Research Institute for Environmental Economics and Management

Discussion Paper Series No.2001

Optimal Dynamic Incentive Contracts between a Principal and Multiple Agents in Controlled Markov Processes: A Constructive Approach

Yasuaki Wasa, Ken-Ichi Akao, and Kenko Uchida

February 2020

WASEDA

Optimal Dynamic Incentive Contracts between a Principal and Multiple Agents in Controlled Markov Processes: A Constructive Approach

Yasuaki Wasa, Ken-Ichi Akao, and Kenko Uchida

Abstract

This paper investigates an optimal dynamic incentive contract between a risk-averse principal (system operator) and multiple risk-averse agents (subsystems) with independently local controllers in continuous-time controlled Markov processes, which can represent various cyber-physical systems. The principal's incentive design and the agents' decision-makings under asymmetric information structure are known as the principal-agent (PA) problems in economic field. However, the standard framework in economics cannot be directly applied to the realistic control systems including large-scale cyber-physical systems and complex networked systems due to some unrealistic assumptions for an engineering perspective. In this paper, using a constructive approach based on the techniques of the classical stochastic control theory, we propose and solve a novel dynamic control/incentive synthesis for the PA problem under moral hazard.

Index Terms

Principal-agent problems, Moral hazard, Cyber-physical systems, Multi-agent systems, Dynamic programming, Risk-sensitive stochastic control, Differential games

Yasuaki Wasa is with the Department of Electrical Engineering and Bioscience, Waseda University, Tokyo 169-8555, Japan. wasa@aoni.waseda.jp.

Ken-Ichi Akao is with the Graduate School of Social Sciences, Waseda University, Tokyo 169-8050, Japan. akao@waseda.jp.

Kenko Uchida is with the Research Institute for Science and Engineering, Waseda University, Tokyo 169-8555, Japan. kuchida@waseda.jp.

This work was supported by JST CREST Grant Number JPMJCR15K2.

I. Introduction

Large-scale infrastructure systems are composed of subsystems whose interests conflict. For analysis, control and synthesis of these systems, there becomes a need to develop a novel system-theoretic framework which requires well-ordered decentralized, distributed and hierarchical network control while taking the different types of decision-makers into account (see e.g. [1], [2]). To achieve this from control systems perspectives, notable examples include the local-action based approach [3], the multi-layered control architecture based on time-scale decomposition [4], [5] and the passivity-based approach [6].

In this paper, we address a novel control/incentive synthesis problem motivated by the contract theory, which is a quite different approach from the above control systems literature [3]–[6]. We focus on standard cyber-physical systems including typical infrastructure systems in the presence of a dynamic principal (system operator and utility) and multiple agents (subsystems) whose interests conflict and are dynamically interdependent. The proposed control/incentive synthesis problem is to maximize the principal's profit while ensuring all the agents' profit maximization. Actually, the principal and the agents independently take control/incentive maximizing their own profit by using mutually different available information, which is known as hidden action type asymmetric information and moral hazard [7], [8]. The decision-making problem under the above physical model and information structures is called principal-agent (PA) problem and their contributions are well-established as contract theory [7], [8].

In this paper, we discuss PA problems such that the infrastructure system obeys a standard continuous-time controlled Markov model [9], [10], the risk-averse multiple agents have independent controls on a finite time interval and the risk-averse principal has incentive variables to give the agents some rewards. Recently, the contract theory for such continuous-time dynamical systems have been remarkably evolved in economic field [11]–[19]. The dynamics discussed in [11]–[19] are economic models, which is rather different from the physical-based control systems with the limited control dimension. Since we cannot directly apply the solution presented in [11]–[19] to our problems taking the physical-based control systems into account, we need to develop a novel control/incentive synthesis introduced below.

The contributions of this paper are summarized as follows.

First, we focus on the framework of the system model handled in the PA problems with the standard continuous-time controlled Markov processes such that the system model is linear in control variables. The Markovian control framework includes various infrastructure systems and cyber-physical systems (see e.g. [9], [10] and therein). In all the papers [11]–[17] except [18], [19], the partial derivative of the value function with respect to the state variable, called the shadow price or the adjoint variable, is expressed as a function of the agents' optimal control led by the so-called first-order condition (FOC).

The FOC corresponds to the stationary condition of the Hamiltonian with respect to the control variable in the maximum principle or dynamic programming. To proceed the FOC approach, the following two technical assumptions of system models are required. The first assumption is that the optimal control must be interior to its (compact) domain. [1] The second assumption is that the partial derivative of the vector field with respect to the control parameter at any time can be arbitrarily changed. Note that the approach proposed in [18] does not use the FOC, while the second assumption above holds. As control systems satisfying the FOC are extremely limited in the real world, this paper presents a novel approach without the FOC in order to theoretically guarantee the optimality of control/incentive.

Second, we formulate our dynamic contract problem in the classical continuous-time Markovian control framework [9], [10], and present a constructive method leading to an optimal contract (controls and incentives). The previous works [11]–[19] for continuous-time models are developed in the weak solution framework based on the measure transformation and the martingale representation [20]–[22]; the weak solution formulation can discuss quite general stochastic control problems, and is suitable for qualitative analyses but not good at constructing actually incentives and controls; actually, the design examples in [12], [13], [19] are re-formulated as the classical Markovian control problem. In view of this, we formulate our contract problem with the so-called strong solution framework dealing with the classical Markovian control problem, so that, though the formulation is less general than the weak solution formulation, we can develop a constructive approach in an intuitive way to our contract problem based on the Hamilton-Jacobi-Bellman (HJB) equations and their classical solutions.

Third, our approach can be applied to multi-agent systems with mutual interests/conflict and dynamic games. On the existing PA problems with continuous-time dynamical systems, as far as the authors know, all the papers [11]–[13], [15], [16], [18], [19] except [14], [17] handle continuous-time dynamical systems composed of a single principal and a *single* agent. In case of *interactive* multi-agent systems, we cannot directly apply the approach for the single-principal single-agent in [11]–[13], [15], [16], [18] to our problem and we must develop a new method to solve our problem. The paper [14], [17] are the literature for interactive multi-agent systems in continuous-time stochastic dynamical systems. However, the paper [14], [17] adopt the FOC approach in the weak solution formulation, and so should face the two issues stated above.

Lastly, regarding the contract problems under moral hazard *in engineering*, there are relatively few papers [23]–[25]. The paper [23] applies a PA-type moral hazard problem to a differential game between

[1] Even if the control range is a compact set, we can eliminate the first assumption by adding a suitable penalty term to the objective function.

single-principal and single-agent in dynamical cybersecurity management. The paper [24] considers the contract problem with both adverse selection and moral hazard between single-principal and single-agent in static systems. Our previous paper [25] proposes an optimal control/incentive synthesis in the electricity regulation market with discrete-time dynamical systems. However, the papers [23]–[25] do not reveal a rigorous constructive method leading to an optimal contract between the principal and the multiple agents, whereas this paper theoretically addresses a model-based optimal control/incentive design in a systematic manner.

In summary, we formulate a novel type of control/incentive synthesis problem based on the PA problems in the presence of a single principal and interactive *multiple* agents with mutual interests/conflict and present a constructive approach without FOC within the classical framework of the continuous-time controlled Markov processes.

II. Mathematical Notation

The n-dimensional Euclidean space is denoted by $\mathbb{R}^n$. The partial derivative operators with respect to a variable x are given by $\nabla_x := \frac{\partial}{\partial x}$ and $\nabla_x^2 := \frac{\partial^2}{\partial x^2}$. A function $f(x)$ is called *of class* $\mathcal{C}^r$ at $x \in \mathcal{X}$ if all its derivatives of orders $\leq r$ are continuous in a neighborhood of x. We denote by $\mathcal{C}^r(\mathcal{X})$ the set of $\mathcal{C}^r$ class functions on $\mathcal{X}$. A function $g(x,y)$ is called of class $\mathcal{C}^{r,l}(\mathcal{X}\times\mathcal{Y})$, sometimes denoted by $g \in \mathcal{C}^{r,l}(\mathcal{X}\times\mathcal{Y})$, if all partial derivatives of orders $\leq r$ at $x \in \mathcal{X}$ and of order $\leq l$ at $y \in \mathcal{Y}$ are continuous on $\mathcal{X}\times\mathcal{Y}$. If $r = l$, then we denote $g \in \mathcal{C}^r(\mathcal{X}\times\mathcal{Y})$. Let $\mathbb{E}_{t,x}$ and $\mathbb{E}_{t,x}[v]$ denote the expectation operator and the conditional expectation of v given (t,x), respectively. The notations X' and $\mathrm{tr}[Y]$ stand for transposition of a matrix or a vector X and the trace of a square matrix Y, respectively. For notational simplicity, the functionals with time t (e.g. $f(t,x)$) will be sometimes written as the functionals with subscript t (i.e. $f_t(x)$). Furthermore, meaningless arguments of functions and functionals will be sometimes omitted.

III. Problem Formulation

In this section, we formulate our problem within a classical framework under the standard technical assumptions in the risk-sensitive stochastic control and differential games [9], [26]. We also develop our arguments based on the fundamental results and notations in [10].

This paper considers N risk-averse agents (subsystems) and a risk-averse principal (system operator). Each agent $i \in \mathcal{N} := \{1,2,\ldots,N\}$ executes $u^i \in \mathbb{R}^{m_i}$ independently. Let u and u^{-i} denote the collection of all control profile and that of the control profile expect u^i, i.e., $u := (u^1,\ldots,u^N)$ and

$u^{-i} := (u^1, \ldots, u^{i-1}, u^{i+1}, \ldots, u^N)$, respectively. Suppose that the state $x(\tau) \in \mathbb{R}^n$ during the time period $\tau \in [0, T]$ satisfies a stochastic differential equation of the form:

$$dx_\tau = f(\tau, x(\tau), u(\tau))d\tau + \sigma(\tau, x(\tau))d\beta_\tau, \tag{1}$$
$$f(\tau, x(\tau), u(\tau)) := f^0(\tau, x(\tau)) + \sum_{i \in \mathcal{N}} f^i(\tau, x(\tau))u^i(\tau),$$

where β_τ is a $q(\geq n)$-dimensional standard Brownian motion, $f^0 : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $f^i : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m_i}$, $i \in \mathcal{N}$, and $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n \times q}$. The system function f is linear in the control variable u^i. The admissible control u^i of agent $i \in \mathcal{N}$ at time $t \in [0,T]$ is given by only the current state x, i.e., a Markov control policy $u^i : [0,T] \times \mathbb{R}^n \to U^i$, where U^i is a compact subset of $\mathbb{R}^{m_i}$. In this paper, let Γ^i denote the set of all such admissible decision rules of agent i, that is, if $u^i \in \Gamma^i$, $u^i(t,x)$ is continuous at $t \in [0,T]$ and Lipschitz continuous at $x \in \mathbb{R}^n$. Let us denote by Γ^{-i} the admissible control set of u^{-i}. With these notations, we will sometimes write $u \in \Gamma$ as $(u^i, u^{-i}) \in \Gamma^i \times \Gamma^{-i}$.

To guarantee the existence of a classical solution of the system equation (1), we assume the standard regularity conditions on the system functions $f^i(t,x)$, $i \in \{0\} \cup \mathcal{N}$, and $\sigma(t,x)$, that is **(A1)** $f^i, \sigma \in \mathcal{C}^1([0,T] \times \mathbb{R}^n)$ *such that* f^i, σ, $\nabla_x f^i$ *and* $\nabla_x \sigma$ *are bounded on* $[0,T] \times \mathbb{R}^n$. Under this assumption, for an admissible control $u \in \Gamma := \Gamma^1 \times \ldots \times \Gamma^N$ and an initial condition $x(t) = x$, $(t,x) \in [0,T] \times \mathbb{R}^n$, we have a unique and continuous sample solution of the equation (1) (see, e.g. [10, Theorem V4.1]). Let us denote by $X_t(x,u)$ the solution, that is the state trajectory along the controlled Markov diffusions (1) with the initial condition $(t,x) \in [0,T] \times \mathbb{R}^n$ and the control $u \in \Gamma$. The set of $X_t(x,u)$ is denoted by $\mathcal{X}_t$. Furthermore, for developing our dynamic programming (HJB equation) based approach in a mathematically sound way, we assume **(A2)** σ *is of class* $\mathcal{C}^2$ *at* $x \in \mathbb{R}^n$ *and* σ^{-1} *is bounded on* $[0,T] \times \mathbb{R}^n$. This assumption implies that the HJB equation is a uniformly parabolic equation (see [10, Chapter VI]).

To implement the dynamical system, each agent $i \in \mathcal{N}$ independently decides its own control $u^i(t,x)$ according to the state x at time t. In this case, the resulting system behavior is generally different from what the principal desires. To incentivize the agents to take a suitable control for the principal's objective, let us formulate the salary functional $W^i : [0,T] \times \mathcal{X}_0 \to \mathbb{R}$ from a principal to an agent $i \in \mathcal{N}$. [2] In this paper, following [11]–[15], the salary functional along the trajectory (1) is given by

$$W^i(t, X_t(x,u)) = w^{iT}(x_T) + w^{i0}(t,x) + \int_t^T w^{it}(\tau, x_\tau)d\tau + \int_t^T w^{ix}(\tau, x_\tau)dx_\tau. \tag{2}$$

[2]As dx_τ includes w^{it} and w^{ix} defined in (2) are mathematically redundant. Meanwhile, (2) is basically the same as the normal literature [11]–[15] and we see from Lemma 2 that (2) is the only formulation satisfying the incentive compatibility constraints (6b). Hence, we use (2).

The objective of our problem is that the principal determines the most suitable salary parameters $w := (w^1, \ldots, w^N)$, $w^i = (w^{iT}, w^{i0}, w^{it}, w^{ix}) \in \Pi$, $i \in \mathcal{N}$, *before the agents' implementation*. The salary parameter w^i is composed of the salary at terminal time $w^{iT} : \mathbb{R}^n \to \mathbb{R}$, the salary at initial time $w^{i0} : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ and two kinds of salaries during the transient period, $w^{it} : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ and $w^{ix} : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{1\times n}$, in the salary functional W^i. The set Π is defined as the set of all feasible salary parameters w^i obeying the following conditions: w^{i0} is continuous on $[0,T] \times \mathbb{R}^n$, $w^{it}, w^{ix} \in \mathcal{C}^1([0,T] \times \mathbb{R}^n)$ such that w^{it}, w^{ix}, $\nabla_x w^{it}$ and $\nabla_x w^{ix}$ are bounded on $[0,T] \times \mathbb{R}^n$, and $w^{iT} \in \mathcal{C}^2(\mathbb{R}^n)$ such that w^{iT} and $\nabla_x w^{iT}$ are bounded on $\mathbb{R}^n$, These conditions are the same as those for the cost functions of the agents and the principal introduced below.

Each agent $i \in \mathcal{N}$ has the two cost functions $\varphi^i : \mathbb{R}^n \to \mathbb{R}$ and $l^i : [0,T] \times \mathbb{R}^n \times \Gamma^i \to \mathbb{R}$. Then, the agent's reward functional $\Psi^i : [0,T] \times \mathcal{X}_0 \to \mathbb{R}$ is written by

$$\Psi^i(t, X_t(x,u); w^i) = \varphi^i(x_T) + \int_t^T l^i(\tau, x_\tau, u^i_\tau) d\tau + W^i(t, X_t(x,u)). \tag{3}$$

Given a salary parameter $w^i \in \Pi$, the risk-averse agent i executes its own control $u^i \in \Gamma^i$ maximizing the following profit functional $J_i : [0,T] \times \mathcal{X}_0 \to \mathbb{R}$, i.e.,

$$J_i(t, X_t(x,u); w^i) := \mathbb{E}_{t,x}\left[\nu_i(\Psi^i(t, X_t(x,u); w^i))\right], \tag{4a}$$

$$\nu_i(z) = -\exp(-r_i z), \tag{4b}$$

where $r_i > 0$ is a risk-aversion coefficient. We will show later the technical conditions **(A3)** for the agents' cost functionals together with those for the principal's cost functionals.

Meanwhile, the principal also has the cost functions $\varphi^0 : \mathbb{R}^n \to \mathbb{R}$ and $l^0 : [0,T] \times \mathbb{R}^n \times \Gamma \to \mathbb{R}$. Then, the profit and reward functionals of the risk-averse principal are given by

$$J_0(t, X_t(x,u); w) = \mathbb{E}_{t,x}\left[\nu_0\left(\Psi^0(t, X_t(x,u); w)\right)\right], \tag{5a}$$

$$\nu_0(z) = -\exp(-Rz), \tag{5b}$$

$$\Psi^0(t, X_t(u); w) := \varphi^0(x_T) + \int_t^T l^0(\tau, x_\tau, u_\tau) d\tau - \sum_{i\in\mathcal{N}} W^i(t, X_t(x,u)), \tag{5c}$$

where $R > 0$ is a risk-aversion coefficient. The agent's cost functions l^i and φ^i, $i \in \mathcal{N}$, and the principal's cost functions l^0 and φ^0 satisfy the following assumptions: **(A3)** *$l^i \in \mathcal{C}^1([0,T] \times \mathbb{R}^n \times U^i)$, $i \in \mathcal{N}$, and $l^0 \in \mathcal{C}^1([0,T] \times \mathbb{R}^n \times U)$ such that l^i, $\nabla_x l^i$ are bounded on $[0,T] \times \mathbb{R}^n \times U^i$ and l^0, $\nabla_x l^0$ are bounded on $[0,T] \times \mathbb{R}^n \times U$, and $\varphi^i \in \mathcal{C}^2(\mathbb{R}^n)$, $i \in \{0\} \cup \mathcal{N}$, such that φ^i, $\nabla_x \varphi^i$ are bounded on $\mathbb{R}^n$.* This assumption imposes standard regularity conditions on the cost functions in the risk-sensitivity stochastic control and differential games (see [9], [26]); under the assumptions **(A1)–(A3)**, we can show that the

HJB equations appearing in this paper have solutions of class $\mathcal{C}^{1,2}([0,T]\times\mathbb{R}^n)$, by modifying slightly Theorem VI6.2 in [10] for our problem setting.

The information structure of our contract problem is as follows:[3] **(B1)** Both the principal and the agents share the physical model information (f^0,σ) as public information. The principal receives all agents' model information $(f^i,\varphi^i,l^i,\nu_i)$ a priori; **(B2)** The principal can observe on-line state information x perfectly but cannot access or operate the agents' on-line control information u directly; **(B3)** Each agent $i\in\mathcal{N}$ receives the model information W^i of salary in advance and the on-line salary parameter w^i led by (6) in real-time from the principal.

Now, we can formulate the following optimal incentive contract problem so that the principal determines the optimal salary parameters w:[4] For a given initial condition $x(t)=x$, $(t,x)\in[0,T]\times\mathbb{R}^n$, and a prescribed profit level $k^i(t,x)$ of agent $i\in\mathcal{N}$,

$$\max_{u^\dagger\in\Gamma, w\in\Pi\times\ldots\times\Pi} J_0(t,X_t(x,u^\dagger);w) \tag{6a}$$

$$\text{subject to}\quad u^{i\dagger}\in\arg\max_{u^i\in\Gamma^i} J_i(t,X_t(x,u^i,u^{-i\dagger});w^i),\ \ i\in\mathcal{N}, \tag{6b}$$

$$J_i(t,X_t(x,u^{i\dagger},u^{-i\dagger});w^i)\geq \nu_i(k_t^i(x)),\ \ i\in\mathcal{N}, \tag{6c}$$

where $k^i:[0,T]\times\mathbb{R}^n\to\mathbb{R}$ is continuous on $[0,T]\times\mathbb{R}^n$. Obviously, the optimization problem (6a) of the principal's design parameters, i.e. salary parameters w, depends on the agents' control u. By solving the problem (6), we acquire the optimal salary functionals and the agents' optimal control policy. In contract theory, (6) is called a *PA problem* under moral hazard. The constraints (6b) and (6c) are, respectively, called *incentive compatibility* constraints and *individual rationality* constraints [11]–[16], [18]. The constraint (6b) claims that the salary incentivizes the agents to use an optimal control maximizing its own net reward Ψ^i. Consequently, the profile of control policies satisfying (6b) constitutes a Nash equilibrium. The constraint (6c) guarantees the prescribed profit level.

Note that the principal solves the optimization problem (6) before the agents' implementation and reports the optimal salary functional W^{i*} to each agent $i\in\mathcal{N}$ in advance. The agent $i\in\mathcal{N}$ does not receive the optimal control u^{i*} derived by (6) and can select its own control u^i from U^i without restraint in real-time. Since the principal selects the salary functional W^i satisfying the incentive compatibility property, the reasonably optimal incentive W^{i*} can incentivize the rational agents to take the optimal

[3]See [27] in the context of the ancillary markets in dynamic power grids.

[4]From (6b) and (6c), an optimal u depending on w is obtained. Therefore, to solve (6a), the principal needs to determine not only the primal parameter w but also u associated with w. See Section IV for more details. Furthermore, we consider the generalized control/incentive problem (6) on the future time interval from the arbitrary current time $t\in[0,T]$ to the terminal time T based on the time-consistency property so that we can discuss possibilities of several incentive options.

control u^{i*} desired by the principal; otherwise the irrational agent will reduce its own reward. As a result, the agents have to implement the optimal control u^* in real-time and the principal can know the optimal controls selected by the agents in advance.

IV. Optimal Incentive Contract

This section solves the optimal incentive contract problem (6) and obtaining optimal salary functionals with and the agents' optimal controls.

To solve the problem (6), let us first consider agents' optimization problem (6b) and (6c). Throughout this section, we suppose that, for any salary parameters $w \in \Pi \times \ldots \times \Pi$, time $t \in [0,T]$ and state $x \in \mathbb{R}^n$, there is a tuple of optimal controls $u^* = (u^{i*}, u^{-i*}) \in \Gamma^i \times \Gamma^{-i}$, that is a *Nash equilibrium* [26], defined by

$$u^{i*} \in \arg\max_{u^i \in \Gamma^i} J_i(t, X_t(x, u^i, u^{-i*}); w^i), \;\; i \in \mathcal{N}. \tag{7}$$

Here, the multiple agents' stochastic differential game (6b) decouples into N-separate risk-sensitive stochastic control problems. In other words, each agent $i \in \mathcal{N}$ controls only $u^i \in \Gamma^i$ under the others' Nash control policy u^{-i*}. Meanwhile, given $(t,x) \in [0,T] \times \mathbb{R}^n$, $w^{i0}(t,x)$ is independent of the selection of u^i. Then, given $(t,x) \in [0,T] \times \mathbb{R}^n$ and $w \in \Pi \times \ldots \times \Pi$, let $V^i : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ be the agent i's value function defined by

$$V^i(t,x) := -\nu_i(-w_t^{i0}(x)) \max_{u^i \in \Gamma^i} J_i(t, X_t(x, u^i, u^{-i*}); w^i). \tag{8}$$

We here obtain the following lemma.

Lemma 1: Suppose that, for any $w \in \Pi \times \ldots \times \Pi$, there are a Nash equilibrium $u^* = (u^{i*}, u^{-i*}) \in \Gamma^i \times \Gamma^{-i}$ and the corresponding $V^i \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$. Then, for $i \in \mathcal{N}$, $u^{-i*} \in \Gamma^{-i}$ and $w \in \Pi \times \ldots \times \Pi$, $u^{i*} \in \Gamma^i$ and V^i obey a solution of the following HJB equation for $\tau \in [0,T]$:

$$\begin{aligned}
&\nabla_\tau V_\tau^i(x) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V_\tau^i(x)\sigma_\tau(x)\sigma_\tau(x)'\right] \\
&\quad + \max_{u^i \in U^i}\Big[\nabla_x V_\tau^i(x) f_\tau(x, u^i, u_\tau^{-i*}) - \nabla_x V_\tau^i(x) r_i \sigma_\tau(x)(w_\tau^{ix}(x)\sigma_\tau(x))') \\
&\quad - r_i V_\tau^i(x)\left[l_\tau^i(x,u^i) + w_\tau^{it}(x) + w_\tau^{ix}(x) f_\tau(x,u^i,u_\tau^{-i*})\right] - r_i V_\tau^i(x)\frac{-r_i}{2}|(w_\tau^{ix}(x)\sigma_\tau(x))'|^2\Big] \\
&= \nabla_\tau V_\tau^i(x) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V_\tau^i(x)\sigma_\tau(x)\sigma_\tau(x)'\right] + \nabla_x V_\tau^i(x)\left[f_\tau(x, u_\tau^{i*}, u_\tau^{-i*}) - r_i\sigma_\tau(x)(w_\tau^{ix}(x)\sigma_\tau(x))')\right] \\
&\quad - r_i V_\tau^i(x)\left[l_\tau^i(x,u_\tau^{i*}) + w_\tau^{it}(x) + w_\tau^{ix}(x) f_\tau(x,u_\tau^{i*},u_\tau^{-i*})\right] - r_i V_\tau^i(x)\frac{-r_i}{2}|(w_\tau^{ix}(x)\sigma_\tau(x))'|^2 = 0,
\end{aligned} \tag{9a}$$

$$V^i(T, x_T) = \nu_i(w^{iT}(x_T) + \varphi^i(x_T)). \tag{9b}$$

Proof: See Appendix A. ■

Lemma1 and the following Lemmas 2 and 3 require that $V^i \in \mathcal{C}^{1,2}([0,T]\times\mathbb{R}^n)$, $i\in\mathcal{N}$. The assumptions **(A1)–(A3)** are a sufficient (but generally not necessary) condition for this requirement [10].

From Lemma 1, the optimal control policy u^* is restricted by the HJB equation (9). Then, we can obtain an alternative form of the salary functional W^i based on V^i by using (9). The original idea of the representation of the salary functionals is from [11]; the next lemma is a modification and extension of the representation result in [12] to our framework of multiple agents.

Lemma 2: Suppose that, for any $w\in\Pi\times\ldots\times\Pi$, there are a Nash equilibrium $u^*=(u^{i*},u^{-i*})\in\Gamma$ and $V^i\in\mathcal{C}^{1,2}([0,T]\times\mathbb{R}^n)$ led by (9). Then, $W^i(t,X_t(x,u^i,u^{-i*}))$ can be expressed in the following form:

$$\begin{aligned} W^i(t,X_t(x,u^i,u^{-i*})) &= -\varphi^i(x_T)+h^{i0}(t,x)\\ &-\int_t^T\left[h^{ix}_\tau(x_\tau)f_\tau(x_\tau,u^*_\tau)+l^i_\tau(x_\tau,u^{i*}_\tau)+\frac{-r_i}{2}|(h^{ix}_\tau(x_\tau)\sigma_\tau(x_\tau))'|^2\right]d\tau+\int_t^T h^{ix}_\tau(x_\tau)dx_\tau, \end{aligned} \tag{10}$$

where

$$h^{i0}(t,x) \;:=\; \nu_i^{-1}(V^i(t,x))+w^{i0}(t,x), \tag{11a}$$

$$h^{ix}(t,x) \;:=\; \frac{\nabla_x V^i(t,x)}{(-r_i)V^i(t,x)}+w^{ix}(t,x). \tag{11b}$$

Note that the function ν_i^{-1} is the inverse function of ν_i, i.e., $\nu_i^{-1}(a)=(-1/r_i)\log_e(-a)$.

Proof: See Appendix B. ■

Lemma 2 implies that the HJB equation (9) on V^i requires the salary functional $W^i(t,X_t(x,u^i,u^{-i*}))$ based on (10) along the state trajectory (1) with a control policy (u^i,u^{-i*}) on interval $[t,T]$. [5] It follows from the functional form of (10) that h^{ix} defined by (11b) should be in the class of w^{ix}. From the certainty equivalence property, by comparison (2) and (10), we afterwards only focus on the class of the salary functional $W^i(t,X_t(x,u^i,u^{-i*}))$ with the new salary parameter $w^{i*}=(w^{iT*},w^{i0*},w^{it*},w^{ix*})$ defined by

$$w^{iT*}(x) = -\varphi^i(x), \tag{12a}$$

$$w^{i0*}(t,x) = h^{i0}(t,x), \tag{12b}$$

$$w^{it*}(t,x) = \frac{r_i}{2}|(h^{ix}(t,x)\sigma(t,x))'|^2-h^{ix}(t,x)f(t,x,u^*(t,x))-l^i(t,x,u^{i*}(t,x)), \tag{12c}$$

$$w^{ix*}(t,x) = h^{ix}(t,x). \tag{12d}$$

The new parameter constraints (12) are a necessary condition for optimal incentive. Furthermore, we obtain the following lemma from Lemma 2.

[5]See [11, Theorem 6], [12] for the detailed economic interpretation of the salary functional $W^i(t,X_t(x,u^*))$ defined by (10).

Lemma 3: Suppose that, for any $w \in \Pi \times \ldots \times \Pi$, there are a Nash equilibrium $u^* \in \Gamma$ and the corresponding $V^i \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^n)$ led by (9). Then, for $w^{i*} \in \Pi$, $i \in \mathcal{N}$, (12) with h^i defined by (11), $V^i(t,x) = -1$ and $\nabla_x V^i(t,x) = 0$ for all $(t,x) \in [0,T] \times \mathbb{R}^n$.

Proof: Insert (12) into (9) and V^i is given by the solution of the HJB equation:

$$0 = \nabla_\tau V^i_\tau(x) + \frac{1}{2}\mathrm{tr}\left[\nabla^2_x V^i_\tau(x)\sigma_\tau(x)\sigma_\tau(x)'\right] + \nabla_x V^i_\tau(x)\left[f_\tau(x,u^*_\tau) - r_i\sigma_\tau(x)(h^{ix}_\tau(x)\sigma_\tau(x))')\right], \quad \text{(13a)}$$

$$V^i(T,x_T) = \nu_i(0) = -1, \quad \text{(13b)}$$

$\tau \in [0,T]$. Since (13a) is a linear parabolic equation, we see from the boundary condition (13b) that $V^i(t,x) = -1$ for all $(t,x) \in [0,T] \times \mathbb{R}^n$ is the unique solution of the HJB equation (13). This completes the proof. ■

We see from (8), (12) and Lemma 3 that, for all $(t,x) \in [0,T] \times \mathbb{R}^n$, $J_i(t, X_t(x,u^*); w^*) = \nu_i(h^{i0}(t,x))$.

Meanwhile, from the form of (12), we can parameterize the salary parameter w^{i*} by using a Nash equilibrium u^* and a parameter $h^i = (h^{i0}, h^{ix})$, $i \in \mathcal{N}$. In view of this and the characterization of the value functions by Lemma 3, let us introduce a new parameterization of the salary functional (2) with the salary functions given by (12) and the parameters $h^i = (h^{i0}, h^{ix}) \in \mathcal{H}^0 \times \mathcal{H}^x =: \mathcal{H}$, $i \in \mathcal{N}$, where $h^{i0} : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{1\times n}$, $h^{ix} : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{1\times n}$, and $(h^{i0}, h^{ix}) \in \mathcal{H}^0 \times \mathcal{H}^x$ implies that h^{i0} is continuous on $[0,T] \times \mathbb{R}^n$ and $h^{ix} \in \mathcal{C}^1([0,T] \times \mathbb{R}^n)$ such that h^{ix}, $\nabla_x h^{ix}$ are bounded on $[0,T] \times \mathbb{R}^n$. Here, note that this parameterization includes the salary functionals of the form (10) with (11) that are necessarily derived from the incentive compatibility constraints (6b) and, in other words, given as a necessary condition for a Nash equilibrium of agents' control defined by (7). Moreover, if $N = 1$, then Lemmas 1, 2, and 3 can be reduced to the results of the single agent problem discussed in [11]–[13], [15], [16], [18].

When a set of parameters $h := (h^1, \ldots, h^N) \in \mathcal{H} \times \ldots \times \mathcal{H}$ is fixed, the salary functions w^* defined by (12) and the Nash equilibrium u^* defined by (7) are determined dependently on the parameter h; so as to clarify the relation between parameters, we sometimes denote $W^i(t, X_t(x, u^i, u^{-i*}); u^*, h^i)$ and $w^{i*}(u^*, h^i)$. From now on, we regard $h \in \mathcal{H} \times \ldots \times \mathcal{H}$ as design parameters of the salary functionals and the agents' controls.

Now we can state our main result for the multiple agents' optimization making the incentive compatibility constraints (6b) be fulfilled, which gives a constructive characterization of the Nash equilibrium in Γ for the class of the salary functional W^i, $i \in \mathcal{N}$, defined by (12) with $h^i \in \mathcal{H}$, $i \in \mathcal{N}$.

Theorem 1: For $(t,x) \in [0,T] \times \mathbb{R}^n$, $\bar{u} \in \Gamma$, $u^i \in \Gamma^i$ and $h^i \in \mathcal{H}$, $i \in \mathcal{N}$, let $W^i(t, X_t(x, u^i, \bar{u}^{-i}); \bar{u}, h^i)$ be the salary functional. Then, $u^\dagger \in \Gamma$ is *implementable* (i.e., $u^\dagger$ is a Nash equilibrium) if and only if

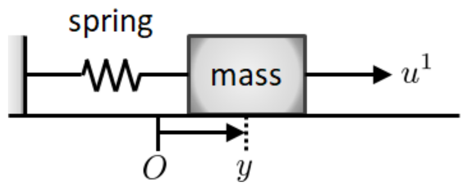

Fig. 1. Mass-spring system.

the following condition holds for all $i \in \mathcal{N}$, $(t,x) \in [0,T] \times \mathbb{R}^n$ and $u^\dagger \in \Gamma$.

$$u_t^{i\dagger}(x) \in \arg\max_{u^i \in U^i} h_t^{ix}(x) f_t(x, u^i, u_t^{-i\dagger}(x)) + l_t^i(x, u^i). \tag{14}$$

Furthermore, for arbitrary $h^i \in \mathcal{H}$, $i \in \mathcal{N}$, the value of the corresponding V^i is the same as one shown in Lemma 3.

Proof: See Appendix C. [6] ■

Under additional convexity assumptions on the control ranges U^i, $i \in \mathcal{N}$, and the cost functions l^i, $i \in \mathcal{N}$, we actually find a Nash equilibrium based on the characterization result of Theorem 1.

Corollary 1: Assume that, for each $i \in \mathcal{N}$, U^i is convex and l^i is of class $\mathcal{C}^2$ at $u^i \in U^i$ such that $\nabla^2_{u^i} l^i < 0$ uniformly. Then, for each $(t,x,p) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^{1\times n}$, there exists a function $\mu^i : [0,T] \times \mathbb{R}^n \times \mathbb{R}^{1\times n} \to U^i$ satisfying

$$\mu^i(t,x,p) \in \arg\max_{u^i \in U^i} (p f^i(t,x) u^i + l^i(t,x,u^i)), \tag{15}$$

such that μ^i is continuous at (t,x,p) and Lipschitz continuous at (x,p). In addition, for each $i \in \mathcal{N}$,

$$u^{i\dagger}(t,x) = \mu^i(t,x,h^{ix}(t,x)), \tag{16}$$

is admissible ($u^\dagger \in \Gamma$) and implementable ($u^\dagger$ is a Nash equilibrium).

Proof: The continuity of μ^i at (t,x,p) follows from the uniqueness of the maximum (15) on the compact set U^i. The Lipschitz continuity at (x,p) is shown by [10, Lemma VI.6.3], since $pf^i(t,x)u^i + l^i(t,x,u^i)$ is Lipschitz continuous at (x,p) and l^i is of class $\mathcal{C}^2$ at u^i. Thus, by noting that h^{ix} is continuous at t and Lipschitz continuous at x, we see that $u^\dagger$ defined by (16) is continuous at t and Lipschitz continuous at x, so that $u^\dagger$ is admissible. The second statement ($u^\dagger$ is a Nash equilibrium) is straightforward from Theorem 1. This completes the proof. ■

We here discuss distinctive features of our approach, compared with the so-called FOC approach [11]–[16]. In our framework and approach, as we remarked just above Theorem 1, a Nash equilibrium $u^* \in \Gamma$,

[6] Theorem 1 does not require the uniqueness of Nash equilibria.

which may be constructed by (14) in Theorem 1, should be given dependently on a design parameter $h \in \mathcal{H} \times \ldots \times \mathcal{H}$ as $u^*(h)$. On the other hand, if we adopt the FOC approach in the same framework as above, we should determine the parameter $h \in \mathcal{H} \times \ldots \times \mathcal{H}$ as a function $h(u^*)$ of a Nash equilibrium $u^* \in \Gamma$ by applying the FOC to the Nash optimality of (14) and regard $u^* \in \Gamma$ as a design parameter. The construction of the parameter $h(u^*)$ in the FOC approach requires a rather strict condition, e.g., the range of $\nabla_{u^i} f$ is equal to that of the state x, while the construction of the Nash equilibrium $u^*(h)$ requires just a solvability condition for the static game (14) (see, e.g., Corollary 1).

For example, let us now consider a well-known physical dynamic system, a mass-spring system with a mass 1, the spring constant 1 and the applied force $u^1 \in U^1 \subset \mathbb{R}$ as shown in Fig. 1. Then, we have $\ddot{y} + y = u^1$, where $y \in \mathbb{R}$ is the displacement of the mass. Using $x = (y, \dot{y})'$, we also have

$$f(\tau, x, u^1) = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x}_{=f^0(\tau,x)} + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix} u^1}_{=f^1(\tau,x)u^1} . \tag{17}$$

Since $\nabla_{u^1} f(t, x, u^1) = f^1(t, x) = (0, 1)'$, u^1 cannot directly operate y and the range of $\nabla_{u^1} f(t, x, u^1)$ is smaller than that of the state $x \in \mathbb{R}^2$. Hence, the FOC approach cannot be applicable to even the mass-spring system stated above.

We see from Theorem 1 that, given w^i (12) based on h^i, each agent $i \in \mathcal{N}$ chooses an implementable control policy $u^{i\dagger} \in \Gamma^i$ (14). To express the parameters used in $u^{i\dagger}$ clearly, let us denote by a function $\mu^i : [0, T] \times \mathbb{R}^n \times \mathcal{H}^x \to U^i$ the alternative representation of $u^{i\dagger}$ satisfying (14), i.e., (16).

We finally consider the selection of $h^i = (h^{i0}, h^{ix})$, $i \in \mathcal{N}$. From the above discussions, we can obtain Theorem 2.

Theorem 2: Assume that the same conditions in Corollary 1 hold. Suppose that, for any $h = (h^1, \ldots, h^N) \in \mathcal{H} \times \ldots \times \mathcal{H}$, there is a Nash equilibrium $u^\dagger = (u^{1\dagger}, \ldots, u^{N\dagger}) \in \Gamma = \Gamma^1 \times \ldots \times \Gamma^N$ satisfying the implementability condition (14). Then, the incentive contract problem (6) is equivalent to the optimal control problem (18) under the salary functional led by (10):

$$\sup_{h^i \in \mathcal{H}, i \in \mathcal{N}} J_0(t, X(t, x, u^\dagger); w(u^\dagger, h)) \tag{18a}$$

$$\text{subject to} \quad u^{i\dagger}(\tau, \xi) = \mu^i(\tau, \xi, h^{ix}(\tau, \xi)), \quad (\tau, \xi) \in [t, T] \times \mathbb{R}^n, \ i \in \mathcal{N}, \tag{18b}$$

$$h^{i0}(t, x) \geq k^i(t, x), \qquad i \in \mathcal{N}. \tag{18c}$$

Proof: For any $h^i = (h^{i0}, h^{ix}) \in \mathcal{H}$, $i \in \mathcal{N}$, the principal selects the salary functional (2) with the salary parameter (12) described by $u^{i*}(t, x) = \mu^i(t, x, h^{ix}(t, x))$, $i \in \mathcal{N}$. Since $\mu^i(\tau, x, h^{ix}_\tau(x))$, $i \in \mathcal{N}$, is implementable from Theorem 1, Lemma 1 is obviously satisfied. This completes the proof. ■

Theorem 2 means that the incentive contract problem (6) can be reduced to the control/incentive synthesis problem with respect to design parameters h from the systems and control perspective. We see from (18) that h^{i0} and h^{ix} can be designed separately for satisfying (6c) and (6b), respectively.

Let us first focus on $h^{i0}(t,x)$. From Theorems 1 and 2, given the initial time and states (t,x), h^{i0} is independent of the selection of u^i and can be set arbitrarily from $\mathcal{H}^0$ under the individual rationality constraint (18c). Actually, given $(t,x) \in [0,T] \times \mathbb{R}^n$, we normally set $h^{i0}(t,x)$ such that $h^{i0}(t,x) = k^i(t,x)$ due to the principal's profit maximization (18a) [11].

We next focus on the selection of h^{ix}, $i \in \mathcal{N}$. The design parameter h^{ix} is determined by the principal's optimization problem (18a) in the presence of the admissible and implementable control policy $u^\dagger$ defined by (18b). From Theorems 1 and 2, once the principal determines an appropriate h^{ix}, $i \in \mathcal{N}$, the optimal u and the corresponding w are automatically fixed from (16) and (12). In other words, by selecting a design parameter h^{ix}, the principal can operate a Nash equilibrium led by agents to some extent.

To find a design parameter h^{ix} in the principal's optimization problem (18a), we first fix the agents' decision functions μ^i, $i \in \mathcal{N}$, defined by (15), and solve the stochastic optimal control problem (18a) with respect to h^{ix}, $i \in \mathcal{N}$, by using, e.g., the HJB equation approach (see [10, Chapter VI]). The existence of μ^i is guaranteed from Corollary 1, but it is generally difficult to acquire the analytical solution of (16) except for limited problem formulation, e.g. the case of cost functions with a quadratic-form on controls in $U^i = \mathbb{R}^{m_i}$, and the framework reduced to so-called bang-bang controls. In the next section, we will consider a special case to solve h^{ix} analytically. The existence of the analytical solution confirms the advantage of our contract problem with the *strong solution framework* dealing with the classical Markovian control problem.

V. Conclusion

This paper have investigated an optimal dynamic incentive contracts between a risk-averse principal and multiple risk-averse agents in continuous-time controlled Markov processes, which include various infrastructure systems and cyber-physical systems. We have mainly proposed a novel control/incentive synthesis problem under moral hazard. Our approach is based on the principal of optimality and the solution of HJB equations. Thanks to the techniques, we have revealed that the proposed dynamic contract incentivizes the agents to take a suitable control composed of a specific Nash equilibrium desired by the principal.

One of the future works is to analyze the qualitative/quantitative property of the design parameter h^{ix} and investigate the trade-off between incentive and risk-sharing in the agency relationship. Another topic is to apply the proposed methodology to the power network systems with fast-regulation electricity markets, whose technical issues are shown in [27].

REFERENCES

[1] M. Amin, A.M. Annaswamy, C.L. DeMarco and T. Samad, *IEEE vision for smart grid controls: 2030 and beyond*, IEEE Press, 2013.

[2] F. Lamnabhi-Lagarrigue et al., "Systems & control for the future of humanity, research agenda: Current and future roles, impact and grand challenges," *Annual Reviews in Control*, vol. 43, pp. 1–64, 2017.

[3] S. Hara, J. Imura, K. Tsumura, T. Ishizaki and T. Sadamoto, "Glocal (global/local) control synthesis for hierarchical networked systems," *Proc. 2015 IEEE Multi-conference on Systems and Control*, pp. 107–112, 2015.

[4] A. Kiani, A. Annaswamy and T. Samad, "A hierachical transactive control architecture for renewables integration in smart grids: Analytical modeling and stability," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 2054–2065, 2014.

[5] Y. Wasa, K. Sakata, K. Hirata and K. Uchida, "Differential game-based load frequency control for power networks and its integration with electricity market mechanisms," *Proc. 1st IEEE Conference on Control Technology and Applications*, pp. 1044–1049, 2017.

[6] T. Hatanaka, X. Zhang, W. Shi, M. Zhu and N. Li, "Physics-integrated hierarchical/distributed HVAC optimization for multiple buildings with robustness against time delays," *Proc. of 56th IEEE Conference on Decision and Control*, pp. 6573–6579, 2017.

[7] D. Fudenberg and J. Tirole, *Game Theory*, The MIT Press, 1991.

[8] P. Bolton and M. Dewatripont, *Contract Theory*, The MIT Press, 2005.

[9] W.H. Fleming and H.M. Soner, *Controlled markov processes and viscosity solutions*, 2nd ed., Springer, 2006.

[10] W.H. Fleming and R.W. Rishel, *Deterministic and stochastic optimal control*, Springer-Verlag, 1975.

[11] B. Holmstrom and P. Milgrom, "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica: J. the Econometric Society*, 55(2):303–328, 1987.

[12] H. Schattler and J. Sung, "The first-order approach to the continuous-time principal-agent problem with exponential utility," *J. Economic Theory*, 61:331–371, 1993.

[13] H. Schattler and J. Sung, "On optimal sharing rules in discrete- and continuous-time principal-agent problems with exponential utility," *J. Economic Dynamics and Control*, 21:551–574, 1997.

[14] H.K. Koo, G. Shim and J. Sung, "Optimal multi-agent performance measures for team contracts," *Mathematical Finance*, 18(4):649–667, 2008.

[15] J. Cvitanic and J. Zhang, *Contract Theory in Continuous-Time Models*, Springer-Verlag, 2013.

[16] N. Williams, "A solvable continuous time dynamic principal-agent model," *J. Economic Theory*, vol. 159, pp. 989–1015, 2015.

[17] R. Elie and D. Possamaï, "Contracting theory with competitive interacting agents," Available via arXiv:1605.08099, 2016.

[18] Y. Sannikov, "A continuous-time version of the principal-agent problem," *Review of Economic Studies*, vol. 75, no. 3, pp. 957–984, 2008.

[19] J. Cvitanić, D. Possamaï and N. Touzi, "Dynamic programming approach to principal-agent problems," *Finance and Stochastics*, vol. 22, no. 1, pp. 1–37, 2018.

[20] M.H.A. Davis and P. Varaiya, "Dynamic programming conditions for partially observable stochastic systems," *SIAM J. Control*, vol. 11, no. 2, pp. 226–261, 1973.

[21] M.H.A. Davis, “On the existence of optimal policies in stochastic control,” *SIAM J. Control*, vol. 11, no. 4, pp. 587–594, 1973.

[22] R.J. Elliott, “The optimal control of a stochastic system,” *SIAM J. Control and Optimization*, vol. 15, no. 5, pp. 756–778, 1977.

[23] J. Chen and Q. Zhu, “A differential game approach to dynamic contract design for systemic cyber risk management under asymmetric information,” *Proc. 56th Annual Allerton Conference on Communication, Control, and Computing*, pp. 575–582, 2018.

[24] P. Venkitasubramaniam and V. Gupta, “Data-driven contract design,” *Proc. 2019 American Control Conference*, pp. 2283–2288, 2019.

[25] Y. Wasa, K. Hirata and K. Uchida, “Optimal agency contract for incentive and control under moral hazard in dynamic electric power networks,” *IET Smart Grid*, vol. 2, no. 4, pp. 594–601, 2019.

[26] T. Basar, “Nash equilibrium of risk-sensitive nonlinear stochastic differential games,” *J. Optimization Theory and Applications*, 100(3):479–498, 1999.

[27] K. Uchida, K. Hirata and Y. Wasa, “Incentivizing market and control for ancillary services in dynamic power grids,” *In Smart Gird Control: Overview and Research Opportunities*, J. Stoustrup, A.M. Annaswamy, A. Chakrabortty and Z. Qu (eds.), Springer-Verlag, pp. 47–58, 2018.

[28] Y. Wasa, K. Hirata and K. Uchida, “Contract theory approach to incentivizing market and control design,” Available via arXiv:1709.09318, 2018.

Appendix

A. Proof of Lemma 1

To derive the HJB equation for a Nash equilibrium $u^* \in \Gamma$, let us convert the Bolza-form original problem to the Mayer form. As the local transient cost function l^i of agent $i \in \mathcal{N}$ and his salary functional W^i satisfying the same regularity conditions as (1), for a Nash equilibrium $u^* \in \Gamma$ and $u^i \in \Gamma^i$, we can introduce an extra state variable $x^e \in \mathbb{R}$ satisfying

$$\begin{aligned}\begin{bmatrix} dx_\tau \\ dx_\tau^e \end{bmatrix} &= \begin{bmatrix} f_\tau(x_\tau, u_\tau^i, u_\tau^{-i*})d\tau + \sigma_\tau(x_\tau)d\beta_\tau \\ l_\tau^i(x_\tau, u_\tau^i)d\tau + w_\tau^{it}(x_\tau)d\tau + w_\tau^{ix}(x_\tau)dx_\tau \end{bmatrix} \\ &= \begin{bmatrix} f_\tau \\ l_\tau^i + w_\tau^{it} + w_\tau^{ix} f_\tau \end{bmatrix} d\tau + \begin{bmatrix} \sigma_\tau \\ w_\tau^{ix}\sigma_\tau \end{bmatrix} d\beta_\tau. \end{aligned} \tag{19}$$

Then, we introduce a value function $\tilde{V}^i : [0,T] \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ given by

$$\tilde{V}^i(t,x,x^e) := -\nu_i(-w_t^{i0}(x)) \max_{u^i \in \Gamma^i} \mathbb{E}_{t,x,x^e}\left[\nu_i(x^e(T))\right], \tag{20}$$

where $x^e(T)$ means the state at time T along (19) during $\tau \in [t,T]$ with $x(t) = x$ and $x^e(t) = x^e$. By using Ito’s differential rule, we can derive the following HJB equation on $\tilde{V}^i$ for all $(\tau, x, x^e) \in$

$[0, T] \times \mathbb{R}^n \times \mathbb{R}$:

$$0 = \nabla_\tau \tilde{V}^i(\tau, x, x^e) + \frac{1}{2}\mathrm{tr}\left\{ \begin{bmatrix} \nabla_x^2 \tilde{V}_\tau^i & \nabla_x \nabla_{x^e} \tilde{V}_\tau^i \\ \nabla_{x^e} \nabla_x \tilde{V}_\tau^i & \nabla_{x^e} \tilde{V}_\tau^i \end{bmatrix} \begin{bmatrix} \sigma_\tau \sigma_\tau' & \sigma_\tau (w_\tau^{ix} \sigma_\tau)' \\ (w_\tau^{ix} \sigma_\tau)\sigma_\tau' & |(w_\tau^{ix} \sigma_\tau)'|^2 \end{bmatrix} \right\}$$

$$+ \max_{u^i \in U^i} \left\{ \begin{bmatrix} \nabla_x \tilde{V}_\tau^i \; \nabla_{x^e} \tilde{V}_\tau^i \end{bmatrix} \begin{bmatrix} f_\tau(x, u^i, u_\tau^{-i*}) \\ l_\tau^i(x, u^i) + w_\tau^{it} + w_\tau^{ix} f_\tau \end{bmatrix} \right\} \tag{21a}$$

$$\tilde{V}^i(T, x_T, x_T^e) = \nu_i(w^{iT}(x_T) + \varphi^i(x_T) + x_T^e) \tag{21b}$$

On the other hand, from (3), (8), (19) and (20), since $\nu_i(a+b) = -\nu_i(a)\nu_i(b)$ for arbitrary $a, b \in \mathbb{R}$ and $x^e(T) = \Psi^i(t, X_t(x, u^i, u^{-i*}); w^i) + x^e$, we have

$$\tilde{V}^i(t, x, x^e) = -\nu_i(x^e) \times V^i(t, x). \tag{22}$$

We substitute (22) for (21a) and obtain

$$\begin{aligned} 0 &= \nabla_\tau V_\tau^i(x)(-\nu_i(x^e)) \\ &+\frac{1}{2}\mathrm{tr}\left\{ \begin{bmatrix} -\nu_i(x^e)\nabla_x^2 V_\tau^i(x) & r_i \cdot \nu_i(x^e) \nabla_x V_\tau^i(x) \\ r_i \cdot \nu_i(x^e) \nabla_x V_\tau^i(x) & r_i^2(-\nu_i(x^e)) V_\tau^i(x) \end{bmatrix} \begin{bmatrix} \sigma_\tau(x)\sigma_\tau(x)' & \sigma_\tau(x)(w_\tau^{ix}(x)\sigma_\tau(x))' \\ (w_\tau^{ix}(x)\sigma_\tau(x))\sigma_\tau(x)' & |(w_\tau^{ix}(x)\sigma_\tau(x))'|^2 \end{bmatrix} \right\} \\ &+ \max_{u^i \in U^i} \left\{ \begin{bmatrix} -\nu_i(x^e)\nabla_x V_\tau^i(x) & r_i \cdot \nu_i(x^e) V_\tau^i(x) \end{bmatrix} \begin{bmatrix} f_\tau(x, u^i, u_\tau^{-i*}) \\ l_\tau^i(x, u^i) + w_\tau^{it}(x) + w_\tau^{ix}(x) f_\tau(x, u^i, u_\tau^{-i*}) \end{bmatrix} \right\} \end{aligned} \tag{23}$$

We here divide both sides of (23) by $-\nu_i(x^e) > 0$ and derive (9a) by using the standard properties of the trace. From (2), (3), (4) (21b) and (22), the boundary condition (9b) on V^i is easily obtained. This completes the proof.

B. Proof of Lemma 2

Following to [11, Theorem 6] and [12, Theorem 4.1], let us consider the salary condition based on the HJB equation (9) for V^i. To derive the representation for the sharing rule, we now define the agent's certainty equivalent $\hat{V}^i$ as

$$\hat{V}^i(t, x) := \nu_i^{-1}(V^i(t, x)) = (-r_i)^{-1} \log_e(-V^i(t, x)). \tag{24}$$

By using Ito's differential rule and (9a), the following equation holds for any $\tau \in [t, T]$:

$$d\hat{V}_\tau^i(x) = \frac{dV_\tau^i(x)}{(-r_i)V_\tau^i(x)} - \frac{1}{2}\frac{dV_\tau^i(x) dV_\tau^i(x)}{(-r_i)(V_\tau^i(x))^2}, \;\; x \in \mathbb{R}^n, \tag{25}$$

where

$$dV_\tau^i = \left(\nabla_\tau V_\tau^i + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V_\tau^i \sigma_\tau \sigma_\tau' \right] \right) d\tau + \nabla_x V_\tau^i dx_\tau. \tag{26}$$

Then, from (9a) and (26),

$$\frac{dV_\tau^i}{(-r_i)V_\tau^i} = \frac{\nabla_x V_\tau^i}{(-r_i)V_\tau^i}dx_\tau$$
$$+\left(\frac{r_i}{2}|(w_\tau^{ix}\sigma_\tau)'|^2 - \left(\frac{\nabla_x V_\tau^i}{(-r_i)V_\tau^i} + w_\tau^{ix}\right) f_\tau(u_\tau^*) - l_\tau^i(u_\tau^{i*}) - w_\tau^{it} + r_i \frac{\nabla_x V_\tau^i}{(-r_i)V_\tau^i}\sigma_\tau(w_\tau^{ix}\sigma_\tau)'\right) d\tau, \tag{27a}$$
$$-\frac{1}{2}\frac{dV_\tau^i dV_\tau^i}{(-r_i)(V_\tau^i)^2} = -\frac{1}{2}(-r_i)\frac{|(\nabla_x V_\tau^i \sigma_\tau)'|^2}{((-r_i)V_\tau^i)^2}d\tau. \tag{27b}$$

We substitute (27) for (25) and derive

$$d\hat{V}_\tau^i = \left(-\frac{1}{2}(-r_i)|(h_\tau^{ix}\sigma_\tau)'|^2 - h_\tau^{ix} f_\tau - l_\tau^i\right) d\tau + h_\tau^{ix} dx_\tau - (w_\tau^{it} d\tau + w_\tau^{ix} dx_\tau) \tag{28}$$

where h^{ix} is defined as (11b). From (9b), (11a), and (24),

$$\hat{V}^i(T, x_T) = w^{iT}(x_T) + \varphi^i(x_T), \tag{29a}$$
$$\hat{V}^i(T, x_T) = \hat{V}^i(t, x) + \int_t^T d\hat{V}_\tau^i = h^{i0}(t, x) - w^{i0}(t, x) + \int_t^T d\hat{V}_\tau^i. \tag{29b}$$

Substituting (28) into (29b), we deform (29) and obtain (12) by using (2). The proof is completed.

C. *Proof of Theorem 1*

(Necessary Condition) We suppose that $u^\dagger \in \Gamma$ is implementable (Nash equilibrium). u^* used in Lemmas 1, 2 and 3 is replaced by $u^\dagger$ from now on. Then, substitute the necessary condition (12) for the replaced HJB equation (9) and obtain

$$0 = \nabla_\tau V_\tau^i(x) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V_\tau^i(x)\sigma_\tau(x)\sigma_\tau(x)'\right]$$
$$+\max_{u^i\in U^i}\Big[\nabla_x V_\tau^i(x) f_\tau(x, u^i, u_\tau^{-i\dagger}) - \nabla_x V_\tau^i(x) r_i \sigma_\tau(x)(h_\tau^{ix}(x)\sigma_\tau(x))')$$
$$-r_i V_\tau^i(x)\left[h_\tau^{ix}(x) f_\tau(x, u^i, u_\tau^{-i\dagger}) + l_\tau^i(x, u^i)\right] + r_i V_\tau^i(x)\left[h_\tau^{ix}(x) f_\tau(x, u_\tau^\dagger) + l_\tau^i(x, u_\tau^{i\dagger})\right]\Big]$$
$$= \nabla_\tau V_\tau^i(x) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V_\tau^i(x)\sigma_\tau(x)\sigma_\tau(x)'\right] + \nabla_x V_\tau^i(x)\left[f_\tau(x, u_\tau^\dagger) - r_i\sigma_\tau(x)(h_\tau^{ix}(x)\sigma_\tau(x))')\right], \tag{30a}$$
$$V^i(T, x_T) = \nu_i(0) = -1. \tag{30b}$$

As we see from Lemma 3 that $V_t^i(x) = \nu_i(0) = -1$ and $\nabla_x V_t^i(x) = 0$ for all $(t, x) \in [0, T] \times \mathbb{R}^n$, (30a) can be replaced by

$$0 = \nabla_\tau V_\tau^i(x) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V_\tau^i(x)\sigma_\tau(x)\sigma_\tau(x)'\right]$$
$$+\max_{u^i\in U^i}\left[(-r_i)V_\tau^i(x)\left[h_\tau^{ix}(x) f_\tau(x, u^i, u_\tau^{-i\dagger}) + l_\tau^i(x, u^i)\right] - (-r_i)V_\tau^i(x)\left[h_\tau^{ix}(x) f_\tau(x, u_\tau^\dagger) + l_\tau^i(x, u_\tau^{i\dagger})\right]\right]$$
$$= \nabla_\tau V_\tau^i(x) + \frac{1}{2}\mathrm{tr}\left[\nabla_x^2 V_\tau^i(x)\sigma_\tau(x)\sigma_\tau(x)'\right]. \tag{31}$$

Therefore, as $-r_i V_\tau^i(x) > 0$ holds, we can obtain

$$\max_{u^i \in U^i} \left[h_\tau^{ix}(x) f_\tau(x, u^i, u_\tau^{-i\dagger}) + l_\tau^i(x, u^i) \right] = h_\tau^{ix}(x) f_\tau(x, u_\tau^{i\dagger}, u_\tau^{-i\dagger}) + l_\tau^i(x, u_\tau^{i\dagger}), \tag{32}$$

which is just (14).

(Sufficient Condition) For $u^\dagger \in \Gamma$ and $h^i \in \mathcal{H}$, we consider the reward functional Ψ^i with the salary functional $W^i(t, X_t(x, u^i, u^{-i\dagger}); u^\dagger, h^i)$ composed of (2) and (12). Then,

$$\begin{aligned}
&\Psi^i(t, X_t(x, u^i, u^{-i\dagger}); u^\dagger, h^i) \\
&= W_t^i(X_t(x, u^i, u^{-i\dagger}); u^\dagger, h^i) + \varphi^i(x_T) + \int_t^T l_\tau^i(x_\tau, u_\tau^i) d\tau \\
&= h_t^{i0}(x) + \int_t^T l_\tau^i(x_\tau, u_\tau^i) d\tau \\
&\quad - \int_t^T \left[\frac{-r_i}{2} |(h_\tau^{ix} \sigma_\tau)'|^2 + h_\tau^{ix} f_t(x_\tau, u_\tau^\dagger) + l_\tau^i(x_\tau, u_\tau^{i\dagger}) \right] d\tau + \int_t^T h_\tau^{ix} (f_\tau(x_\tau, u_\tau^i, u_\tau^{-i\dagger}) d\tau + \sigma_\tau d\beta_\tau) \\
&= h_t^{i0}(x) - \frac{-r_i}{2} \int_t^T |(h_\tau^{ix} \sigma_\tau)'|^2 d\tau + \int_t^T h_\tau^{ix} \sigma_\tau d\beta_\tau \\
&\quad + \int_t^T [h_\tau^{ix} f_\tau(x_\tau, u_\tau^i, u_\tau^{-i\dagger}) + l_\tau^i(x_\tau, u_\tau^i))] d\tau - \int_t^T [h_\tau^{ix} f_\tau(x_\tau, u_\tau^{i\dagger}, u_\tau^{-i\dagger}) + l_\tau^i(x_\tau, u_\tau^{i\dagger})] d\tau.
\end{aligned} \tag{33}$$

Actually, as the profit functional of agent i is given by (4), we substitute (33) for (4) and derive

$$\begin{aligned}
&J_i(t, X_t(x, u^i, u^{-i\dagger}); u^\dagger, h^i) = \nu_i(h_t^{i0}(x)) \mathbb{E}_{t,x} \left[\exp\left[-\frac{1}{2} \int_t^T |(-r_i h_\tau^{ix} \sigma_\tau)'|^2 d\tau + \int_t^T -r_i h_\tau^{ix} \sigma_\tau d\beta_\tau \right] \right] \\
&\times \exp\left[(-r_i) \int_t^T [h_\tau^{ix} f_\tau(x_\tau, u_\tau^i, u_\tau^{-i\dagger}) + l_\tau^i(x_\tau, u_\tau^i)] d\tau - (-r_i) \int_t^T [h_\tau^{ix} f_\tau(x_\tau, u_\tau^{i\dagger}, u_\tau^{-i\dagger}) + l_\tau^i(x_\tau, u_\tau^{i\dagger})] d\tau \right].
\end{aligned} \tag{34}$$

Now, we suppose that $u^\dagger \in \Gamma$ satisfies (14). Then, from (34),

$$J_i(t, X_t(x, u^i, u^{-i\dagger}); u^\dagger, h^i) \le \nu_i(h_t^{i0}(x)) \mathbb{E}_{t,x} \left[\exp\left[-\frac{1}{2} \int_t^T |(-r_i h_\tau^{ix} \sigma_\tau)'|^2 d\tau + \int_t^T -r_i h_\tau^{ix} \sigma_\tau d\beta_\tau \right] \right]. \tag{35}$$

Since $h_\tau^{ix}(x) \sigma_\tau(x)$ is bounded on $\mathbb{R}^n$, the expectation term in (35) always takes 1. Hence,

$$J_i(t, X_t(x, u^i, u^{-i\dagger}); u^\dagger, h^i) \le \nu_i(h_t^{i0}(x)) = J_i(t, X_t(x, u^{i\dagger}, u^{-i\dagger}); u^\dagger, h^i) \tag{36}$$

is obtained and $u^\dagger$ is implementable (a Nash equilibrium).

From the above results, it is obvious that V^i takes the value shown in Lemma 3. This completes the proof.