_{T-1}\\
& I\in\{0,1\}\\
= \max_{\beta_T^{\pm},\tau}\ & \mathbb{E}\left[\sum_{t=T}^{\tau} Rx_T-\beta_T \mid N_{T-1}\right]\\
\text{s.t. }\ & w_0=\mathbb{E}\left[\sum_{s=T}^{\tau}\beta_s-b\mid N_{T-1}\right]\\
& w_k=\max_{\boldsymbol{\nu}}\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{s=T}^{\tau}\beta_s-b\nu_s\mid N_{T-1},k\right],\\
& \quad \forall k=0,\ldots,T-1-N_{T-1}\\
& \tau\in\{T-1,T\}.
\end{aligned}$$

If for any t, N_t, and $\mathbf{w}$, we use the following induction hypothesis,

$$\begin{aligned}
J(t+1,N_t,\mathbf{w}) = \max_{\beta_{t+1}^{\pm},\ldots,\beta_T^{\pm},\tau}\ & \mathbb{E}\left[\sum_{s=t+1}^{\tau} Rx_s-\beta_s\mid N_t\right]\\
\text{s.t. }\quad & w_0=\mathbb{E}\left[\sum_{s=t+1}^{\tau}\beta_s-b\mid N_t\right]\\
& w_k=\max_{\boldsymbol{\nu}}\mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{s=t+1}^{\tau}\beta_s-b\nu_s\mid N_t,k\right],\\
& \quad \forall k=0,\ldots,t-N_t\\
& \tau\in\{t,t+1,\ldots,T\}.
\end{aligned}$$

Then for t with any N_{t-1} and $\mathbf{w}$,

$$J(t,N_{t-1},\mathbf{w})=\max_{\mathbf{w}^{\pm},\beta^{\pm},I} I\left\{\lambda_{t,N_{t-1}}\left[(R-\beta^{+})+\max_{\beta_{t+1}^{'\pm},\ldots,\beta_T^{'\pm},\tau'\in\Pi'}\mathbb{E}\left[\sum_{s=t+1}^{\tau}Rx_t-\beta_t^{'}\mid N_{t-1}+1\right]\right]\right.$$

$$
\begin{aligned}
&+(1-\lambda_{t,N_{t-1}})\left[-\beta^- + \max_{\beta''^{\pm}_{t+1},\dots,\beta''^{\pm}_{T},\tau''\in\Pi''} \mathbb{E}\left[\sum_{s=t+1}^{\tau} Rx_t - \beta''_t \mid N_{t-1}-1\right]\right]\Bigg\}\\
\text{s.t.}\qquad & w_0 = I(\lambda_{t,N_{t-1}}(\beta^+ + w_0^+) + (1-\lambda_{t,N_{t-1}})(\beta^- + w_0^-) - b) \qquad \text{(EC.1.13)}\\
& w_k = I\max\left\{\lambda_{t-k,N_{t-1}}(\beta^+ + w_k^+) + (1-\lambda_{t-k,N_{t-1}})(\beta^- + w_k^-) - b,\right.\\
&\qquad\qquad \left.\beta^- + w_{k+1}^-\right\},\ \text{when } t>1,\ \forall k=0,\dots,t-1-N_{t-1} \qquad \text{(EC.1.14)}\\
& I\in\{0,1\},
\end{aligned}
$$

in which the feasible set Π' is defined by the following constraints

$$
\begin{aligned}
& w_0^+ = \mathbb{E}\left[\sum_{s=t+1}^{\tau} \beta'_s - b \mid N_{t-1}+1\right] \qquad \text{(EC.1.13a)}\\
& w_k^+ = \max_{\boldsymbol{\nu}} \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{s=t+1}^{\tau} \beta'_s - b\nu_s \mid N_{t-1}+1, k\right], \qquad \text{(EC.1.14a)}\\
& \qquad \forall k=0,\dots,t-1-N_t,\\
& \tau'\in\{t,t+1,\dots,T\},
\end{aligned}
$$

and the feasible set Π'' is defined by the following constraints

$$
\begin{aligned}
& w_0^- = \mathbb{E}\left[\sum_{s=t+1}^{\tau} \beta''_s - b \mid N_{t-1}-1\right] \qquad \text{(EC.1.13b)}\\
& w_k^- = \max_{\boldsymbol{\nu}} \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{s=t+1}^{\tau} \beta''_s - b\nu_s \mid N_{t-1}-1, k\right], \qquad \text{(EC.1.14b)}\\
& \quad \forall k=0,\dots,t-N_t\\
& \tau''\in\{t,t+1,\dots,T\}.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
J(t,N_{t-1},\mathbf{w}) = &\max_{\beta_t^{\pm},\dots,\beta_T^{\pm},\tau} \mathbb{E}\left[\sum_{s=t}^{\tau} Rx_s - \beta_s \mid N_{t-1}\right]\\
\text{s.t.}\qquad & w_0 = \mathbb{E}\left[\sum_{s=t}^{\tau} \beta_s - b \mid N_{t-1}\right] \qquad \text{(EC.1.15)}\\
& w_k = \max_{\boldsymbol{\nu}} \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{s=t}^{\tau} \beta_s - b\nu_s \mid N_t, k\right], \qquad \text{(EC.1.16)}\\
& \quad \forall k=0,\dots,t-1-N_{t-1}\\
& \tau\in\{t-1,t,\dots,T\},
\end{aligned}
$$

where (EC.1.15) is from combining (EC.1.13a), (EC.1.13b) and (EC.1.13); and (EC.1.16) from (EC.1.14a), (EC.1.14b) and (EC.1.14). Therefore, for $t=1$, we have

$$
J(1,0,w) = \max_{\boldsymbol{\beta},\tau} \mathbb{E}\left[\sum_{s=1}^{\tau} Rx_s - \beta_s\right]
$$

$$
\begin{aligned}
\text{s.t.} \qquad & w = \mathbb{E}\left[\sum_{s=1}^{\tau} \beta_s - b\right] \\
& w = \max_{\boldsymbol{\nu}} \mathbb{E}_{\boldsymbol{\nu}}\left[\sum_{s=1}^{\tau} \beta_s - b\nu_s\right], \\
& \tau \in \{0,1,\dots,T\}. \\
= \hat{J}(w), &
\end{aligned}
$$

which completes the proof. □

EC.2. Proofs of statements in Section 3

EC.2.1. Proof of Lemma 2

Lemma 2. *Given any trajectory $\mathcal{N}$ and effort process $\boldsymbol{\nu}$, we have*

$$\mathbb{P}_{\boldsymbol{\nu}}\left(\mathcal{N} \mid \overline{\lambda}\right) \geq C^{\sum_{t=1}^{T} \nu_t(\mathcal{N})} \mathbb{P}_{\boldsymbol{\nu}}\left(\mathcal{N} \mid \underline{\lambda}\right). \tag{24}$$

Proof: Let $H := \sum_{t=1}^{T} \nu_t(\mathcal{N})$ be the total number of periods in which the agent exerts effort under effort process ν following trajectory $\mathcal{N}$. Let $N := N_T$ be the total number of arrivals following trajectory $\mathcal{N}$.

If there exists $t \in [T]$ such that $x_t = 1$ and $\nu_t(\mathcal{N}) = 0$, then

$$\mathbb{P}_{\nu}\left(\mathcal{N} \mid \overline{\lambda}\right) = \mathbb{P}_{\nu}\left(\mathcal{N} \mid \underline{\lambda}\right) = 0,$$

which directly implies (24). If for any $s \in [t]$ such that $x_s = 1$ we have $\nu_s(\mathcal{N}) = 1$, then

$$\mathbb{P}_{\nu}\left(\mathcal{N} \mid \overline{\lambda}\right) = \overline{\lambda}^{N} (1 - \overline{\lambda})^{H-N},$$

and

$$\mathbb{P}_{\nu}\left(\mathcal{N} \mid \underline{\lambda}\right) = \underline{\lambda}^{N} (1 - \underline{\lambda})^{H-N},$$

Therefore,

$$\mathbb{P}_{\nu}\left(\mathcal{N} \mid \overline{\lambda}\right) = \left(\frac{\overline{\lambda}}{\underline{\lambda}}\right)^{N} \left(\frac{1-\overline{\lambda}}{1-\underline{\lambda}}\right)^{H-N} \cdot \mathbb{P}_{\nu}\left(\mathcal{N} \mid \underline{\lambda}\right) \geq \left(\frac{1-\overline{\lambda}}{1-\underline{\lambda}}\right)^{H} \mathbb{P}_{\nu}\left(\mathcal{N} \mid \underline{\lambda}\right),$$

which completes the proof. □

EC.2.2. Proof of Lemma 3

LEMMA 3. *For any contract and its best-response effort process that satisfies* (23), *we have*

$$\mathbb{E}_{\hat{\nu}}\left[\sum_{t=1}^{T}(1-\hat{\nu}_t)\mid\overline{\lambda}\right]\geq\frac{1}{2}T^{-1/e}(T-C\ln T). \tag{27}$$

Proof: Following (26) and Lemma 2, for any trajectory $\mathcal{N}\in\mathcal{A}$, we have

$$\mathbb{P}_{\hat{\nu}}\left(\mathcal{N}\mid\overline{\lambda}\right)\geq C^{C\ln T}\mathbb{P}_{\hat{\nu}}\left(\mathcal{N}\mid\underline{\lambda}\right)=T^{C\ln C}\mathbb{P}_{\hat{\nu}}\left(\mathcal{N}\mid\underline{\lambda}\right),$$

which implies, due to (25), that,

$$\sum_{\mathcal{N}\in\mathcal{A}}\mathbb{P}_{\hat{\nu}}\left(\mathcal{N}\mid\overline{\lambda}\right)\geq\frac{1}{2}T^{C\ln C}. \tag{EC.2.1}$$

Inequality (26) further implies

$$\sum_{t=1}^{T}(1-\hat{\nu}_t(\mathcal{N}))>T-C\ln T,\ \forall\mathcal{N}\in\mathcal{A}. \tag{EC.2.2}$$

Then the expected number of shirking periods for a high type agent is at least

$$\begin{aligned}
&\mathbb{E}_{\hat{\nu}}\left[\sum_{t=1}^{T}(1-\hat{\nu}_t)\mid\overline{\lambda}\right]\\
=&\sum_{\mathcal{N}}\left(\mathbb{P}_{\hat{\nu}}\left(\mathcal{N}\mid\overline{\lambda}\right)\sum_{t=1}^{T}(1-\hat{\nu}_t(\mathcal{N}))\right)\\
\geq&\sum_{\mathcal{N}\in\mathcal{A}}\left(\mathbb{P}_{\hat{\nu}}\left(\mathcal{N}\mid\overline{\lambda}\right)\sum_{t=1}^{T}(1-\hat{\nu}_t(\mathcal{N}))\right)\\
\geq&\frac{1}{2}T^{C\ln C}(T-C\ln T)\\
\geq&\frac{1}{2}T^{-1/e}(T-C\ln T),
\end{aligned} \tag{EC.2.3}$$

where the first inequality is because for any $\mathcal{N}$, $\mathbb{P}_{\hat{\nu}}\left(\mathcal{N}\mid\overline{\lambda}\right)\sum_{t=1}^{T}(1-\hat{\nu}_t(\mathcal{N}_{t-1}))\geq 0$, the second inequality is from (EC.2.2) and (EC.2.1), and the last inequality is because $x\ln x\geq -1/e$ for all $x>0$, which implies (27). □

EC.2.3. Proof of Theorem 1

THEOREM 1. *For any contract* Γ, *we have*

$$Reg(\Gamma,T)\geq\min\left\{\frac{1}{2}(1-P_0)(b-\underline{\lambda}R)C\ln T,\ \frac{1}{2}P_0(\overline{\lambda}R-b)T^{-1/e}(T-C\ln T)\right\}=\Omega(\ln T).$$

Proof: The proof follows directly from Lemma 2 and Lemma 3. □

EC.3. Proofs of statements in Section 4

EC.3.1. Proof of Proposition 4

PROPOSITION 4. *Let Γ be the contract generated according to Algorithm 1. This contract Γ satisfies (IC).*

Proof: Assume that there is an effort process $\hat{\nu} \neq \bar{\nu}$, such that $W(\hat{\Gamma}, \hat{\nu}) > W(\hat{\Gamma}, \bar{\nu})$. Then there must be a trajectory $\mathcal{N}$ such that $\mathbb{P}_{\hat{\nu}}(\mathcal{N}) > 0$, and for some $t \leq \tau(\mathcal{N})$, we have $\hat{\nu}_t(\mathcal{N}) = 0$. Denote $s(\mathcal{N}) := \max\{t \mid t \leq \tau(\mathcal{N}), \hat{\nu}_t(\mathcal{N}) = 0\}$. Furthermore, denote $\tilde{\mathcal{N}} := \operatorname{argmax}_{\mathcal{N}}\{s(\mathcal{N}) \mid \mathbb{P}_{\hat{\nu}}(\mathcal{N}) > 0\}$, and $\tilde{s} := s(\tilde{\mathcal{N}})$. According to (14) and (30), we have

$$\mathbb{P}_{\hat{\nu}}\left(\lambda = \overline{\lambda} \mid \tilde{\mathcal{N}}_{\tilde{s}-1}\right) \geq \mathbb{P}_{\bar{\nu}}\left(\lambda = \overline{\lambda} \mid \tilde{\mathcal{N}}_{\tilde{s}-1}\right) \geq P^{\tilde{s}-1}. \tag{EC.3.1}$$

Define $\tilde{\nu}$ as a new effort process same as $\hat{\nu}$ except that $\tilde{\nu}_{\tilde{s}}(\tilde{\mathcal{N}}) = 1$. In the following, we will prove that $W(\hat{\Gamma}, \tilde{\nu}) \geq W(\hat{\Gamma}, \hat{\nu})$. Therefore, if we repeat this process of defining a new effort process, we will finally get $\bar{\nu}$ and $W(\hat{\Gamma}, \bar{\nu}) \geq W(\hat{\Gamma}, \hat{\nu})$. Then by contradiction, we get the result.

Define the forward-looking utility at period t with history $\mathcal{N}_{t-1}$ under contract Γ following effort process ν as

$$W_t(\Gamma, \nu, \mathcal{N}_{t-1}) := \mathbb{E}_{\nu}\left[\sum_{s=t}^{\tau} \beta_t - b\nu_t \mid \mathcal{N}_{t-1}\right]. \tag{EC.3.2}$$

Then under contract $\hat{\Gamma}$, for any effort process ν,

$$\begin{aligned} W_t(\hat{\Gamma}, \nu, \mathcal{N}_{t-1}) = \mathbb{P}_{\nu}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_{t-1}\right) \sum_{m=t}^{\tau} (\overline{\lambda}\beta^m - b)\mathbb{E}\left[\nu_m \mid \overline{\lambda}, \mathcal{N}_{t-1}\right] \\ + \left(1 - \mathbb{P}_{\nu}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_{t-1}\right)\right) \sum_{m=t}^{\tau} (\underline{\lambda}\beta^m - b)\mathbb{E}\left[\nu_m \mid \underline{\lambda}, \mathcal{N}_{t-1}\right]. \end{aligned}$$

Since $\hat{\nu}$ and $\tilde{\nu}$ are only different at period $\tilde{s}$ given $\tilde{\mathcal{N}}_{\tilde{s}-1}$, and $\hat{\nu}_t(\mathcal{N}) = \tilde{\nu}_t(\mathcal{N}) = 1$ for any $\mathcal{N} \supseteq \tilde{\mathcal{N}}_{\tilde{s}}$ and $\tilde{s} < t \leq \tau(\mathcal{N})$, we have

$$\begin{aligned} W(\hat{\Gamma}, \tilde{\nu}) - W(\hat{\Gamma}, \hat{\nu}) &= \mathbb{P}_{\hat{\nu}}\left(\tilde{\mathcal{N}}_{\tilde{s}-1}\right)\left(W_{\tilde{s}}(\hat{\Gamma}, \tilde{\nu}, \tilde{\mathcal{N}}_{\tilde{s}-1}) - W_{\tilde{s}}(\hat{\Gamma}, \hat{\nu}, \tilde{\mathcal{N}}_{\tilde{s}-1})\right) \\ &= \mathbb{P}_{\hat{\nu}}\left(\tilde{\mathcal{N}}_{\tilde{s}-1}\right)\left[\mathbb{P}_{\hat{\nu}}\left(\lambda = \overline{\lambda} \mid \tilde{\mathcal{N}}_{\tilde{s}-1}\right)(\overline{\lambda}\beta^{\tilde{s}} - b) + \left(1 - \mathbb{P}_{\hat{\nu}}\left(\lambda = \overline{\lambda} \mid \tilde{\mathcal{N}}_{\tilde{s}-1}\right)\right)(\underline{\lambda}\beta^{\tilde{s}} - b)\right] \\ &\geq \mathbb{P}_{\hat{\nu}}\left(\tilde{\mathcal{N}}_{\tilde{s}-1}\right)\left(P^{\tilde{s}-1}\overline{\lambda}\beta^{\tilde{s}} + (1 - P^{\tilde{s}-1})\underline{\lambda}\beta^{\tilde{s}} - b\right) \\ &= 0, \end{aligned}$$

where the inequality is according to (EC.3.1) and the last equation is according to (32). This together with the statement above completes the proof. □

EC.3.2. Technical Lemma EC.2

Lemma EC.2. *Let Γ be the contract generated by Algorithm 1, then we have*

$$Reg(\Gamma;T) \le P_0\overline{\lambda}\sum_{t=1}^{T}(\beta^{(t)}-\beta) + P_0\left(T-\mathbb{E}\left[\tau \mid \overline{\lambda}\right]\right)\overline{\lambda}(R-\beta) + (1-P_0)\mathbb{E}\left[\tau \mid \underline{\lambda}\right]\underline{\lambda}(\bar{\beta}-R).$$

Proof: First, according to (5) and (12), we can rewrite the regret as

$$\mathrm{Reg}(\Gamma;T) = \ P_0\left(\overline{\lambda}(R-\beta)T-\mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau}Rx_t-\beta_t \mid \overline{\lambda}\right]\right)-(1-P_0)\mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau}Rx_t-\beta_t \mid \underline{\lambda}\right], \tag{EC.3.3}$$

where $\bar{\beta}=b/\underline{\lambda}$ is defined in Section 4. For the first term in (EC.3.3),

$$\begin{aligned}
&\overline{\lambda}(R-\beta)T-\mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau}Rx_t-\beta_t \mid \overline{\lambda}\right]\\
&=\overline{\lambda}(R-\beta)T-\sum_{t=1}^{T}\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)\overline{\lambda}(R-\beta^{(t)})\\
&=\sum_{t=1}^{T}\overline{\lambda}\left[(R-\beta)-\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)(R-\beta^{(t)})+\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)(R-\beta)-\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)(R-\beta)\right]\\
&=\sum_{t=1}^{T}\overline{\lambda}\left[\left(\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)(R-\beta)-\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)(R-\beta^{(t)})\right)+\left((R-\beta)-\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)(R-\beta)\right)\right]\\
&\le\sum_{t=1}^{T}\overline{\lambda}(\beta^{(t)}-\beta)+\sum_{t=1}^{T}\mathbb{P}\left(\tau< t \mid \overline{\lambda}\right)\overline{\lambda}(R-\beta),
\end{aligned} \tag{EC.3.4}$$

where the inequality follows from $\mathbb{P}\left(\tau\ge t \mid \overline{\lambda}\right)\le 1$ and (32). For the second term in (EC.3.3),

$$\begin{aligned}
&-\mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau}Rx_t-\beta_t \mid \underline{\lambda}\right]\\
=&\ -\sum_{t=1}^{T}(R-\beta^{(t)})\underline{\lambda}\mathbb{P}\left(\tau\ge t \mid \underline{\lambda}\right) \qquad\qquad \text{(EC.3.5)}\\
\le&\ \underline{\lambda}(\bar{\beta}-R)\sum_{t=1}^{T}\mathbb{P}\left(\tau\ge t \mid \underline{\lambda}\right),
\end{aligned}$$

where $\bar{\beta}=b/\underline{\lambda}$, such that $\beta^{(t)}\le\bar{\beta}$ and $R\le\bar{\beta}$, following (1) and (32). Combining (EC.3.4) and (EC.3.5) we get the desired result. □

EC.3.3. Proof of Theorem 2

Theorem 2. *For any instance of the online dynamic contract design problem with T periods, let Γ be a contract as described in Algorithm 1. The total expected regret of implementing contract Γ satisfies*

$$Reg(\Gamma,T)=O(\ln T). \tag{33}$$

Proof: We can upper bound the regret by a summation of three parts. First, we have from Lemma EC.2 that

$$\mathrm{Reg}(\Gamma;T) \le P_0\overline{\lambda}\sum_{t=1}^{T}(\beta^{(t)}-\beta) + P_0\left(T-\mathbb{E}\left[\tau \mid \overline{\lambda}\right]\right)\overline{\lambda}(R-\beta) + (1-P_0)\mathbb{E}\left[\tau \mid \underline{\lambda}\right]\underline{\lambda}(\bar{\beta}-R).$$

Then combining the results in Lemma 5 and Lemma 7, we can upper bound the expected total regret in the following:

$$\begin{aligned}\mathrm{Reg}(\Gamma;T) \le\ & P_0\overline{\lambda}\sum_{t=1}^{T}(\beta^{(t)}-\beta) + P_0\left(T-\mathbb{E}\left[\tau \mid \overline{\lambda}\right]\right)\overline{\lambda}(R-\beta) + (1-P_0)\mathbb{E}\left[\tau \mid \underline{\lambda}\right]\underline{\lambda}(\bar{\beta}-R)\\ \le\ & P_0\overline{\lambda}\left(c_1+\frac{1}{\overline{\lambda}^2}+\frac{3c_1}{1-T^{c_2}}\right)c_3\ln T + P_0\overline{\lambda}(R-\beta)(\ln T+1) + (1-P_0)\underline{\lambda}(\bar{\beta}-R)\frac{4\ln T+2e/T^2}{\left(\overline{\lambda}-\underline{\lambda}\right)^2},\end{aligned}$$

which completes the proof. □

EC.3.4. Proof of Lemma 4

Lemma 4. *let Γ be a corresponding contract as illustrated in Algorithm 1. For any $k \ge 0$, and $t \in [T]$ such that*

$$t \ge \max\left\{c_1(1+3k), 1/\overline{\lambda}^2\right\}\ln T,$$

we have

$$\beta^{(t)} - \beta \le c_3 T^{-c_2 k}. \tag{37}$$

Proof: According to the calculation of $\beta^{(t)}$, we have $\beta \le \beta^{(t)} \le \overline{\beta}$ as $P^t \in [0,1]$ for all $t \in [T]$. According to the definition we have

$$\begin{aligned}\beta^{(t)} - \beta &= \frac{b}{P^{(t-1)}\overline{\lambda} + (1-P^{(t-1)})\,\underline{\lambda}} - \frac{b}{\overline{\lambda}}\\ &= \frac{b\left(\overline{\lambda} - P^{(t-1)}\overline{\lambda} - (1-P^{(t-1)})\underline{\lambda}\right)}{\overline{\lambda}\left(P^{(t-1)}\overline{\lambda} + (1-P^{(t-1)})\,\underline{\lambda}\right)}\\ &= \frac{b\left(\overline{\lambda}-\underline{\lambda}\right)\left(1-P^{(t-1)}\right)}{P^{(t-1)}\overline{\lambda}^2 + (1-P^{(t-1)})\,\overline{\lambda}\underline{\lambda}}\\ &= \frac{b\left(\overline{\lambda}-\underline{\lambda}\right)}{\overline{\lambda}^2 + (1/P^{(t-1)}-1)\,\overline{\lambda}\underline{\lambda}}\left(\frac{1}{P^{(t-1)}}-1\right)\\ &\le \frac{b\left(\overline{\lambda}-\underline{\lambda}\right)}{\overline{\lambda}^2}\left(\frac{1}{P^{(t-1)}}-1\right),\end{aligned} \tag{EC.3.6}$$

where the inequality is because $\left(1/P^{(t-1)}-1\right) > 0$. Thus we have

$$\beta^{(t)} - \beta \le \min\left\{\overline{\beta}-\beta, \frac{b\left(\overline{\lambda}-\underline{\lambda}\right)}{\overline{\lambda}^2}\left(\frac{1}{P^{(t-1)}}-1\right)\right\}. \tag{EC.3.7}$$

Then to prove (37), it is sufficient to prove

$$\frac{b\left(\overline{\lambda}-\underline{\lambda}\right)}{\overline{\lambda}^2}\left(\frac{1}{P^{(t)}}-1\right)\le c_3T^{-c_2k}. \tag{EC.3.8}$$

Substituting $P^{(t)}$ with equation (31), we have that Equation (EC.3.8) holds only if

$$\frac{b\left(\overline{\lambda}-\underline{\lambda}\right)}{\overline{\lambda}^2}\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{a_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-a_t}\left(\frac{1}{P_0}-1\right)\le c_3T^{-c_2k},$$

which is equivalent to

$$\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{a_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-a_t}\le T^{-c_2k}, \tag{EC.3.9}$$

by the definition of c_3. By the definition we have $a_t=t\max\left\{0,\overline{\lambda}-\sqrt{\ln(T-t+1)/t}\right\}$. Note that we only consider when $t\ge\ln T/\overline{\lambda}^2$, such that $a_t=t\left(\overline{\lambda}-\sqrt{\ln(T-t+1)/t}\right)\ge t\left(\overline{\lambda}-\sqrt{\ln T/t}\right)$. Note that the left-hand side of (EC.3.9) is decreasing in a_t. Next we prove a result stronger than (EC.3.9) in the following:

$$\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{t\left(\overline{\lambda}-\sqrt{\ln T/t}\right)}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t\left(1-\overline{\lambda}+\sqrt{\ln T/t}\right)}\le T^{-c_2k}. \tag{EC.3.10}$$

Taking ln on both sides we have

$$t\left(\overline{\lambda}-\sqrt{\ln T/t}\right)\ln\frac{\underline{\lambda}}{\overline{\lambda}}+t\left(1-\overline{\lambda}+\sqrt{\ln T/t}\right)\ln\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\le -c_2k\ln T.$$

Rearranging the term, we have

$$\sqrt{t}^2\left(\overline{\lambda}\ln\frac{\underline{\lambda}}{\overline{\lambda}}+\left(1-\overline{\lambda}\right)\ln\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)+\sqrt{t}\cdot\sqrt{\ln T}\ln\left(\frac{\overline{\lambda}}{\underline{\lambda}}\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)+c_2k\ln T\le 0,$$

which is a quadratic inequality of $\sqrt{t}$. Recall that Equation (EC.3.8) holds if and only if the quadratic inequality of $\sqrt{t}$ holds. Denote

$$a:=\overline{\lambda}\ln\frac{\underline{\lambda}}{\overline{\lambda}}+\left(1-\overline{\lambda}\right)\ln\frac{1-\underline{\lambda}}{1-\overline{\lambda}},$$
$$b:=\sqrt{\ln T}\ln\left(\frac{\overline{\lambda}}{\underline{\lambda}}\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right),$$
$$\text{and } c_k:=c_2k\ln T.$$

We have that $a<0$ according to Lemma EC.3, and $b,c_k>0$. Note that $a=c_2/c_1$ and $b=-\sqrt{\ln T}\cdot c_2/\sqrt{c_1}$. When $k=0$, we have $c_k=0$. The inequality holds when

$$\sqrt{t}\ge-\frac{b_T}{a}=\sqrt{\ln T}\cdot\sqrt{c_1},$$

that is, $t\ge c_1\ln T$. Thus the result holds when $k=0$. When $k\ge1$, we have $c_k>0$. Then there will be a positive root and a negative root. According to the formula of the root of quadratic equation, the inequality holds when

$$\sqrt{t}\ge\frac{b+\sqrt{b^2-4ac_k}}{-2a}.$$

Using the fact that $\sqrt{\alpha+\gamma} \le \sqrt{\alpha}+\sqrt{\gamma}$ for any $\alpha, \beta > 0$, the above inequality holds if only

$$\sqrt{t} \ge \frac{b+b+\sqrt{-4ac_k}}{-2a} = \frac{b+\sqrt{-ac_k}}{-a},$$

that is

$$t \ge \left(\frac{b+\sqrt{-ac_k}}{-a}\right)^2 = \left(\frac{b}{a}\right)^2 - \frac{c_k}{a} - \frac{2b}{a}\sqrt{-\frac{c_k}{a}}.$$

Taking into a, b and c_k we have

$$\begin{aligned}\left(\frac{b}{a}\right)^2 - \frac{c_k}{a} - \frac{2b}{a}\sqrt{-\frac{c_k}{a}} &= c_1 \ln T - \frac{-c_2 k \ln T}{c_2/c_1} + 2\sqrt{c_1 \ln T}\sqrt{-\frac{-c_2 k \ln T}{c_2/c_1}} \\ &\le c_1 \ln T(1+3k),\end{aligned}$$

where the inequality is because $\sqrt{k} \le k$ when $k \ge 1$. Therefore, when the result holds $k \ge 1$. This together with the result when $k = 0$ completes the proof. □

EC.3.5. Proof of Lemma 5

LEMMA 5. *Under contract Γ, for any $T \ge 2$, we have*

$$\sum_{t=1}^{T}(\beta^{(t)} - \beta) \le \left(c_1 + \frac{1}{\overline{\lambda}^2} + \frac{3c_1}{1-T^{-c_2}}\right) c_3 \ln T \le \left(c_1 + \frac{1}{\overline{\lambda}^2} + \frac{3c_1}{1-2^{-c_2}}\right) c_3 \ln T. \tag{38}$$

Proof: Note that we have

$$\beta^{(t)} - \beta \le \min\left\{\overline{\beta} - \beta, \frac{b\left(\overline{\lambda} - \underline{\lambda}\right)}{\overline{\lambda}^2}\left(\frac{1}{P^{(t-1)}} - 1\right)\right\}. \tag{EC.3.11}$$

From Lemma 4, for any $k = 0, 1, \ldots$, if $t \ge \max\left\{c_1(1+3k), 1/\overline{\lambda}^2\right\} \ln T$, we have

$$\beta^{(t)} - \beta \le c_3 T^{c_2 k}, \tag{EC.3.12}$$

where c_1, c_2 and c_3 are positive constants defined in Lemma 4. Then we can divide the time horizon T with k, such that

$$\begin{aligned}\sum_{t=1}^{T}(\beta^{(t)} - \beta) \le& \max\left\{c_1, \frac{1}{\overline{\lambda}^2}\right\} \ln T \cdot \min\left\{\overline{\beta} - \beta, c_3\right\} \\ &+ \sum_{k=0}^{\lfloor T/(3c_1 \ln T) - 1/3\rfloor} \sum_{t=\lceil c_1(1+3k)\ln T\rceil}^{\lfloor c_1(1+3(k+1))\ln T\rfloor} \left(c_3 T^{c_2 k}\right).\end{aligned} \tag{EC.3.13}$$

We bound the first term as

$$\max\left\{c_1, \frac{1}{\overline{\lambda}^2}\right\} \ln T \cdot \min\left\{\overline{\beta} - \beta, c_3\right\} \le \left(c_1 + \frac{1}{\overline{\lambda}^2}\right) c_3 \ln T.$$

Then we bound the second term of the right-hand side, that is

$$\begin{aligned}
&\sum_{k=0}^{\lfloor T/(3c_1\ln T)-1/3\rfloor}\sum_{t=\lceil c_1(1+3k)\ln T\rceil}^{\lfloor c_1(1+3(k+1))\ln T\rfloor}\left(c_3T^{c_2k}\right)\\
\leq&\sum_{k=0}^{\lfloor T/(3c_1\ln T)-1/3\rfloor}\left(3c_1\ln T\cdot c_3T^{c_2k}\right)\\
=&3c_1c_3\ln T\sum_{k=0}^{\lfloor T/(3c_1\ln T)-1/3\rfloor}T^{c_2k}\\
=&3c_1c_3\ln T\cdot\frac{1-T^{-c_2\lfloor T/(3c_1\ln T)-1/3\rfloor}}{1-T^{c_2}}\\
\leq&3c_1c_3\ln T\cdot\frac{1}{1-T^{c_2}}
\end{aligned}$$

Then combine the above two inequalities we can derive the result. □

EC.3.6. Proof of Lemma 6

Lemma 6. *For $\epsilon_t\leq\overline{\lambda}-\underline{\lambda}$, we have*

$$\begin{aligned}
&\mathbb{P}\left(N_t<a_t\mid\overline{\lambda}\right)\leq e^{-2t\epsilon_t^2}=\frac{1}{(T-t+1)^2},\ \textit{and}\\
&\mathbb{P}\left(N_t\geq a_t\mid\underline{\lambda}\right)\leq e^{-2t(\overline{\lambda}-\underline{\lambda}-\epsilon_t)^2}.
\end{aligned}\tag{39}$$

Proof: Apply the *Hoeffding's Inequality*, we have

$$\begin{aligned}
&\mathbb{P}\left(N_t<\overline{\lambda}t-t\epsilon_t\mid\overline{\lambda}\right)=\mathbb{P}\left(\overline{\lambda}t-N_t>t\epsilon_t\mid\overline{\lambda}\right)\leq e^{-2t\epsilon_t^2}=\frac{1}{(T-t+1)^2},\ \text{and}\\
&\mathbb{P}\left(N_t\geq\overline{\lambda}t-t\epsilon_t\mid\underline{\lambda}\right)=\mathbb{P}\left(N_t-\underline{\lambda}t>t(\overline{\lambda}-\underline{\lambda}-\epsilon_t)\mid\underline{\lambda}\right)\leq e^{-2t(\overline{\lambda}-\underline{\lambda}-\epsilon_t)^2},\ \text{for }\epsilon_t\leq\overline{\lambda}-\underline{\lambda}.
\end{aligned}$$

This completes the proof. □

EC.3.7. Proof of Lemma 7

Lemma 7. *For the termination time τ of our contract, we have*

$$\mathbb{E}\left[\tau\mid\overline{\lambda}\right]\geq T-\ln T-1,\ \textit{and}\ \mathbb{E}\left[\tau\mid\underline{\lambda}\right]\leq\frac{4\ln T+2e/T^2}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}.\tag{40}$$

Proof: First, we prove the first inequality. We have $\{\tau<t\}\subseteq\cup_{s<t}\{\tau=s\}$, which implies that $\mathbb{P}(\tau<t)\leq\sum_{s<t}\mathbb{P}(\tau=s)$. Therefore,

$$\sum_{t=1}^{T}\mathbb{P}(\tau<t)\leq\sum_{t=1}^{T}\sum_{s=1}^{t-1}\mathbb{P}(\tau=s)=\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\mathbb{P}(\tau=s)=\sum_{t=1}^{T}\mathbb{P}(\tau=t)(T-t).\tag{EC.3.14}$$

The Hoeffding's inequality implies

$$\mathbb{P}\left(\overline{\lambda}t-N_t>t\epsilon_t\mid\overline{\lambda}\right)\leq e^{-2t\epsilon_t^2}=\frac{1}{(T-t+1)^2}.\tag{EC.3.15}$$

Together with (30), we have

$$\mathbb{P}\left(\tau=t\,|\,\overline{\lambda}\right)=\mathbb{P}\left(\bigcap_{s=1}^{t-1}\{N_s\geq s(\overline{\lambda}-\epsilon_s)\}\bigcap\{N_t<t(\overline{\lambda}-\epsilon_t)\}\,|\,\overline{\lambda}\right)\leq\mathbb{P}\left(\{N_t<t(\overline{\lambda}-\epsilon_t)\}\,|\,\overline{\lambda}\right)\leq\frac{1}{(T-t+1)^2}.$$

Therefore, we have

$$\sum_{t=1}^{T}\mathbb{P}\left(\tau<t\,|\,\overline{\lambda}\right)\leq\sum_{t=1}^{T}\frac{T-t}{(T-t+1)^2}<\sum_{t=1}^{T}\frac{1}{T-t+1}<\ln T+1, \tag{EC.3.16}$$

in which the first inequality follows from (EC.3.14), the second inequality from (EC.3.15), and the last on from the fact that $\sum_{n=1}^{N}(1/n)<\ln N+1$ for any integer N. Then we have

$$\mathbb{E}\left[\tau\,|\,\overline{\lambda}\right]=\sum_{t=1}^{T}\mathbb{P}\left(\tau\geq t\right)=\sum_{t=1}^{T}(1-\mathbb{P}\left(\tau<t\right))=T-\sum_{t=1}^{T}\mathbb{P}\left(\tau<t\,|\,\overline{\lambda}\right)\geq T-\ln T-1.$$

Next we prove the second inequality. When $t\geq\ln T/(\overline{\lambda}-\underline{\lambda})^2$, we have $\epsilon_t\leq\overline{\lambda}-\underline{\lambda}$. Thus according to the terminating rule and Lemma 6, we have

$$\begin{aligned}\mathbb{E}\left[\tau\,|\,\underline{\lambda}\right]=&\sum_{t=1}^{T}\mathbb{P}\left(\tau\geq t\,|\,\underline{\lambda}\right)\\ =&\sum_{t=1}^{T}\mathbb{P}\left(N_t\geq\overline{\lambda}t-t\epsilon_t\,|\,\underline{\lambda}\right)\\ \leq&\frac{\ln T}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}+\sum_{t=\left\lceil\ln T/\left(\overline{\lambda}-\underline{\lambda}\right)^2\right\rceil}^{T}e^{-2t\left(\overline{\lambda}-\underline{\lambda}-\epsilon_t\right)^2}.\end{aligned}$$

Furthermore, when $t\geq 4\ln T/\left(\overline{\lambda}-\underline{\lambda}\right)^2$, we have that

$$\overline{\lambda}-\underline{\lambda}-\epsilon_t=\overline{\lambda}-\underline{\lambda}-\sqrt{\frac{\ln(T-t)}{t}}\geq\frac{\overline{\lambda}-\underline{\lambda}}{2}.$$

Thus we have

$$\begin{aligned}&\sum_{t=\left\lceil\ln T/\left(\overline{\lambda}-\underline{\lambda}\right)^2\right\rceil}^{T}e^{-2t\left(\overline{\lambda}-\underline{\lambda}-\epsilon_t\right)^2}\\ \leq&\frac{3\ln T}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}+\sum_{t=\left\lceil 4\ln T/\left(\overline{\lambda}-\underline{\lambda}\right)^2\right\rceil}^{T}e^{-t\left(\overline{\lambda}-\underline{\lambda}\right)^2/2}\\ \leq&\frac{3\ln T}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}+e^{-2\ln T}\frac{1-e^{-T\left(\overline{\lambda}-\underline{\lambda}\right)^2/2}}{1-e^{-\left(\overline{\lambda}-\underline{\lambda}\right)^2/2}}\\ \leq&\frac{3\ln T}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}+\frac{1}{T^2}\cdot\frac{1}{1-e^{-\left(\overline{\lambda}-\underline{\lambda}\right)^2/2}}\\ \leq&\frac{3\ln T}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}+\frac{1}{T^2}\frac{2e}{\left(\overline{\lambda}-\underline{\lambda}\right)^2}\end{aligned}$$

using the fact that e^{-x} is decreasing in x, and $1/(1-e^{-x})\leq e/x$ for any $x\in(0,1)$. Combining the above two inequalities we can complete the proof. □

EC.3.8. Proof of Proposition 5

Proposition 5. *The proposed "explore-then-commit" contract satisfies* (IC).

Proof: Denote $\hat{\Gamma}$ to represent the "explore-then-commit" contract. Following the proof of Proposition 4, we can define the corresponding effort process $\hat{\nu}$, trajectory $\tilde{N}$, and time period $\tilde{s}$. If the agent has not been terminated by $\tilde{s} > \bar{t}$, we must have

$$\mathbb{P}_{\hat{\nu}}\left(\lambda = \overline{\lambda} \,|\, \tilde{\mathcal{N}}_{\tilde{s}-1}\right) \geq \mathbb{P}_{\bar{\nu}}\left(\lambda = \overline{\lambda} \,|\, \tilde{\mathcal{N}}_{\tilde{s}-1}\right) \geq P_a. \tag{EC.3.17}$$

The remaining proof follows the same steps as the proof of Proposition 4 after (EC.3.1), except that we replace $P^{(t)}$ and $\beta^{(t)}$ in that proof with 0 and $\bar{\beta}$, respectively, for $t \leq \bar{t}$; and with P_a and β_a, respectively, for $t > \bar{t}$. □

EC.3.9. Proof of Theorem 3

Theorem 3. *For any instance of the online dynamic contract design problem with T periods, let Γ be an "explore-then-commit" contract as defined in Algorithm 2. The total expected regret can be upper bounded as*

$$Reg(\Gamma, T) = O(\ln T). \tag{42}$$

Proof: First, according to (5) and (12), we can rewrite the regret as

$$\mathrm{Reg}(\Gamma;T) = \; P_0\left(\overline{\lambda}(R-\beta)T - \mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau} Rx_t - \beta_t \,|\, \overline{\lambda}\right]\right) - (1-P_0)\mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau} Rx_t - \beta_t \,|\, \underline{\lambda}\right], \tag{EC.3.18}$$

where $\bar{\beta} = b/\underline{\lambda}$ is defined in Section 4. For the first term

$$\begin{aligned}
&\overline{\lambda}(R-\beta)T - \mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau} Rx_t - \beta_t \,|\, \overline{\lambda}\right] \\
&= \overline{\lambda}(R-\beta)T - \sum_{t=1}^{\bar{t}} \overline{\lambda}(R-\bar{\beta}) - \sum_{t=\bar{t}+1}^{T} \mathbb{P}\left(\tau \geq t \,|\, \overline{\lambda}\right) \overline{\lambda}(R-\beta_a) \\
&= \overline{\lambda}(R-\beta)T - \sum_{t=1}^{\bar{t}} \overline{\lambda}(R-\bar{\beta}) - \sum_{t=\bar{t}+1}^{T} \left(1 - \mathbb{P}\left(\tau < t \,|\, \overline{\lambda}\right)\right) \overline{\lambda}(R-\beta_a) \\
&= \overline{\lambda}(R-\beta)T - \sum_{t=1}^{\bar{t}} \overline{\lambda}(R-\bar{\beta}) - \sum_{t=\bar{t}+1}^{T} \overline{\lambda}(R-\beta_a) + \sum_{t=\bar{t}+1}^{T} \mathbb{P}\left(\tau < t \,|\, \overline{\lambda}\right) \overline{\lambda}(R-\beta_a) \\
&\leq \overline{\lambda}\left[(\bar{\beta}-\beta)\bar{t} + (\beta_a - \beta)(T-\bar{t})\right] + T\overline{\lambda}(R-\beta)\mathbb{P}\left(\tau < T \,|\, \overline{\lambda}\right).
\end{aligned} \tag{EC.3.19}$$

For the second term,

$$-\mathbb{E}_{\hat{\nu}(\Gamma)}\left[\sum_{t=1}^{\tau} Rx_t - \beta_t \,|\, \underline{\lambda}\right] = \sum_{t=1}^{\bar{t}} \underline{\lambda}\left(\bar{\beta} - R\right) - \sum_{t=\bar{t}+1}^{T} \mathbb{P}\left(\tau \geq t\right) \underline{\lambda}(R-\beta_a) \leq \underline{\lambda}\left(\bar{\beta} - R\right)\bar{t}. \tag{EC.3.20}$$

Take (EC.3.19) and (EC.3.20) into (EC.3.18) we have

$$\mathrm{Reg}(\Gamma;T) \le P_0\overline{\lambda}\left[(\bar{\beta}-\beta)\bar{t} + (\beta_a-\beta)(T-\bar{t})\right] + P_0 T\overline{\lambda}(R-\beta)\mathbb{P}\left(\tau < T \mid \overline{\lambda}\right) + (1-P_0)\underline{\lambda}(\bar{\beta}-R)\bar{t},$$

which is (43) in the main text. Then take the results in Lemma 8 and Lemma 9 into the above inequality, we have

$$\mathrm{Reg}(\Gamma;T) \le P_0\overline{\lambda}\left[(\bar{\beta}-\beta)\bar{t} + c_3\right] + P_0\overline{\lambda}(R-\beta) + (1-P_0)\underline{\lambda}(\bar{\beta}-R)\bar{t} = O(\ln T),$$

since $\bar{t} = O(\ln T)$. This completes the proof. □

EC.3.10. Proof of Lemma 8

Lemma 8. *We have*

$$\beta_a - \beta \le \frac{c_3}{T}. \tag{44}$$

Proof: By definition we have $\bar{t} \ge \max\left\{c_1(1+3/c_2),\ 1/\overline{\lambda}^2\right\}\ln T$. Then following Lemma 4 and taking $k = 1/c_2$ we have

$$\beta^{(\bar{t})} - \beta \le c_3 T^{-c_2 k} = \frac{c_3}{T}.$$

Note that $\beta_a = \beta^{(\bar{t})}$ by definition, thus we complete the proof.

□

EC.3.11. Proof of Lemma 9

Lemma 9. *The probability of terminating a capable agent before T satisfies*

$$\mathbb{P}\left(\tau < T \mid \overline{\lambda}\right) \le \frac{1}{T}. \tag{45}$$

Proof: When $t < \bar{t}$, the agent is not terminated.

When $t \ge \bar{t}$,

$$\begin{aligned}
\mathbb{P}(P_t \le P_a) &= \mathbb{P}\left(\left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-N_t}\left(\frac{1}{P_0}-1\right)+1\right)^{-1} \le \left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{a}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{\bar{t}-a}\left(\frac{1}{P_0}-1\right)+1\right)^{-1}\right) \\
&= \mathbb{P}\left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-N_t} \ge \left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{a}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{\bar{t}-a}\right) \\
&= \mathbb{P}\left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{t-N_t} \ge \left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}\bar{t}-\sqrt{\bar{t}\ln T}}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{(1-\overline{\lambda})\bar{t}+\sqrt{\bar{t}\ln T}}\right) \qquad \text{(EC.3.21)}
\end{aligned}$$

Define

$$f(t) := \left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}t-\sqrt{t\ln T}}\left(\frac{1-\underline{\lambda}}{1-\overline{\lambda}}\right)^{(1-\overline{\lambda})t+\sqrt{t\ln T}}.$$

Take ln on both sides we have

$$\begin{aligned}\ln f(t) = &\left(\bar{\lambda}t - \sqrt{t\ln T}\right)\ln\left(\frac{\underline{\lambda}}{\bar{\lambda}}\right) + \left((1-\bar{\lambda})t + \sqrt{t\ln T}\right)\ln\left(\frac{1-\underline{\lambda}}{1-\bar{\lambda}}\right)\\ = &\sqrt{t}^2 \cdot \ln\left[\left(\frac{\underline{\lambda}}{\bar{\lambda}}\right)^{\bar{\lambda}}\left(\frac{1-\underline{\lambda}}{1-\bar{\lambda}}\right)^{(1-\bar{\lambda})}\right] + \sqrt{t}\cdot\ln\left(\frac{1-\underline{\lambda}}{1-\bar{\lambda}}\frac{\bar{\lambda}}{\underline{\lambda}}\right)\sqrt{\ln T}.\end{aligned}$$

Then we can write $\ln f(t)$ as a function of $\sqrt{t}$. Specifically, define function g as

$$g(u) = \alpha' u^2 + \alpha\sqrt{\ln T}\cdot u,$$

in which α and α' are defined in (36), and $\alpha' < 0$ according to Lemma EC.3. We have $\ln f(t) = g(\sqrt{t})$.

According to the quadratic formulation of $g(u)$, when $u \geq -\alpha\sqrt{\ln T}/(2\alpha') = \sqrt{(c_1/4)\ln T}$, $g(u)$ is decreasing in u, which implies that when $t \geq (c_1/4)\ln T$, $f(t)$ is decreasing in t. Since $\bar{t} > (c_1/4)\ln T$, for any $t \geq \bar{t}$, we have $f(t) \leq f(\bar{t})$, which, together with (EC.3.21) implies that

$$\begin{aligned}\mathbb{P}\left(P_t \leq P_a\right) \leq &\,\mathbb{P}\left(\left(\frac{\underline{\lambda}}{\bar{\lambda}}\right)^{N_t}\left(\frac{1-\underline{\lambda}}{1-\bar{\lambda}}\right)^{t-N_t} \geq \left(\frac{\underline{\lambda}}{\bar{\lambda}}\right)^{\bar{\lambda}t-\sqrt{t\ln T}}\left(\frac{1-\underline{\lambda}}{1-\bar{\lambda}}\right)^{(1-\bar{\lambda})t+\sqrt{t\ln T}}\right)\\ = &\,\mathbb{P}\left(N_t \leq \bar{\lambda}t - \sqrt{t\ln T}\right) \leq \frac{1}{T^2},\end{aligned}$$

in which the last inequality follows the Hoeffding's Inequality.

Therefore, we have

$$\mathbb{P}\left(\tau < T \mid \bar{\lambda}\right) = \sum_{t=\bar{t}}^{T}\mathbb{P}\left(\tau = t \mid \bar{\lambda}\right) \leq \sum_{t=\bar{t}}^{T}\mathbb{P}\left(P_t < P_a \mid \bar{\lambda}\right) \leq \sum_{t=\bar{t}}^{T}\frac{1}{T^2} \leq \frac{1}{T}.$$

This completes the proof. □

EC.3.12. Technical Lemma EC.3

Lemma EC.3. *For any $0 < \underline{\lambda} < \bar{\lambda} < 1$, we have*

$$\bar{\lambda}\ln\frac{\underline{\lambda}}{\bar{\lambda}} + \left(1-\bar{\lambda}\right)\ln\frac{1-\underline{\lambda}}{1-\bar{\lambda}} < 0. \tag{EC.3.22}$$

Proof: Denote $\Delta_\lambda := \bar{\lambda} - \underline{\lambda}$, then $\underline{\lambda} = \bar{\lambda} - \Delta_\lambda$. Take $\underline{\lambda} = \bar{\lambda} - \Delta_\lambda$ into the left-hand side of Equation (EC.3.22) and define a new function of Δ_λ, that is

$$f(\Delta_\lambda) := \bar{\lambda}\ln\frac{\bar{\lambda}-\Delta_\lambda}{\bar{\lambda}} + \left(1-\bar{\lambda}\right)\ln\frac{1-\left(\bar{\lambda}-\Delta_\lambda\right)}{1-\bar{\lambda}}$$

Then we have

$$\begin{aligned}\frac{\partial f}{\partial\Delta_\lambda} &= \frac{\bar{\lambda}^2}{\bar{\lambda}-\Delta_\lambda}\left(-\frac{1}{\bar{\lambda}}\right) + \frac{\left(1-\bar{\lambda}\right)^2}{1-\left(\bar{\lambda}-\Delta_\lambda\right)}\frac{1}{1-\bar{\lambda}}\\ &= -\frac{\bar{\lambda}}{\bar{\lambda}-\Delta_\lambda} + \frac{1-\bar{\lambda}}{1-\left(\bar{\lambda}-\Delta_\lambda\right)}\\ &< -1+1 = 0,\end{aligned}$$

where the inequality is because $\Delta_\lambda \in (0, \overline{\lambda})$. Thus we have

$$f(\Delta_\lambda) < f(0) = 0.$$

This completes the proof. $\square$

EC.4. Proofs of statements in Section 5

EC.4.1. Proof of Lemma 10

Lemma 10. *After any time t with the full-effort process and history $\mathcal{N}_t$ such that the number of good arrivals is N_t, the posterior belief that the agent is of type $\overline{\lambda}$, denoted by $P_t(N_t)$, is*

$$P_t(N_t) := \mathbb{P}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_t\right) = \mathbb{P}\left(\lambda = \overline{\lambda} \mid N_t\right) = \left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{N_t} \cdot e^{(\overline{\lambda}-\underline{\lambda})t} \cdot \left(\frac{1}{P_0} - 1\right) + 1\right)^{-1}. \tag{48}$$

Proof: According to (53), we have

$$\mathbb{P}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_t\right) = \frac{\mathbb{P}\left(\lambda = \overline{\lambda}\right) f(\mathcal{N} \mid \overline{\lambda})}{\mathbb{P}\left(\lambda = \overline{\lambda}\right) f(\mathcal{N} \mid \overline{\lambda}) + \mathbb{P}\left(\lambda = \underline{\lambda}\right) f(\mathcal{N} \mid \underline{\lambda})} = \frac{P_0 \overline{\lambda}^{N_t} e^{-\overline{\lambda}t}}{P_0 \overline{\lambda}^{N_t} e^{-\overline{\lambda}t} + (1-P_0)\underline{\lambda}^{N_t} e^{-\underline{\lambda}t}}$$

Besides, according to the probability mass function of Poisson distribution, we have

$$\mathbb{P}\left(\lambda = \overline{\lambda} \mid N_t\right) = \frac{\mathbb{P}\left(\lambda = \overline{\lambda}\right) \mathbb{P}\left(N_t \mid \overline{\lambda}\right)}{\mathbb{P}\left(N_t \mid \mathbb{P}\left(\lambda = \overline{\lambda}\right) \overline{\lambda}\right) + \mathbb{P}\left(\lambda = \underline{\lambda}\right) \mathbb{P}\left(N_t \mid \underline{\lambda}\right)} = \frac{P_0 (\overline{\lambda}t)^{N_t} e^{-\overline{\lambda}t}}{P_0 (\overline{\lambda}t)^{N_t} e^{-\overline{\lambda}t} + (1-P_0)(\underline{\lambda}t)^{N_t} e^{-\underline{\lambda}t}}.$$

By rearranging the term we verify that $\mathbb{P}\left(\lambda = \overline{\lambda} \mid \mathcal{N}_t\right) = \mathbb{P}\left(\lambda = \overline{\lambda} \mid N_t\right)$, and both equals to the right-hand side of equation (48), which completes the proof. $\square$

EC.4.2. A Derivation of (53)

The inter-arrival times of the Poisson process with rate λ are i.i.d. exponential random variables with parameter λ. For a Poisson process with rate λ over time interval $[0, T)$, the joint distribution of observing n arrivals and arrival time epochs $\{t_i\}_{i=1}^n$ such that $0 \le t_1 < t_2 < \ldots, < t_n < T$ can be expressed as

$$f(n, t_1, t_2, \ldots, t_n; \lambda) = \Pi_{i=1}^n \left(\lambda e^{-\lambda(t_i - t_{i-1})}\right) e^{-\lambda(T - t_n)} = \lambda^n e^{-\lambda \sum_{i=1}^{n+1}(t_i - t_{i-1})} = \lambda^n e^{-\lambda T}.$$

where for ease of exposition we denote $t_0 = 0$, and $t_{n+1} = T$. The following verification serves as a sanity check,

$$\begin{aligned}
&\sum_{n=0}^{\infty} \int \cdots \int_{0 \le t_1 \le \ldots \le t_n < T} f(n, t_1, \ldots, t_n; \lambda) \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
=& \sum_{n=0}^{\infty} \lambda^n e^{-\lambda T} \int \cdots \int_{0 \le t_1 \le \ldots \le t_n < T} \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
=& \sum_{n=0}^{\infty} \lambda^n e^{-\lambda T} \cdot \frac{t^n}{n!} \\
=& \, 1,
\end{aligned}$$

where the second equality is because $t_1, \dots, t_n$ given n can be seen as order statistics for a sample from n i.i.d. uniform distributions, and the third equality is according to the probability mass function of the Poisson distribution.

EC.4.3. Proof of Lemma 11

Lemma 11. *Given any trajectory $\mathcal{N}$ and effort process $\boldsymbol{\nu}$, we have*

$$f_{\boldsymbol{\nu}}(\mathcal{N} \mid \overline{\lambda}) \geq e^{-(\overline{\lambda}-\underline{\lambda})\mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N})} f_{\boldsymbol{\nu}}(\mathcal{N} \mid \underline{\lambda}). \tag{57}$$

Proof: Let N be the number of arrivals in trajectory $\mathcal{N}$. From (54) and (55) we have

$$f_{\boldsymbol{\nu}}(\mathcal{N} \mid \overline{\lambda}) = \left(\frac{\overline{\lambda}}{\underline{\lambda}}\right)^N e^{-(\overline{\lambda}-\underline{\lambda})\mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N})} f_{\boldsymbol{\nu}}(\mathcal{N} \mid \underline{\lambda}),$$

which implies the desired result since $\overline{\lambda} > \underline{\lambda}$. □

EC.4.4. Proof of Lemma 12

Lemma 12. *For any contract and its best-response effort process that satisfies (52), we have*

$$\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=0}^{T} (1-\hat{\nu}_t)\mathrm{d}t \mid \overline{\lambda}\right] \geq \frac{1}{2} T^{-1/2}(T - C\ln T). \tag{59}$$

Proof:

$$\begin{aligned}
&\mathbb{E}_{\hat{\boldsymbol{\nu}}}\left[\int_{t=1}^{T} (1-\hat{\nu}_t)\mathrm{d}t \mid \overline{\lambda}\right] \\
\geq\ & \int_{\mathcal{N}\in\mathcal{A}} \left[\int_{t=1}^{T} (1-\hat{\nu}_t(\mathcal{N}))\mathrm{d}t\right] f_{\hat{\boldsymbol{\nu}}}(\mathcal{N} \mid \overline{\lambda})\mathrm{d}\mathcal{N} \\
\geq\ & (T - C\ln T) \int_{\mathcal{N}\in\mathcal{A}} f_{\hat{\boldsymbol{\nu}}}(\mathcal{N} \mid \overline{\lambda})\mathrm{d}\mathcal{N} \\
\geq\ & (T - C\ln T) \int_{\mathcal{N}\in\mathcal{A}} e^{-(\overline{\lambda}-\underline{\lambda})C\ln T} f_{\hat{\boldsymbol{\nu}}}(\mathcal{N} \mid \underline{\lambda})\mathrm{d}\mathcal{N} \\
\geq\ & \frac{1}{2}(T - C\ln T)T^{-\frac{1}{2}},
\end{aligned}$$

where the first inequality is because $1 - \hat{\nu}_t(\mathcal{N}) \geq 0$, the second inequality uses (58), the third inequality uses (59), and the last inequality follows from (58). This completes the proof. □

EC.4.5. Proof of Proposition 6

We introduce some mathematical notations for this section. For any $\mathcal{N}$-adapted process $\{X_t\}$, define

$$X_{t-} := \lim_{s\uparrow t} X_s.$$

Correspondingly, we extend the definition (56) to

$$\int g(\mathcal{N}_{t-}) f_{\boldsymbol{\nu}}(\mathcal{N}_{t-} \mid \lambda) \mathrm{d}\mathcal{N}_{t-} := \sum_{n=0}^{\infty} \underset{t_1,\ldots,t_n \in [0..t)}{\int \cdots \int} g(n, t_1, \ldots, t_n) f_{\boldsymbol{\nu}}(n, t_1, \ldots, t_n; \lambda, t) \mathrm{d}t_1 \ldots \mathrm{d}t_n. \tag{EC.4.1}$$

We first show the following technical result.

Lemma EC.4. *For any function $g : \{\mathcal{N}_{t-}\} \to \mathbb{R}$, we have*

$$\int g(\mathcal{N}_{t-}) f_{\boldsymbol{\nu}}(\mathcal{N}_{t-}) \mathrm{d}\mathcal{N}_{t-} = \int g\big(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})\big) f_{\tilde{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-}) \mathrm{d}\hat{\mathcal{N}}_{t-}.$$

Proof: Let $\mathcal{N}_{t-}$ be represented by n, the number of arrivals, with arrival time epochs $\{t_i\}_{i=1}^n$. Following (53), (54), (55) and (EC.4.1), we have

$$\begin{aligned}
&\int g(\mathcal{N}_{t-}) f_{\boldsymbol{\nu}}(\mathcal{N}_{t-} \mid \lambda) \mathrm{d}\mathcal{N}_{t-} \\
&= \sum_{n=0}^{\infty} \underset{t_1,\ldots,t_n \in [0..t)}{\int \cdots \int} g(\mathcal{N}_{t-}) f_{\boldsymbol{\nu}}(n, t_1, \ldots, t_n; \lambda, t) \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
&= \sum_{n=0}^{\infty} \underset{\substack{t_1,\ldots,t_n \in [0..t): \\ \nu_{t_i}(\mathcal{N}_{t-}) = 1, \forall i \in [n]}}{\int \cdots \int} g(\mathcal{N}_{t-}) \lambda^n e^{-\lambda \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-})} \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
&= \sum_{n=0}^{\infty} \underset{\substack{t_1,\ldots,t_n \in [0..t): \\ \nu_{t_i}(\mathcal{N}_{t-}) = 1, \forall i \in [n]}}{\int \cdots \int} g(\mathcal{N}_{t-}) \lambda^n e^{-\lambda \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-})} \\
&\quad \cdot \left[\sum_{\tilde{n}=0}^{\infty} \underset{\tilde{t}_1,\ldots,\tilde{t}_{\tilde{n}} \in [0..t - \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-}))}{\int \cdots \int} f(\tilde{n}, \tilde{t}_1, \ldots, \tilde{t}_{\tilde{n}}); \lambda, t - \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-})) \mathrm{d}\tilde{t}_1 \ldots \mathrm{d}\tilde{t}_{\tilde{n}} \right] \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
&= \sum_{n=0}^{\infty} \underset{\substack{t_1,\ldots,t_n \in [0..t): \\ \nu_{t_i}(\mathcal{N}_{t-}) = 1, \forall i \in [n]}}{\int \cdots \int} g(\mathcal{N}_{t-}) \lambda^n e^{-\lambda \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-})} \left[\sum_{\tilde{n}=0}^{\infty} \underset{\tilde{t}_1,\ldots,\tilde{t}_{\tilde{n}} \in [0..t - \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-}))}{\int \cdots \int} \lambda^{\tilde{n}} e^{-\lambda (t - \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-}))} \mathrm{d}\tilde{t}_1 \ldots \mathrm{d}\tilde{t}_{\tilde{n}} \right] \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
&= \sum_{n=0}^{\infty} \underset{\substack{t_1,\ldots,t_n \in [0..t): \\ \nu_{t_i}(\mathcal{N}_{t-}) = 1, \forall i \in [n]}}{\int \cdots \int} g(\mathcal{N}_{t-}) \lambda^n e^{-\lambda \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-})} \left[\sum_{\tilde{n}=0}^{\infty} \lambda^{\tilde{n}} e^{-\lambda (t - \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-}))} \underset{\tilde{t}_1,\ldots,\tilde{t}_{\tilde{n}} \in [0..t - \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-}))}{\int \cdots \int} 1 \cdot \mathrm{d}\tilde{t}_1 \ldots \mathrm{d}\tilde{t}_{\tilde{n}} \right] \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
&\overset{(*)}{=} \sum_{n=0}^{\infty} \underset{\substack{t_1,\ldots,t_n \in [0..t): \\ \nu_{t_i}(\mathcal{N}_{t-}) = 1, \forall i \in [n]}}{\int \cdots \int} g(\mathcal{N}_{t-}) \lambda^n e^{-\lambda \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-})} \left[\sum_{\tilde{n}=0}^{\infty} \lambda^{\tilde{n}} e^{-\lambda (t - \mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-}))} \underset{\substack{\tilde{t}_1,\ldots,\tilde{t}_{\tilde{n}} \in [0..t): \\ \nu_{\tilde{t}_i}(\mathcal{N}_{t-}) = 0, \forall i \in [\tilde{n}]}}{\int \cdots \int} 1 \cdot \mathrm{d}\tilde{t}_1 \ldots \mathrm{d}\tilde{t}_{\tilde{n}} \right] \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
&= \sum_{n=0}^{\infty} \sum_{\tilde{n}=0}^{\infty} \underset{\substack{t_1,\ldots,t_n \in [0..t): \\ \nu_{t_i}(\mathcal{N}_{t-}) = 1, \forall i \in [n]}}{\int \cdots \int} \underset{\substack{\tilde{t}_1,\ldots,\tilde{t}_{\tilde{n}} \in [0..t): \\ \nu_{\tilde{t}_i}(\mathcal{N}_{t-}) = 0, \forall i \in [\tilde{n}]}}{\int \cdots \int} g(\mathcal{N}_{t-}) \lambda^{n + \tilde{n}} e^{-\lambda t} \mathrm{d}\tilde{t}_1 \ldots \mathrm{d}\tilde{t}_{\tilde{n}} \, \mathrm{d}t_1 \ldots \mathrm{d}t_n,
\end{aligned} \tag{EC.4.2}$$

where $(*)$ follows because $\int_{s\in[0,t)} \mathbb{1}\{\nu_s(\mathcal{N}_{t-})=0\}\,\mathrm{d}s = t-\mathcal{T}_{\boldsymbol{\nu}}(\mathcal{N}_{t-})$. Let $\hat{\mathcal{N}}_{t-}$ be the unique trajectory represented by the number of arrivals $\hat{n}=n+\tilde{n}$ with arrival time epochs $\{\hat{t}_i\}_{i\in[\hat{n}]}=\{t_i\}_{i\in n}\cup\{\tilde{t}_i\}_{i\in[\tilde{n}]}$. Conversely, for any trajectory $\hat{\mathcal{N}}_{t-}$ with $\hat{n}$ arrivals and arrival epochs $\{\hat{t}_i\}_{i\in[\hat{n}]}$, there exists a unique trajectory $\mathcal{N}_{t-}=\mathcal{F}(\hat{\mathcal{N}}_{t-})$, implying a unique partition of $\{\hat{t}_i\}_{i\in[\hat{n}]}$ into $\{\hat{t}_i\}_{i\in[\hat{n}]}=\{t_i\}_{i\in[n]}\cup\{\tilde{t}_i\}_{i\in[\tilde{n}]}$, where $\hat{n}=n+\tilde{n}$, $\nu_{t_i}(\mathcal{N}_{t-})=1,\forall i\in[n]$ and $\nu_{\tilde{t}_i}(\mathcal{N}_{t-})=0,\forall i\in[\tilde{n}]$. Therefore, we have

$$
\begin{aligned}
&\sum_{n=0}^{\infty}\sum_{\tilde{n}=0}^{\infty}\underset{\substack{t_1,\ldots,t_n\in[0..t):\\ \nu_{t_i}(\mathcal{N}_{t-})=1,\forall i\in[n]}}{\int\cdots\int}\quad\underset{\substack{\tilde{t}_1,\ldots,\tilde{t}_{\tilde{n}}\in[0..t):\\ \nu_{\tilde{t}_i}(\mathcal{N}_{t-})=0,\forall i\in[\tilde{n}]}}{\int\cdots\int}\ g(\mathcal{N}_{t-})\lambda^{n+\tilde{n}}e^{-\lambda t}\mathrm{d}\tilde{t}_1\ldots\mathrm{d}\tilde{t}_{\tilde{n}}\ \mathrm{d}t_1\ldots\mathrm{d}t_n\\
&=\sum_{\hat{n}=0}^{\infty}\underset{\hat{t}_1,\ldots,\hat{t}_{\hat{n}}\in[0..t)}{\int\cdots\int}\ g(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-}))\lambda^{\hat{n}}e^{-\lambda t}\mathrm{d}\hat{t}_1\ldots\mathrm{d}\hat{t}_{\hat{n}}\\
&=\int g\big(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})\big)f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-})\mathrm{d}\hat{\mathcal{N}}_{t-},
\end{aligned}
$$

which completes the proof. □

Proposition 6. *Consider a contract* Γ *that terminates according to* τ *defined in* (60)*, pays the agent* $\beta^{(t)}$ *if there is an arrival at time* t*, and zero if not. That is,* $\mathrm{d}L_t=\beta^{(t)}\mathrm{d}N_t$*. This contract* Γ *satisfies* (IC)*, in which the full-effort process in the continuous time is defined as*

$$
\bar{\boldsymbol{\nu}} := \{\nu_t = 1\}_{t\in[0,T)}.
$$

Proof: For any effort process $\boldsymbol{\nu}$ and trajectory $\mathcal{N}_t$, define $\mathcal{F}_{\boldsymbol{\nu}}(\mathcal{N}_t)$ as a trajectory generated by the counting process $\{\hat{N}_s\}$, such that $\mathrm{d}\hat{N}_s=\mathrm{d}N_s$ if $\nu_s(\mathcal{N}_{s-})=1$, and $\mathrm{d}N_s=0$ if $\nu_s(\mathcal{N}_{s-})=0$. The function $\mathcal{F}_{\boldsymbol{\nu}}$ is well defined because for any $\hat{\mathcal{N}}_t$, there exists a unique $\mathcal{N}_t$ such that $\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_t)=\mathcal{N}_t$ according to the generation process. Fix any effort process $\boldsymbol{\nu}$ and our contract Γ, following (46) we have

$$
\begin{aligned}
W(\Gamma,\boldsymbol{\nu}) &= \mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^{\tau}\big(\beta^{(t)}\mathrm{d}N_t-b\nu_t\mathrm{d}t\big)\right]\\
&= P_0\mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^{\tau}\big(\beta^{(t)}\mathrm{d}N_t-b\nu_t\mathrm{d}t\big)\mid\overline{\lambda}\right]+(1-P_0)\mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^{\tau}\big(\beta^{(t)}\mathrm{d}N_t-b\nu_t\mathrm{d}t\big)\mid\underline{\lambda}\right].
\end{aligned}
\tag{EC.4.3}
$$

For any $\lambda\in\{\overline{\lambda},\underline{\lambda}\}$, we have

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^{\tau}\big(\beta^{(t)}\mathrm{d}N_t-b\nu_t\mathrm{d}t\big)\mid\lambda\right]\\
&=\mathbb{E}_{\boldsymbol{\nu}}\left[\int_0^{T}\mathbb{1}\{\tau\geq t,\nu_t=1\}\big(\beta^{(t)}\lambda\mathrm{d}t-b\mathrm{d}t\big)\mid\lambda\right]\\
&=\int_0^T\mathbb{E}_{\boldsymbol{\nu}}\left[\mathbb{1}\{\tau\geq t,\nu_t=1\}\mid\lambda\right](\beta^{(t)}\lambda-b)\mathrm{d}t\\
&=\int_0^T\mathbb{E}_{\boldsymbol{\nu}}\left[\mathbb{E}\left[\mathbb{1}\{\tau\geq t,\nu_t=1\}\mid\mathcal{N}_{t-}\right]\mid\lambda\right](\beta^{(t)}\lambda-b)\mathrm{d}t
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^T \mathbb{E}_{\boldsymbol{\nu}}\left[\mathbb{1}\left\{\tau(\mathcal{N}_{t-}) \geq t, \nu_t(\mathcal{N}_{t-}) = 1\right\} \mid \lambda\right](\beta^{(t)}\lambda - b)\mathrm{d}t \\
&= \int_0^T \int_{\mathcal{N}_{t-}} \mathbb{1}\left\{\tau(\mathcal{N}_{t-}) \geq t, \nu_t(\mathcal{N}_{t-}) = 1\right\} f_{\boldsymbol{\nu}}(\mathcal{N}_{t-} \mid \lambda)\mathrm{d}\mathcal{N}_{t-}(\beta^{(t)}\lambda - b)\mathrm{d}t \\
&\overset{(a)}{=} \int_0^T \int_{\hat{\mathcal{N}}_{t-}} \mathbb{1}\left\{\tau(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) \geq t, \nu_t(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) = 1\right\} f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \lambda)\mathrm{d}\hat{\mathcal{N}}_{t-}(\beta^{(t)}\lambda - b)\mathrm{d}t, \qquad \text{(EC.4.4)}
\end{aligned}
$$

where $f_{\boldsymbol{\nu}}(\mathcal{N} \mid \lambda)$ is as defined in (55). The first equality follows from $\left(\beta^{(t)}\mathrm{d}N_t - b\nu_t\mathrm{d}t\right) = 0$ when the agent exerts no effort and the fact that the agent exerts no effort after being terminated. The second equality follows from the Fubini's theorem, given that the expectation of the integral must be finite. The third equality follows from the Tower rule of calculating conditional expectation, and the forth from that both τ and ν_t adapt to $\mathcal{N}_t$. The sixth equality follows from the re-writing the expectation, and finally, (a) follows from the Lemma EC.4.

Substituting (EC.4.4) into (EC.4.3) we have

$$
\begin{aligned}
W(\Gamma, \boldsymbol{\nu}) = &\int_0^T \int_{\hat{\mathcal{N}}_{t-}} \mathbb{1}\left\{\tau(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) \geq t, \nu_t(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) = 1\right\} \\
&\left[\beta^{(t)}\left(P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda})\overline{\lambda} + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\underline{\lambda}\right) - \left(P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda}) + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\right) b\right] \mathrm{d}\hat{\mathcal{N}}_{t-}\mathrm{d}t \\
\overset{(b)}{\leq} &\int_0^T \int_{\hat{\mathcal{N}}_{t-}} \mathbb{1}\left\{\tau(\hat{\mathcal{N}}_{t-}) \geq t, \bar{\nu}_t(\hat{\mathcal{N}}_{t-}) = 1\right\} \\
&\left[\beta^{(t)}\left(P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda})\overline{\lambda} + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\underline{\lambda}\right) - \left(P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda}) + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\right) b\right] \mathrm{d}\hat{\mathcal{N}}_{t-}\mathrm{d}t \\
= \ &P_0\mathbb{E}_{\bar{\boldsymbol{\nu}}}\left[\int_0^\tau \left(\beta^{(t)}\mathrm{d}N_t - b\mathrm{d}t\right) \mid \overline{\lambda}\right] + (1-P_0)\mathbb{E}_{\bar{\boldsymbol{\nu}}}\left[\int_0^\tau \left(\beta^{(t)}\mathrm{d}N_t - b\mathrm{d}t\right) \mid \underline{\lambda}\right] = \ W(\Gamma, \bar{\boldsymbol{\nu}}).
\end{aligned}
$$

In order to show the inequality (b), we need to argue (1)

$$
\mathbb{1}\left\{\tau(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) \geq t, \nu_t(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) = 1\right\} \leq \mathbb{1}\left\{\tau(\hat{\mathcal{N}}_t) \geq t, \bar{\nu}_t(\hat{\mathcal{N}}_t) = 1\right\}, \qquad \text{(EC.4.5)}
$$

and (2)

$$
\begin{aligned}
&\mathbb{1}\left\{\tau(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) \geq t, \nu_t(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) = 1\right\} = 1 \Rightarrow \\
&\quad \beta^{(t)}\left(P_0 f_{\bar{\nu}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda})\overline{\lambda} + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\underline{\lambda}\right) - \left(P_0 f_{\bar{\nu}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda}) + (1-P_0) f_{\bar{\nu}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\right) b \geq 0. \qquad \text{(EC.4.6)}
\end{aligned}
$$

We first prove (EC.4.5). Let $\mathcal{N}_t = \mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_t)$. According to the definition of $\mathcal{F}_{\boldsymbol{\nu}}$, we have $N_s \leq \hat{N}_s$ for any $s \in [t]$, which implies that $\tau(\mathcal{N}_{t-}) \leq \tau(\hat{\mathcal{N}}_t)$ according to the termination rule defined in (60). As a result, we have

$$
\begin{aligned}
\mathbb{1}\left\{\tau(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) \geq t, \nu_t(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) = 1\right\} &\leq \mathbb{1}\left\{\tau(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) \geq t\right\} \\
&= \mathbb{1}\left\{\tau(\mathcal{N}_{t-}) \geq t\right\} \\
&\leq \mathbb{1}\left\{\tau(\hat{\mathcal{N}}_{t-}) \geq t\right\} \\
&= \mathbb{1}\left\{\tau(\hat{\mathcal{N}}_{t-}) \geq t, \bar{\nu}_t(\hat{\mathcal{N}}_{t-}) = 1\right\},
\end{aligned}
$$

where the last equality holds because $\bar{\nu}_t(\hat{\mathcal{N}}_t) = 1$ if $\tau(\hat{\mathcal{N}}_t) \geq t$.

Next we prove (EC.4.6).

$$\begin{aligned}
\text{LHS} = &\left[\beta^{(t)}\left(\frac{P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda})\overline{\lambda}}{P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda}) + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})} + \frac{(1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\underline{\lambda}}{P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda}) + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})}\right) - b\right] \\
&\left(P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda}) + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\right) \mathrm{d}\hat{\mathcal{N}}_{t-}\mathrm{d}t \\
= &\left[\beta^{(t)}\left(\overline{\lambda}\mathbb{P}\left(\lambda = \overline{\lambda} \mid \hat{\mathcal{N}}_{t-}\right) + \underline{\lambda}\mathbb{P}\left(\lambda = \underline{\lambda} \mid \hat{\mathcal{N}}_{t-}\right)\right) - b\right]\left(P_0 f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \overline{\lambda}) + (1-P_0) f_{\bar{\boldsymbol{\nu}}}(\hat{\mathcal{N}}_{t-} \mid \underline{\lambda})\right) \mathrm{d}\hat{\mathcal{N}}_{t-}\mathrm{d}t
\end{aligned}$$

Since we have proved when $\tau(\mathcal{F}_{\boldsymbol{\nu}}(\hat{\mathcal{N}}_{t-})) \geq t$, we have $\tau(\hat{\mathcal{N}}_{t-}) \geq t$. According to (60), we have $\hat{N}_{t-} \geq a_{t-}$, which, together with (48) and (62), implies that $\mathbb{P}\left(\lambda = \overline{\lambda} \,|\, \hat{\mathcal{N}}_{t-}\right) \geq P^{(t^-)}$. Thus according to the definition of $\beta^{(t)}$ following (62), we have

$$\begin{aligned}
\beta^{(t)}\left(\overline{\lambda}\mathbb{P}\left(\lambda = \overline{\lambda} \mid \hat{\mathcal{N}}_{t-}\right) + \underline{\lambda}\mathbb{P}\left(\lambda = \underline{\lambda} \mid \hat{\mathcal{N}}_{t-}\right)\right) - b &\geq \beta^{(t)}(P^{(t^-)}\overline{\lambda} + (1-P^{(t^-)})\underline{\lambda}) - b \\
&= \beta^{(t)}(P^{(t)}\overline{\lambda} + (1-P^{(t)})\underline{\lambda}) - b = 0 \ ,
\end{aligned}$$

which means that (EC.4.6) is true. This completes the proof. □

EC.4.6. Proof of Lemma 13

Lemma 13. *For any $t \in [0, T)$, we have*

$$\begin{aligned}
&\mathbb{P}\left(N_t < a_t \,|\, \overline{\lambda}\right) \leq e^{-t\varepsilon_t^2/((e-1)^2\overline{\lambda})} = \frac{1}{T^2}, \ \textit{for } \varepsilon_t \leq \overline{\lambda}, \ \textit{and} \\
&\mathbb{P}\left(N_t \geq a_t \,|\, \underline{\lambda}\right) \leq e^{-t(\overline{\lambda}-\underline{\lambda}-\varepsilon_t)^2/((e-1)^2\underline{\lambda})}, \ \textit{for } \varepsilon_t \leq \min\{\overline{\lambda}-\underline{\lambda}, (e-1)\underline{\lambda}\}.
\end{aligned} \tag{65}$$

Proof: Given the type of the agent, N_t obeys a Poisson distribution with parameter $\lambda \in \{\overline{\lambda}, \underline{\lambda}\}$. According to *Bennett's Inequality* (Bennett 1962), for the Poisson random variable N_t with parameter λ, we have

$$\begin{aligned}
&\mathbb{P}\left(N_t - \lambda t \geq t\varepsilon \,|\, \lambda\right) \leq \exp\left(-t\lambda h\left(\frac{\varepsilon}{\lambda}\right)\right), \text{ and} \\
&\mathbb{P}\left(\lambda t - N_t \geq t\varepsilon \,|\, \lambda\right) \leq \exp\left(-t\lambda h\left(\frac{\varepsilon}{\lambda}\right)\right)
\end{aligned}$$

where $h(u) := (1+u)\ln(1+u) - u$. When $u \in [0, \ e-1]$, we have $h(u) \geq \frac{u^2}{(e-1)^2}$, which implies that when $\varepsilon \leq (e-1)\lambda$, we have

$$\begin{aligned}
&\mathbb{P}\left(N_t - \lambda t \geq t\varepsilon \,|\, \lambda\right) \leq \exp\left(-\frac{t\varepsilon^2}{\lambda(e-1)^2}\right), \text{ and} \\
&\mathbb{P}\left(\lambda t - N_t \geq t\varepsilon \,|\, \lambda\right) \leq \exp\left(-\frac{t\varepsilon^2}{\lambda(e-1)^2}\right)
\end{aligned}$$

Then we have when $\varepsilon_t \leq (e-1)\overline{\lambda}$,

$$\mathbb{P}\left(N_t < a_t \,|\, \overline{\lambda}\right) = \mathbb{P}\left(N_t < \overline{\lambda}t - t\varepsilon_t \,|\, \overline{\lambda}\right) \leq e^{-t\varepsilon_t^2/((e-1)^2\overline{\lambda})} = \frac{1}{T^2}, \tag{EC.4.7}$$

and when $\varepsilon_t \le \min\{\overline{\lambda}-\underline{\lambda}, (e-1)\underline{\lambda}\}$,

$$\mathbb{P}\left(N_t \ge a_t \mid \underline{\lambda}\right) = \mathbb{P}\left(N_t \ge \overline{\lambda}t - t\varepsilon_t \mid \underline{\lambda}\right) = \mathbb{P}\left(N_t - \underline{\lambda}t \ge t(\overline{\lambda}-\underline{\lambda}-\varepsilon_t) \mid \underline{\lambda}\right) \le e^{-t(\overline{\lambda}-\underline{\lambda}-\varepsilon_t)^2/((e-1)^2\underline{\lambda})}.$$

This completes the proof. □

EC.4.7. Proof of Lemma 14

Lemma 14. *We have*

$$\mathbb{E}\left[\tau \mid \overline{\lambda}\right] \ge T - \overline{\lambda}, \text{ and } \mathbb{E}\left[\tau \mid \underline{\lambda}\right] \le \alpha'' \ln T + 1/T, \tag{66}$$

where $\alpha'' := \max\{8(e-1)^2/[\overline{\lambda}(\overline{\lambda}-\underline{\lambda})^2],\ 2/[\overline{\lambda}\underline{\lambda}^2]\}$.

Proof: First, we prove the first inequality. From (61) and Lemma 13, we have

$$\mathbb{P}\left(\tau = s_i \mid \overline{\lambda}\right) = \mathbb{P}\left(\bigcap_{j\in[i-1]}\{N_{s_j} \ge a_{s_j}\} \bigcap \{N_{s_i} < a_{s_i}\} \mid \overline{\lambda}\right) \le \mathbb{P}\left(N_{s_i} < a_{s_i} \mid \overline{\lambda}\right) \le \frac{1}{T^2}. \tag{EC.4.8}$$

Then we have

$$\begin{aligned}
\int_{t=0}^{T} \mathbb{P}\left(\tau < t \mid \overline{\lambda}\right) \mathrm{d}t &= \sum_{i\in[a_T]} (s_i - s_{i-1})\mathbb{P}\left(\tau < s_i \mid \overline{\lambda}\right) \\
&= \sum_{i\in[a_T]} (s_i - s_{i-1})\mathbb{P}\left(\bigcup_{k\in[i-1]}\{\tau = s_k\} \mid \overline{\lambda}\right) \\
&\le \sum_{i\in[a_T]} (s_i - s_{i-1}) \sum_{k\in[i-1]} \mathbb{P}\left(\tau = s_k \mid \overline{\lambda}\right) \\
&\le \sum_{i\in[a_T]} (s_i - s_{i-1}) \sum_{k\in[i-1]} \frac{1}{T^2} \\
&= \frac{1}{T^2} \sum_{i\in[a_T]} (s_i - s_{i-1})(i-1) \\
&\le \frac{1}{T^2} \sum_{i\in[a_T]} (s_i - s_{i-1})\overline{\lambda}T \\
&\le \overline{\lambda}
\end{aligned}$$

where the first inequality uses the union bound, the second inequality follows (EC.4.8), the third inequality follows from the fact that $a_T < T$, and the last inequality follow from $s_{a_T} < T$.

Next we prove the second inequality. When $t \ge \alpha'' \ln T$, we have $\varepsilon_t \le \min\{(\overline{\lambda}-\underline{\lambda})/2,\ (e-1)\underline{\lambda}\}$, implying that $\mathbb{P}\left(N_t \ge a_t \mid \underline{\lambda}\right) \le e^{-t(\overline{\lambda}-\underline{\lambda})/(4\underline{\lambda}(e-1)^2)}$ according to (65). Thus according to the terminating rule and Lemma 13, we have

$$\mathbb{E}\left[\tau \mid \underline{\lambda}\right] = \int_{t=0}^{T} \mathbb{P}\left(\tau \ge t \mid \underline{\lambda}\right) \mathrm{d}t$$

$$\begin{aligned}
&= \sum_{i\in[a_T]} (s_i - s_{i-1})\mathbb{P}\left(\bigcap_{k\in[i]}\{N_{s_k} \geq a_{s_k}\} \mid \underline{\lambda}\right) \\
&\leq \sum_{i\in[a_T]} (s_i - s_{i-1})\mathbb{P}\left(N_{s_i} \geq a_{s_i} \mid \underline{\lambda}\right) \\
&\leq \alpha'' \ln T + \sum_{i\in[a_T]: s_i > \alpha'' \ln T} (s_i - s_{i-1}) e^{-s_i(\overline{\lambda}-\underline{\lambda})^2/(4(e-1)^2\underline{\lambda})} \\
&\leq \alpha'' \ln T + \sum_{i\in[a_T]: s_i > \alpha'' \ln T} (s_i - s_{i-1}) e^{-\alpha'' \ln T(\overline{\lambda}-\underline{\lambda})^2/(4(e-1)^2\underline{\lambda})} \\
&\leq \alpha'' \ln T + T e^{-\alpha'' \ln T(\overline{\lambda}-\underline{\lambda})^2/(4(e-1)^2\underline{\lambda})} \\
&\leq \alpha'' \ln T + \frac{1}{T}, \qquad \text{(EC.4.9)}
\end{aligned}$$

where the last inequality uses $\alpha'' \geq 8(e-1)^2/[\overline{\lambda}(\overline{\lambda}-\underline{\lambda})^2$. This completes the proof. □

EC.4.8. Technical Lemma EC.5

Lemma EC.5. *For any $0 < \underline{\lambda} < \overline{\lambda}$, we have*

$$\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}} e^{(\overline{\lambda}-\underline{\lambda})} < 1. \qquad \text{(EC.4.10)}$$

Proof: It is equivalent to prove that $\overline{\lambda}\left(\ln \underline{\lambda} - \ln \overline{\lambda}\right) + \overline{\lambda} - \underline{\lambda} < 0$. Let $\Delta := \overline{\lambda} - \underline{\lambda} > 0$. Define a function $g(\Delta) := (\underline{\lambda} + \Delta)(\ln \underline{\lambda} - \ln(\underline{\lambda} + \Delta)) + \Delta$. Then take the derivative of g we have

$$g'(\Delta) = \ln \underline{\lambda} - \ln(\underline{\lambda} + \Delta) + (\underline{\lambda} + \Delta)(-\frac{1}{(\underline{\lambda} + \Delta)}) + 1 = \ln \underline{\lambda} - \ln(\underline{\lambda} + \Delta) < 0,$$

which implies that $g(\Delta)$ is decreasing in Δ when $\Delta \geq 0$. Consequently, when $\Delta > 0$, we have

$$g(\Delta) < g(0) = 0,$$

which completes the proof. □

EC.4.9. Proof of Lemma 15

Lemma 15. *For any $k = 1, 2\ldots$, if $t \geq (c_2'/c_1')^2(1+3k)\ln T$, we have*

$$\beta^{(t)} - \beta \leq c_3 e^{c_3' k \ln T}, \qquad (67)$$

where c_3 is defined in (36), and c_1', c_2', c_3' are negative constants defined as follows:

$$c_1' := \ln\left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}} e^{(\overline{\lambda}-\underline{\lambda})}\right),\ c_2' := 2(e-1)\sqrt{\overline{\lambda}}\ln\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right),\ c_3' := \frac{c_2'^2}{c_1'}.$$

Proof: According to (EC.3.7), we have

$$\beta^{(t)} - \beta \leq \min\left\{\overline{\beta} - \beta, \frac{b\left(\overline{\lambda} - \underline{\lambda}\right)}{\overline{\lambda}^2}\left(\frac{1}{P^{(t-1)}} - 1\right)\right\}.$$

From (62) we have

$$\begin{aligned}\frac{1}{P^{(t)}} - 1 &= \left(\frac{1}{P_0} - 1\right)\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}t(1-\varepsilon_t)} e^{(\overline{\lambda}-\underline{\lambda})t} \\ &= \left(\frac{1}{P_0} - 1\right)\left[\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}} e^{(\overline{\lambda}-\underline{\lambda})}\right]^t \left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{-2(e-1)\sqrt{\overline{\lambda}t\ln(T)}}\end{aligned}$$

Take ln on both sides we have

$$\begin{aligned}\ln\left(\frac{1}{P^{(t)}} - 1\right) \leq\ & \ln\left(\frac{1}{P_0} - 1\right) + t\ln\left(\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right)^{\overline{\lambda}} e^{(\overline{\lambda}-\underline{\lambda})}\right) - \sqrt{t}\cdot 2(e-1)\sqrt{\overline{\lambda}\ln T}\ln\left(\frac{\underline{\lambda}}{\overline{\lambda}}\right) \\ =\ & \ln\left(\frac{1}{P_0} - 1\right) + \sqrt{t}^2 c_1' - \sqrt{t}c_2'\sqrt{\ln T}.\end{aligned} \tag{EC.4.11}$$

For simplicity, let $c_{2,T}' := c_2'\sqrt{\ln T}$, and $c_{3,T}' := c_3' \ln T$. Denote function $C : \mathbb{R} \to \mathbb{R}$ as

$$C(u) = c_1' u^2 - c_{2,T}' u.$$

According to (EC.4.10) we have $c_1' < 0$, thus $C(u)$ is decreasing in u when $u \geq c_{2,T}'/2c_1'$. Then we will prove that, for any $k = 1, 2, \ldots$, if $u \geq (c_{2,T}'/c_1')\sqrt{1+3k}$, then $C(u) \leq c_{3,T}'k$. Using the formula of roots of a quadratic equation, we have $C(u) \leq c_{3,T}'k$ if

$$u \geq \frac{-c_{2,T}' + \sqrt{c_{2,T}'^2 + c_1' c_{3,T}' k}}{-2c_1'}. \tag{EC.4.12}$$

Using the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for any $a, b > 0$, we have

$$\sqrt{c_{2,T}'^2 + c_1' c_{3,T}' k} \leq -c_{2,T}' + \sqrt{c_1' c_{3,T}' k} = -c_{2,T}'(1 + \sqrt{k}).$$

Thus we have

$$\frac{-c_{3,T}' + \sqrt{c_{2,T}'^2 + c_1' c_{3,T}' k}}{-2c_1'} \leq \frac{-c_{2,T}' - c_{2,T}'(1+\sqrt{k})}{-2c_1'} = \frac{c_{2,T}'}{c_1'}(1+\sqrt{k}).$$

Since $\sqrt{k} \leq k$ when $k \geq 1$, we have

$$1 + \sqrt{k} = \sqrt{1 + k + 2\sqrt{k}} \leq \sqrt{1+3k}.$$

Combining the above two inequalities, we have

$$\frac{-c_{2,T}' + \sqrt{c_{2,T}'^2 + c_1' c_{3,T}' k}}{-2c_1'} \leq \frac{c_{2,T}'}{c_1'}\sqrt{1+3k},$$

which together with (EC.4.12) implies that if $u \geq (c_2'\sqrt{\ln T}/c_1')\sqrt{1+3k}$, then $C(u) \leq c_3' k \ln T$. □

EC.4.10. Proof of Lemma 16

Lemma 16. *For any $T \geq e$, we have*

$$\int_{t=0}^{T} (\beta^{(t)} - \beta)\mathrm{d}t \leq 4c_3 \left(\frac{c_2'}{c_1'}\right)^2 \ln T + \frac{3c_3}{-c_3'}. \tag{68}$$

Proof: From Lemma 15 we have

$$\begin{aligned}
\int_{t=0}^{T} (\beta^{(t)} - \beta)\mathrm{d}t &\leq 4c_3 \left(\frac{c_2'}{c_1'}\right)^2 \ln T + \int_{k=1}^{(T(c_1'/c_2')^2-1)/3} 3c_3 e^{c_3'(k-1)\ln T} dk \\
&= 4c_3 \left(\frac{c_2'}{c_1'}\right)^2 \ln T + 3c_3 \int_{k=0}^{(T(c_1'/c_2')^2/\ln T-4)/3} e^{c_3' k \ln T} dk \\
&\leq 4c_3 \left(\frac{c_2'}{c_1'}\right)^2 \ln T + 3c_3 \cdot \frac{1}{-c_3' \ln T}.
\end{aligned}$$

Since $\ln T \geq 1$ when $T \geq e$, we complete the proof. □

EC.4.11. Proof of Theorem 5

Theorem 5. *For any instance of the online dynamic contract design problem in continuous time with a time horizon of length T, let Γ be the contract designed for the continuous time situation, that terminates according to τ defined in (60), and pays the agent $\beta^{(t)}$ according to (63) if and only if there is an arrival at time t. The total expected regret of implementing contract Γ satisfies*

$$\mathit{Reg}(\Gamma, T) = O(\ln T).$$

Proof: The proof is similar to that of Theorem 2, and thus is omitted for brevity. □