УДК 519.21;519.714.3;681.518

DOI: http://doi.org/10.25728/pu.2020.1.4

СОГЛАСОВАННОЕ УПРАВЛЕНИЕ МНОГОЭЛЕМЕНТНЫМИ ДИНАМИЧЕСКИМИ ОРГАНИЗАЦИОННЫМИ СИСТЕМАМИ

Ч. 1. Динамическая организационная система в составе одного центра и множества агентов

М.В. Белов

Аннотация. Рассмотрены постановка и вопросы решения задач согласованного управления многоэлементными динамическими организационными/активными системами (ОС) с ограничениями на совместную деятельность их элементов в виде технологических сетей. Рассмотрена ОС, состоящая из одного центра и множества подчиненных ему агентов. Доказаны утверждения о том, что для любой допустимой траектории результатов может быть построена согласованная компенсаторная система стимулирования, которая реализует (как равновесие в доминантных стратегиях) траекторию действий агентов, приводящих к требуемой траектории результатов; декомпозирует задачу управления по агентам и по периодам времени; гарантированно обеспечивает (по всем возможным дальновидностям агентов) минимальные затраты управляющего органа (центра) на реализацию данной траектории результатов. Показано, что в таких системах стимулирования размеры платежей зависят только от соответствующих значений функций затрат, которые, в свою очередь, косвенно учитывают технологические функции, структуру сетей и структуру ОС в целом. Поставлена задача оптимального планирования и указан алгоритм ее решения.

Ключевые слова: стимулирование, многоуровневые динамические активные системы, согласованное управление.

ВВЕДЕНИЕ

Принцип согласованного управления представляет собой один из основных в *теории активных систем* [1—4] и в *теории управления организационными системами* [5], причем для случая полной информированности в активных системах принцип был впервые сформулирован в работах [1, 2], где были получены основополагающие результаты.

Рассмотрим объект управления — иерархию взаимосвязанных *агентов*, которые образуют многоуровневую *организационную систему* (ОС), подсистемы которой также представляют собой ОС, каждая из которых — сложная *сетевая структура* [6—9].

В книге [5] приведен *принцип декомпозиции игры агентов*, в рамках которого управляющий орган — *центр* — всегда может (без потери эффективности) в детерминированной двухуровневой многоэлементной статической ОС применять систему стимулирования агентов, при которой выбор требуемого для центра действия составляет доминантную стратегию каждого агента. В настоящей работе эта идея, следуя изложению статьи [10] обобщается на случай динамических и многоуровневых детерминированных ОС. В первой части настоящей статьи рассматривается двухуровневая (центр — агенты) многоэлементная динамическая ОС, в которой ограничение на совместную деятельность агентов задано в форме технологической сети (ориентированного графа без контуров). Для этой ОС строится оптимальная относительно затрат центра компенсаторная (основывающаяся на принципе компенсации затрат агентов [5]) система стимулирования, декомпозирующая игру агентов по агентам и периодам. Фактор динамики рассматривается в смысле многошагового процесса принятия решений, в то время как динамика управляемой организационной системы задается технологией совместной деятельности агентов и специально не исследуется.

1. МОДЕЛЬ ДИНАМИЧЕСКОЙ СЕТЕВОЙ ОРГАНИЗАЦИОННОЙ СИСТЕМЫ

Рассмотрим многоэлементную детерминированную динамическую организационную систему [11] с одним центром и множеством агентов (обозначим через $N = \{1, 2, ..., n\}$ конечное множество агентов, $n \geq 2$). Определим *технологию* совместной деятельности агентов, которую опишем *сетью* $G = (N, E)$ с правильной нумерацией [12] (которая всегда возможна в графе без контуров), и будем считать, что агенты помечены именно этими номерами. Вершины сети соответствуют действиям агентов, а множество дуг $E \subseteq N \times N$ отражает «технологические» связи между ними. Через $N(i) = \{j \in N | (j, i) \in E\}$ обозначим множество *предшественников* i-го агента в сети G, $i \in N$. Считаем, что сеть имеет множество *входов* $M_0 \subseteq N$ и единственный *выход* — n-ю вершину, отличающуюся тем, что конечной (например, рыночной) ценностью обладает только результат деятельности выхода сети — результат динамической сетевой ОС в целом, который приносит центру доход.

Через M_k обозначим множество вершин, в которые входят дуги только из вершин, принадлежащих множествам $\{M_j\}$, $j \in \{1, ..., k - 1\}$; число k назовем *рангом* вершины, принадлежащей множеству M_k.

Рассмотрим функционирование данной ОС в течение $t \in \{1, 2, ..., T\}$ периодов дискретного времени, где T — *горизонт времени*. В периоде t i-й агент выбирает и реализует свое *действие* (целенаправленную активность), которое характеризуется набором параметров $y_i(t) \in Y_i$, где Y_i — множество возможных значений параметров его действий, включающее в себя в качестве одного из допустимых действий отказ от участия в ОС. Действие i-го агента (вместе с результатами его предшествеников) определяет *результат* его деятельности, который, в свою очередь, описывается набором параметров $z_i(t) \in Z_i$, где Z_i — множество возможных значений параметров его результатов. Будем считать, что Y_i и Z_i — подмножества конечномерного евклидова пространства или некоторого конечного множества. В технологических сетях часто множества Y_i и Z_i зависят от выбора действий предшественниками i-го агента, т. е. множества допустимых значений представляют собой функции от результатов его предшественников $Y_i(\{Z_k;\ k \in N(i)\})$ (где $N(i)$ — множество предшественников i-го агента), что подробно рассматривается, например, в статье [8]. Однако в данной постановке, как будет видно далее, наличие или отсутствие этих зависимостей не сказывается на полученных результатах, поэтому не будем указывать эти зависимости явно там, где это не приводит к разночтениям.

В пределах каждого периода агенты выбирают и выполняют действия, не вступая в коалиции, в последовательности соответственно номерам в сети G, соблюдая таким образом технологию совместной деятельности. Обозначим через $y_D(t)$ вектор действий агентов с номерами из множества $D \subseteq N$, через $z_D(t)$ — вектор результатов деятельности агентов с номерами из этого множества, через $y_{-i} = (y_1, ..., y_{i-1}, y_{i+1}, ..., y_n)$ — вектор действий всех агентов, кроме i-го.

Обозначим $\tilde{W}_i^t$ множество возможных значений действий i-го агента и его предшественников в течение периодов $\overline{1, t}$, заметим $\tilde{W}_i^t \subseteq (Y_i)^t \times \prod_{k \in N(i)} (Z_k)^t$. Связь результата деятельности агента с его действием и используемыми им в процессе этой деятельности результатами других агентов определяется «*технологической функцией*» $Q_i^t\colon \tilde{W}_i^t \to Z_i$, т. е. $z_i(t) = Q_i^t(y_i[1{*}t],\ z_{N(i)}[1{*}t])$ (здесь и далее с помощью нотации $\xi[1{*}t]$ обозначим наборы всех значений величины $\xi(\tau)$ (возможно, многомерной) для $\tau = \overline{1, t}$). Для всех $i \in M_0$ имеет место $N(i) = \varnothing$, поэтому положим $z_i(t) = Q_i^t(y_i[1{*}t],\ z_0[1{*}t])$, где z_0 — известный L-мерный вектор *«входов»* сети.

Сетевые структуры ОС исследуются в работах [6—9]; в настоящей статье сеть задает технологические связи между агентами, при этом рассматривается многошаговое выполнение деятельности.

Задачу управления рассмотрим и решим в двух постановках, когда верны предположение П1 или предположение П2.

Предположение П1. Центр наблюдает только значение выхода сети $z_n(t)$, а функции $Q_i^t(\cdot)$ взаимно однозначны относительно действий самого агента и результатов его предшественников в текущем периоде, т. е. при заданных результатах деятельности предшественников агента в текущем периоде $z_{N(i)}(t)$ (и известных $y_i[1{*}(t - 1)]$ и $z_{N(i)}[1{*}(t - 1)]$ в предыдущих периодах) текущее действие агента $y_i(t)$ однозначно определяет результат его деятельности $z_i(t)$, и наоборот — зная действия $y_i[1^*(t-1)]$, результат деятельности агента $z_i(t)$ и результаты его предшественников $z_{N(i)}(t)$, можно однозначно восстановить его действие в текущем периоде.

Предположение П2. Центр в каждом периоде наблюдает фактические действия агентов (в этом периоде).

Обсудим условия, при которых справедливы предположения П1 и П2.

Прежде всего отметим, что предположение П2 достаточно редко выполняется в сложной практической деятельности потому, что наблюдение действий требует от центра временны́х затрат, сравнимых с собственно выполнением действий. Как правило, временно́й ресурс управляющего центра существенно более дорогой, чем соответствующий ресурс управляемых агентов, поэтому контроль в той или иной форме обычно осуществляется на основании результатов. Непосредственное наблюдение действий возможно и оправданно в случае достаточно простых и стандартизованных видов деятельности, допускающих бинарное тестирование — выполнены или нет технологические требования.

Предположение П1 характеризует технологическую функцию совместной деятельности агентов, когда действие каждого из них влияет на результат таким образом, что не может быть замещено каким-либо другим, т. е. технология каждого из агентов уникальна в рамках ОС. Уникальность технологии может проявляться, например, в том, что действие каждого агента может реализовываться только при определенных сочетаниях результатов его предшественников (хотя в общем случае наличие связей между агентами не рассматривается как необходимое условие уникальности технологии каждого из них). Под агентами могут пониматься как отдельные индивиды — сотрудники предприятий, так и целые организации. В качестве содержательных примеров можно привести работу сборочного цеха (или его участка) аэрокосмического предприятия или кораблестроительной верфи, сооружение нефтехимического цеха или изготовление деталей, сборку из них некоторого изделия и его наладку.

Технология именно комплексной деятельности характеризуется важной особенностью: каждый агент в своей деятельности использует результаты предшественников, потому его результат является более сложным объектом, чем результаты предшественников (и нижестоящих подсистем) и потому требует более «богатого» описания. Это объясняет рост размерностей множеств возможных значений результатов агентов по сравнению с результатами предшественников. Принцип усложнения результата противопоставляет данную постановку задачи традиционному подходу к рассмотрению многоуровневых систем [13], предполагающему «агрегирование» результатов предшествующих агентов и нижестоящих подсистем, что не увеличивает сложность (и соответственно размерность вектора характеристик) результата. В случаях, когда действия агентов не усложняют результат, целесообразно пользоваться различными механизмами стимулирования [15], например, «теоремой об идеальном агрегировании».

Практически во всех упомянутых случаях множества значений возможных действий агентов и их результатов целесообразно считать конечными: агенты выполняют «типовые операции», т. е. значения характеристик (возможно, многомерных) их действий и результатов принадлежат конечному числу областей, внутри каждой из которых значения характеристик неразличимы с точки зрения центра.

Рассмотрим i-го агента на промежутке времени $t = \{1, 2, ..., T\}$ и $N(i)$ — множество его предшественников. Согласно определению технологической функции, приведенному выше, результат $z_i(t)$ действий i-го агента (и его предшественников) зависит от кортежа из $t \times (|N(i)| + 1)$ величин $\{y_i(\tau), z_k(\tau)\colon \tau \in \{1, 2, ..., t\}, k \in N(i)\}$, причем зависимость от $|N(i)| + 1$ из них — взаимно-однозначная. В случае, когда агенты выполняют «типовые операции», множества Y_i значений их возможных действий конечны с точки зрения центра, тогда и множества Z_i значений возможных результатов также конечны. В этом случае технологическая функция $Q_i^t(\cdot)$ представляет собой некоторое правило нумерации сочетаний $\{y_i(\cdot),\ z_k(\cdot)\}$, допустимых относительно специфики выполняемых агентами действий.

Будем считать, что выбор действия $y_i(t)$ требует от i-го агента затрат, зависящих в общем случае от действий всех агентов в ОС, $c_i^t(y_N[1^*t])$ — неотрицательные *функции затрат*, $c_i^t(\cdot)\colon \prod_{k \in N} (Y_k)^t \to \Re_+^1$. Предложенный вид зависимости затрат от предыстории действий агентов также характерен для сложной практической деятельности, примеры которой рассмотрены выше, в частности, (трудо)затраты группы слесарей-сборщиков на линии конечной сборки изделия авиакосмической фирмы зависят от предыстории их действий и действий их смежников[1].

[1] В большинстве традиционных постановок аналогичных задач затраты агента зависят от действий не всех агентов, а только предшественников. Однако зависимость «от всех» — более общая и не сужает получаемые результаты.

Тогда *технология деятельности* i-го агента описывается кортежем $\langle N(i),\ c_i^t(\cdot),\ Q_i^t(\cdot),\ t \in \{1, 2, ..., T\}\rangle$.

Сеть G и технологические функции $Q_i^t(\cdot)$ для каждой траектории действий $y_N[1{*}t]$ задают единственную траекторию выхода сети, будем обозначать эту траекторию через $z_n^t(y_N[1{*}t])$.

Если верно предположение П1, центр может применять $\sigma_i^t(\cdot)$ — неотрицательные функции *стимулирования* i-го агента в периоде t, в качестве аргументов которых выступают значения выхода сети $\sigma_i^t(z_n[1{*}t])$: $(Z_n)^t \to \Re^1_+$. В случае справедливости предположения П2 центр может применять неотрицательные функции стимулирования, в качестве аргументов которых выступают наблюдаемые им действия агентов: $\tilde{\sigma}_i^t(y_N[1{*}t])$: $\prod\limits_{k \in N}(Y_k)^t \to \Re^1_+$.

Для того чтобы описать игру агентов, необходимо доопределить порядок функционирования и информированность — кто из них, когда и какие решения (из каких множеств) принимает, и что каждый знает на момент принятия решений. Будем считать, что центр, зная сеть, технологические функции и функции затрат агентов, сначала сообщает им функции стимулирования $\{\sigma_i^\tau\}$, $\tau = \overline{1, T}$, на все последующие периоды времени (центр тем самым осуществляет «*программное управление*»), не имея возможности менять стимулирование в дальнейшем. Сеть, множества допустимых действий, дальновидности, технологические функции и функции затрат всех агентов в каждом периоде представляют собой общее знание для них и центра [5]. Зная текущую историю игры (совокупность выбранных в предыдущих периодах действий всех агентов и результатов их деятельности) и функции стимулирования всех агентов, последние выбирают свои действия в этом периоде однократно, не вступая в коалиции, в последовательности соответственно номерам в сети G, соблюдая таким образом технологию совместной деятельности. Пусть агенты дальновидны. Выбирая действие $y_i(t)$ и прогнозируя свои действия $y_i(\tau)$, $\tau \in \{t + 1, ..., T\}$, i-й агент стремится максимизировать свою целевую функцию

$$F_i(\{\sigma_i^\tau\},\ \tau = \overline{t, T};\ y_N[1{*}T]) = \sum_{\tau = t}^{T} \delta_i(t, \tau) f_i^\tau(\sigma_i^\tau,\ y_N[1{*}\tau]), \qquad (1)$$

где $\delta_i(\cdot) \geq 0$ — распределение дальновидностей агента, $f_i^t(\cdot)$ — целевая функция агента в периоде t, равная разности между вознаграждением и затратами (получение агентом резервной полезности при отказе от участия в ОС учтено в функциях стимулирования $\{\sigma_i^\tau\}$ — см. далее соотношения (4) и (7)):

$$f_i^t(\sigma_i^t,\ y_N[1{*}t]) = -c_i^t(y_N[1{*}t]) + \begin{cases} \sigma_i^t(z_n(y_N[1{*}t])), & \text{если верно П1}, \\ \tilde{\sigma}_i^t(y_N[1{*}t]), & \text{если верно П2}, \end{cases}$$
$$i \in N, \quad t = \overline{1, T}.$$

Описанную модель — совокупность множества агентов N (с их интересами и предпочтениями), сети G и набора $\{Q_i^t(\cdot)\}$ технологических функций — назовем *динамической сетевой ОС* (ДСОС), которая является динамическим обобщением как модели сетевой активной системы, так и модели «производственных цепочек» на случай сетевой структуры [6].

В данной повторяющейся игре агентов *доминантной стратегией* i-го агента (при системе стимулирования $\{\sigma_N^\tau\}$, $\tau = \overline{1, T}$) назовем $y_i^{д}[1{*}T]$ траекторию его действий такую, что

$$\forall y_{-i}[1{*}T] \in \prod_{k \in N;\ k \neq i}(Y_k)^T;\ \ \forall t_1 \in \overline{1, T};\ \ \forall t_2 \in \overline{t_1, T};$$
$$\forall s[t_1{*}t_2] \in (Y_i)^{t_2 - t_1 + 1};\quad s[t_1{*}t_2] \neq y_i^{д}[t_1{*}t_2]:$$
$$F_i(\{\sigma_N^\tau\};\ \{y_i^{д}[1{*}T],\ y_{-i}[1{*}T]\}) \geq F_i(\{\sigma_N^\tau\};$$
$$\{y_i^{д}[1{*}(t_1 - 1)],\ s[t_1{*}t_2],\ y_i^{д}[(t_2 + 1){*}T],\ y_{-i}[1{*}T]\}).$$

Доминантная стратегия $y_i^{д}[1{*}T]$ обеспечивает максимум целевой функции i-го агента независимо от выбора других агентов $y_{-i}[1{*}T]$.

Равновесием в доминантных стратегиях (РДС) называется совокупность доминантных стратегий всех агентов: $y_N^{д}[1{*}T] = \{y_1^{д}[1{*}T],\ ...,\ y_i^{д}[1{*}T],\ ...,\ y_n^{д}[1{*}T]\}$.

Сформулируем *задачу стимулирования в ДСОС*. Обозначим через σ_D^t вектор-функцию стимулирования агентов из множества $D \subseteq N$ в периоде времени $t = \overline{1, T}$. Рассмотрим повторяющуюся игру агентов в течение периодов от 1 до T. Обозначим через $E(\{\sigma_D^\tau\},\ \tau = \overline{1, T})$ множество траекторий действий $y_N[1{*}T]$ — множество решений игры агентов (равновесий в любом смысле, например, совершенных по подыграм равновесий), реализуемых системой стимулирования $\{\sigma_D^\tau\}$.

Предположим, что центр полностью *дальновиден* (степень его дальновидности равна T) [11]. Тогда *целевая функция ДСОС* — критерий эффективности управления (стимулирования) ДСОС — представляет собой разность между доходом центра и его суммарными затратами на стимулирование всех агентов:

$$\Phi(\{\sigma_N^\tau\}, \tau = \overline{1, T}) =$$
$$= \max_{y_N[1^*T] \in E(\{\sigma_N^t\},\ t = \overline{1, T})} \left\{ \sum_{t=1}^{T} \delta(1, t)\phi(z_n^t(y_N[1^*t])) \right\}, \quad (2)$$

где $\delta(\cdot) \ge 0$ — распределение дальновидностей центра, а функция $\phi(\cdot)$ имеет вид $h^t(z_n^t) \ge 0$ — доход центра в периоде t:

$$\phi(z_i^t(y_N[1^*t])) = h^t(z_n^t(y_N[1^*t])) -$$
$$- \begin{cases} \sum_{i=1}^{n} \sigma_i^t(z_n^t(y_N[1^*t])), \text{если верно П1}, \\ \sum_{i=1}^{n} \tilde{\sigma}_i^t(y_N[1^*t]), \text{если верно П2}. \end{cases}$$

Вычисление максимума в формуле (2) по множеству равновесий $E(\cdot)$ соответствует гипотезе благожелательности [5]: из множества равновесий $E(\cdot)$ агенты выберут наиболее выгодное для центра. Если гипотеза благожелательности не соблюдается, следует пользоваться принципом гарантированного результата и брать минимум среди траекторий, принадлежащих множеству равновесий $E(\cdot)$.

Задача управления заключается в построении системы стимулирования $\{\sigma_N^t\}$, имеющей максимальную эффективность, т. е. обеспечивающей максимум целевой функции центра $\Phi(\cdot)$:

$$\Phi(\{\sigma_N^t\}) \to \max_{\sigma_N^t}. \quad (3)$$

Значение функционала $\Phi(\cdot)$, получаемое в результате решения задачи (3), назовем *эффективностью ДСОС.*

Задача (1)—(3) не тривиальна — «прямой» поиск ее решения не представляется возможным. Поэтому воспользуемся свойствами ДСОС, позволяющими применить общие теоремы о декомпозиции игры агентов [5, 6].

2. ДЕКОМПОЗИЦИЯ ИГРЫ АГЕНТОВ В ДИНАМИЧЕСКОЙ СЕТЕВОЙ ОС

Исследования реализуемых в повторяющихся играх траекторий действий агентов следуют одной общей идее (см. обзоры результатов по так называемым «народным теоремам» (folk theorem) для повторяющихся игр в нормальной форме [15—18] и для иерархических игр [19—22]), а именно, идее «достаточно сильного» *наказания* любого агента за любое (даже однократное) отклонение от требующегося от него выбора. Отметим, что на этой же идее основывается *теорема Гермейера* для игр Γ_2 [23], лежащая в основе всех задач стимулирования, начиная, например, с рассмотренной в статье [24]. На этой же идее базируются результаты о декомпозиции игры агентов в работах [5, 6]. Применим ее для ДСОС.

Построим компенсаторную систему стимулирования в ДСОС, оптимальную в смысле затрат центра (минимизирующую его суммарные (по периодам и агентам) затраты на стимулирование агентов — $\sigma_i^\tau(\cdot)$) в двух постановках, когда верны предположение П1 или предположение П2.

Зафиксируем некоторую желательную для центра траекторию выхода сети $x[1^*t]$, траекторию *планов по результату*, и траекторию *планов по действию* $y_N(x[1^*t])$ — набор векторов действий агентов, приводящих к траектории-выходу $x[1^*t]$.

При справедливости предположения П1 несложно показать, что для любой осуществимой траектории выходов $x[1^*t]$ последовательно по $t = 1, 2, ..., T$ может быть построена единственная (!) траектория планов по действиям всех агентов $y_N(x[1^*t])$, обеспечивающая данный выход сети $z_n[1^*t] = x[1^*t]$.

Обозначим:

$y_N^{\text{п}}[1^*t]$ — траектория планов по действиям всех агентов (запись $x[1^*t]$ будем опускать там, где это не приводит к неоднозначности);

$y_i^{\text{п}}[1^*t]$ — траектория планов по действию i-го агента;

$y_i^{\text{п}}(t)$ — t-й элемент траектории планов по действию $y_i^{\text{п}}[1^*t]$ i-го агента;

$Z_i^t(y) \subseteq Z_n$ — подмножество множества Z_n возможных выходов сети, каждый элемент которого соответствует выходу $z_n(t)$ при условии, что i-й агент в t-м периоде выбрал действие y, а остальные агенты — любые допустимые для них действия при любых траекториях действий $y_N[1*(t-1)]$ всех агентов в течение предыдущих периодов, т. е.

$$Z_i^t(y) = \Big\{ z_n(\{y,\ y_{-i}(t),\ y_N[1*(t-1)]\}):$$

$$\forall y_{-i}(t) \in \prod_{k \in N;\ k \neq i} Y_k;\ \forall y_N[1*(t-1)] \in \prod_{k \in N} (Y_k)^{t-1} \Big\};$$

$u_i > 0$ — *резервная полезность i*-го агента (выигрыш агента при отказе от участия в ОС) в течение одного периода.

Сформируем функцию стимулирования для каждого агента $i \in N$ и $t \in \overline{1, T}$, предполагая соблюдение гипотезы доброжелательности:

$$\hat{\sigma}_i^t(z_n[1*t]) =$$
$$= \begin{cases} c_i^t(y_N(z_n[1*t])) + u_i, & z_n(t) \in Z_i^t(y_i^{\text{п}}(t)), \\ 0, & z_n(t) \notin Z_i^t(y_i^{\text{п}}(t)). \end{cases} \quad (4)$$

С содержательной точки зрения система стимулирования (4) такова, что в каждом периоде каждому агенту гарантируется компенсация его фактических затрат $c_i^t(\cdot)$ и обеспечение резервной полезности u_i в том случае, если он выбрал требуемое от него (плановое) действие $y_i^{\text{п}}(t)$ при любой обстановке $y_{-i}(t) = (y_1(t), ..., y_{i-1}(t), y_{i+1}(t), ..., y_n(t))$ игры для него в текущем периоде (любых фактических действиях других агентов) и при любой истории игры $y_N[1*(t-1)]$ (любой траектории действий всех агентов в течение предыдущих периодов). В противном случае стимулирование равно нулю.

Утверждение 1. *Если выполнено предположение П1, то система стимулирования* (4) *реализует траекторию планов по действию $y_N(x[1*t])$ как РДС игры агентов с минимальными затратами центра на стимулирование. Выигрыши всех агентов в каждом периоде в этом равновесии тождественно равны их резервным полезностям. Затраты центра при этом*

$$C(x[1*T]) = \sum_{t=1}^{T} \delta(1,\ t) \sum_{i=1}^{n} (c_i^t(y_N(x[1*t])) + u_i). \quad (5)$$

Доказательство. Рассмотрим произвольный период времени $t \in \overline{1, T}$, зафиксируем произвольный номер агента $i \in N$. Пусть $y_N[1*(t-1)]$ — некоторая история игры, а $y_{-i}(t)$ — некоторая обстановка игры для этого агента. Рассмотрим все возможные альтернативы поведения i-го агента.

А. Пусть в периоде t агент i выбрал плановое действие $y_i^{\text{п}}(t)$. Тогда $y_N[1*t] = \{y_i^{\text{п}}(t),\ y_{-i}(t),\ y_N[1*(t-1)]\}$ и $z_n(y_N[1*t]) \in Z_i^t(y_i^{\text{п}}(t))$. Согласно функции стимулирования (4) значение его целевой функции

$$f_i(\hat{\sigma}_N^t;\ y_N[1*t];\ t) = \hat{\sigma}_i^t(z_n(y_N[1*t])) - c_i^t(y_N[1*t]) =$$
$$= c_i^t(y_N[1*t]) + u_i - c_i^t(y_N[1*t]) = u_i > 0.$$

Б. Пусть в периоде t агент i отказался от участия в ОС, тогда значение его целевой функции $f_i(\hat{\sigma}_N^t;\ y_N[1*t]) = u_i > 0$. Отметим существенность того, что последнее неравенство строгое, так как если оба неравенства нестрогие ($u_i \geq 0$ и $c_i(\cdot) \geq 0$), то из целевой функции агента — см. далее формулу (6) — последует, что множество РДС содержит не только «стратегии любого агента, состоящие из произвольного сочетания плановых действий и отказа от участия в ОС в каждом периоде», но и, возможно, еще какие-то; поэтому необходимо потребовать, чтобы одно из неравенств было строгим: или $u_i > 0$, или $c_i(\cdot) > 0$ (последнее предположение — слишком сильное, так как обычно считается, что при отказе от участия в ОС агент несет нулевые затраты).

В. Пусть в периоде t агент i выбрал действие $s \in Y_i$, отличающееся от планового: $s \neq y_i^{\text{п}}(t)$. Тогда история игры такова: $y_N[1*t] = \{s;\ y_{-i}(t);\ y_N[1*(t-1)]\}$ и $z_n(y_N[1*t] \notin Z_i^t(y_i^{\text{п}}(t))$ (в силу предположения П1). Согласно выражению (4), значение его целевой функции

$$f_i^t(\hat{\sigma}_N^t;\ y_N[1*t]) = \hat{\sigma}_i^t(z_n(y_N[1*t])) - c_i^t(y_N[1*t]) =$$
$$= 0 - c_i^t(y_N[1*t]) \leq 0.$$

Отметим, что опции А — В для каждого агента и для каждого периода t получены независимо от предыдущих и последующих действий каждого из агентов.

В итоге, целевая функция агента примет вид:

$$F_i(\{\hat{\sigma}_N^t\};\ y_N[1*T]) =$$
$$= u_i\Delta_i - \sum_{\tau \in \Theta_i} \delta_i(1,\ \tau)[u_i + c_i^{\tau}(y_N[1*\tau])], \quad (6)$$

где Θ_i — множество периодов, в которых агент участвовал в ОС, но отклонялся от плановой траектории $y_i^{\text{п}}[1*T]$, а $\Delta_i = \sum_{\tau=1}^{T} \delta_i(1,\ \tau)$.

В формуле (6) уменьшаемое $u_i\Delta_i$ не зависит от выбора агента, а так как $c_i(\cdot) \geq 0$, $u_i > 0$, $\delta_i(\cdot) \geq 0$, то разность (6) достигает максимума, когда множество $\Theta_i = \varnothing$.

Отсюда следует, что стратегия любого агента, состоящая из произвольного сочетания плановых действий и отказа от участия в ОС в каждом периоде, — доминантная. В частности, траектория выполнения планов по действию представляет собой доминантную стратегию: для нее $\Theta_i = \varnothing$, тогда $F_i(\{\hat{\sigma}_N^t\};\ \{y_i^{\text{п}}[1*T];\ y_{-i}[1*T]\}) = u_i\Delta_i$.

Проверим это. Рассмотрим траекторию действий i-го агента, когда на промежутке периодов от первого до t_1-го (не включая его) он реализовывал плановые действия $y_i^{\text{п}}[1*(t_1-1)]$, на промежутке от t_1-го до t_2-го периодов принимал участие в ОС, но реализовывал некоторые действия $s[t_1*t_2]$, отличающиеся от плана: $s[t_1*t_2] \neq y_i^{\text{п}}\,[t_1*t_2]$, далее снова следовал плану по действиям $y_i^{\text{п}}\,[(t_2+1)*T]$.

Значение целевой функции агента при такой траектории действий соотносится с максимальным значением при плановой траектории как

$$F_i(\{\hat{\sigma}_N^t\};\ \{y_i^{\text{п}}\,[1*(t_1-1)],\ s[t_1*t_2],\ y_i^{\text{п}}\,[(t_2+1)*T],$$
$$y_{-i}[1*T]\}) = u_i\Delta_i - \sum_{\tau=t_1}^{t_2} \delta_i(1,\tau)[u_i + c_i^{\tau}(\{y_i^{\text{п}}\,[1*(t_1-1)],$$
$$s[t_1*\tau],\ y_{-i}[1*\tau]\})] < u_i\Delta_i = F_i(\{\hat{\sigma}_N^t\};\ \{y_i^{\text{п}}\,[1*T],\ y_{-i}[1*T]\}).$$

Аналогично, отказ i-го агента от участия в ОС на любом промежутке от t_1-го до t_2-го периодов не увеличивает значение целевой функции.

Согласно выражению (6), максимальные значения целевой функции i-го агента ($u_i\Delta_i$) не зависят от действий других агентов, поэтому траектория $y_N(x[1*T])$, когда все агенты следуют планам по действиям в каждом периоде, представляет собой (одно из) РДС. Тогда в предположении выполнения гипотезы доброжелательности все агенты выберут $y_N(x[1*T])$. Следовательно, система стимулирования (4) реализует плановую траекторию — вектор действий всех агентов $y_N(x[1*T])$ — как РДС, при этом суммарные затраты центра определяются по формуле (5).

Покажем что система стимулирования (4) обеспечивает затраты центра на минимально возможном уровне (5) среди всех систем стимулирования, реализующих выход сети $x[1*T]$.

Пусть существует другая система стимулирования $\{\sigma_N''^{\,t}\}$, реализующая выход сети $x[1*T]$ и, следовательно, тот же вектор действий агентов $y_N(x[1*T])$ и характеризующаяся строго меньшими суммарными затратами центра на стимулирование. Тогда найдется по крайней мере один агент $j \in N$, для которого хотя бы в одном периоде $t \in \overline{1, T}$ выполнено $\sigma_j''^{\,t}(y_N(x[1*t])) < c_j^t(y_N(x[1*t])) + u_j$, т. е. значение его целевой функции строго меньше резервной полезности. Но тогда для этого агента в таком периоде целесообразно отказаться от участия в данной системе и получения выигрыша, равного резервной полезности. Получаем противоречие, утверждение 1 доказано. ♦

Содержательно утверждение 1 означает, что компенсаторная система стимулирования (4) реализует траекторию действий агентов (как РДС их игры), приводящих к требуемой траектории результатов сети, декомпозирует задачу по агентам и по периодам и обеспечивает гарантированно (по всем возможным дальновидностям агентов) минимальные затраты центра на реализацию траектории результатов сети. Система стимулирования (4) при этом такова, что агент получает компенсацию затрат и резервной полезности, если точно следует плану независимо от всех остальных агентов и не получает ничего, отклоняясь от плана. Применяя эту систему стимулирования, центр делает невыгодным для агента даже единичное отклонение от плановой траектории, независимо от дальновидности агента (сколько будущих периодов времени он учитывает, принимая текущее решение). К тому же центр «берет на себя» всю тяжесть решения задачи планирования, в том числе учет взаимовлияния решений, принимаемых в различные периоды с учетом дисконтирования и пр.

Пусть теперь выполнено предположение П2 (центр наблюдает фактические действия агентов в каждом периоде). В этом случае центр может формировать систему стимулирования, пользуясь своей информированностью о действиях агентов, на основе траектории плана по действиям.

Зафиксируем некоторую желательную для центра траекторию *планов по действию* $y_N^{\text{п}}\,[1*t]$. Сформируем функцию стимулирования для каждого $i \in N$ и $t \in \overline{1, T}$, предполагая справедливость гипотезы доброжелательности:

$$\hat{\sigma}_i^t(y_N^{\text{п}}\,[1*t]) = \begin{cases} c_i^t(y_N[1*t]) + u_i, & y_i(t) = y_i^{\text{п}}(t,) \\ 0, & y_i(t) \neq y_i^{\text{п}}(t). \end{cases} \quad (7)$$

Доказательство утверждения 1 может быть распространено на данный случай с точностью до обозначений, поэтому справедливо

Утверждение 2. *Если выполнено предположение П2, то система стимулирования* (7) *реализует траекторию планов по действию* $y_N^{\text{п}}\,[1*t]$ *как равновесие в доминантных стратегиях игры агентов с минимальными затратами центра на стимулирование. Выигрыши всех агентов в каждом периоде в этом равновесии тождественно равны их резервным полезностям. Затраты центра при этом*

$$C(y_N^{\text{п}}\,[1*t]) = \sum_{t=1}^{T} \delta(1,t) \sum_{i=1}^{n} (c_i^t(y_N^{\text{п}}\,[1*t]) + u_i). \ ♦ \quad (8)$$

Если гипотеза доброжелательности не соблюдается, то следует дополнить функции стимулирования (4) и (7) стимулирующими надбавками $\varepsilon_i > 0$, которые соответствующим образом увеличат затраты центра:

— если выполнено предположение П1, то:

$$\hat{\sigma}_i^t(z_n[1*t]) =$$
$$= \begin{cases} c_i^t(y_N(z_n[1*t])) + u_i + \varepsilon_i, & z_n(t) \in Z_i^t(y_i^{\text{п}}(t)), \\ 0, & z_n(t) \notin Z_i^t(y_i^{\text{п}}(t)), \end{cases}$$

$$C(x[1*T]) =$$
$$= \sum_{t=1}^{T} \delta(1, t) \sum_{i=1}^{n} (c_i^t(y_N^*(x[1*t])) + u_i + \varepsilon_i), \quad (9)$$

— если справедливо предположение П2, то:

$$\hat{\sigma}_i^t(y_N^{\text{п}}[1*t]) = \begin{cases} c_i^t(y_N[1*t]) + u_i + \varepsilon_i, & y_i(t) = y_i^{\text{п}}(t), \\ 0, & y_i(t) \neq y_i^{\text{п}}(t), \end{cases}$$

$$C(y_N^{\text{п}}[1*T]) =$$
$$= \sum_{t=1}^{T} \delta(1, t) \sum_{i=1}^{n} (c_i^t(y_N^{\text{п}}[1*t]) + u_i + \varepsilon_i). \quad (10)$$

В этом случае единственным будет РДС, в котором все агенты в каждом периоде выбирают требуемые от них действия $y_i^{\text{п}}(t)$.

ЗАКЛЮЧЕНИЕ

Отметим, что в динамической сетевой организационной системе для любой осуществимой траектории результатов сети $z_n[1*T]$ может быть построена компенсаторная система стимулирования, определяемая выражением (4) или (7) ((9) или (10)) и утверждениями 1 или 2, которая:

— реализует траекторию действий агентов (как РДС их игры), приводящих к требуемой траектории результатов $z_n[1*T]$,

— декомпозирует задачу по агентам и по периодам,

— обеспечивает гарантированно (по всем возможным дальновидностям агентов) минимальные затраты центра на реализацию траектории $z_n[1*T]$, определяемые выражениями (5) или (8).

Важно отметить, что декомпозиция задачи по агентам и периодам не отменяет взаимодействий агентов. При соблюдении предположений П1 или П2 и других условий задачи центру удается построить такую оптимальную систему стимулирования ((4) или (7), или (9), или (10)), которая обеспечивает равновесие в доминантных стратегиях игры агентов независимо от их взаимосвязей — технологии динамической сетевой организационной системе: структуры сети G и совокупности технологических функций.

Во второй части статьи полученные результаты будут распространены на случай многоуровневых ОС и неопределенных затрат агентов.

ЛИТЕРАТУРА

1. *Еналеев А.К.* Разработка механизмов стимулирования и управления в двухуровневых активных системах: автореф. дис. канд. техн. наук. — М.: МФТИ, 1980. — 18 с. [*Enaleev, A.K.* Razrabotka mekhanizmov stimulirovaniya i upravleniya v dvukhurovnevykh aktivnykh sistemakh: avtoref. dis. kand. tekhn. nauk. — M.: MFTI, 1980. — 18 s. (In Russian)]
2. *Бурков В.Н., Еналеев А.К., Кондратьев В.В.* Двухуровневые активные системы. IV. Цена децентрализации механизмов функционирования // Автоматика и телемеханика. — 1980. — № 6. — С. 110—116. [*Burkov, V.N., Enaleev, A.K., Kondrat'ev, V.V.* Dvukhurovnevye aktivnye sistemy. IV. Tsena detsentralizatsii mekhanizmov funktsionirovaniya // Avtomatika i telemekhanika. — 1980. — No. 6. — S. 110—116. (In Russian)]
3. *Ашимов А.А., Бурков В.Н., Джапаров Б.А., Кондратьев В.В.* Согласованное управление активными производственными системами. — М.: Наука, 1986. — 248 с. [*Ashimov, A.A., Burkov, V.N., Dzhaparov, B.A., Kondrat'ev, V.V.* Soglasovannoe upravlenie aktivnymi proizvodstvennymi sistemami. — M.: Nauka, 1986. — 248 s. (In Russian)]
4. *Бурков В.Н., Кондратьев В.В.* Механизмы функционирования организационных систем. — М.: Наука, 1981. — 384 с. [*Burkov, V.N., Kondrat'ev, V.V.* Mekhanizmy funktsionirovaniya organizatsionnykh sistem. — M.: Nauka, 1981. — 384 s. (In Russian)]
5. *Новиков Д.А.* Теория управления организационными системами: 3-е изд. — М.: Физматлит, 2012. — 604 с. [*Novikov, D.A.* Teoriya upravleniya organizatsionnymi sistemami: 3-e izd. — M.: Fizmatlit, 2012. — 604 s. (In Russian)]
6. *Белов М.В., Новиков Д.А.* Сетевые активные системы: модели планирования и стимулирования // Проблемы управления. — 2018. — № 1. — С. 54—64. [*Belov, M.V., Novikov, D.A.* Network Active Systems: Planning and Simulation Models // Control Sciences. — 2018. — No. 1. — P. 54—64. (In Russian)]
7. *Еналеев А.К.* Модели и методы построения механизмов стимулирования в сетевых организационных структурах / Материалы 11-й Междунар. конф. «Управление развитием крупномасштабных систем» (MLSD'2018, Москва). — М.: ИПУ РАН. — 2018. — Т. 1. — С. 150—152. [*Enaleev, A.K.* Optimal Mechanism at Network Active Systems / 2018 Eleventh International Conference «Management of large-scale system development» (MLSD). — Moscow: IPU RAN. — 2018. — Vol. 1. — P. 150—152. (In Russian)]
8. *Еналеев А.К.* Оптимальность согласованных механизмов в сетевых организационных структурах // Проблемы управления. — 2020. — № 1. — С. 24—38. [*Enaleev, A.K.* Optimality of the Incentive Compatible Mechanisms in Network Organizational Structures // Control Sciences. — 2020. — No. 1. — P. 24—38. (In Russian)]
9. *Новиков Д.А.* Сетевые структуры и организационные системы. — М.: ИПУ РАН, 2003. — 102 с. [*Novikov, D.A.* Setevye struktury i organizatsionnye sistemy. — M.: IPU RAN, 2003. — 102 s. (In Russian)]
10. *Новиков Д.А., Шохина Т.Е.* Механизмы стимулирования в динамических активных системах // Автоматика и телемеханика. — 2003. — № 12. — С. 94—104. [*Novikov, D.A., Shokhina, T.E.* Incentive mechanisms in dynamic active systems // Automation and Remote Control. — 2003. — Vol. 64, no. 12. — P. 1912—1921.]
11. *Новиков Д.А. Смирнов И.М., Шохина Т.Е.* Механизмы управления динамическими активными системами. — М.:

ИПУ РАН, 2002. — 124 с. [*Novikov, D.A. Smirnov, I.M., Shokhina, T.E.* Mekhanizmy upravleniya dinamicheskimi aktivnymi sistemami. — M.: IPU RAN, 2002. — 124 s. (In Russian)]
12. *Бурков В.Н., Горгидзе И.А., Ловецкий С.Е.* Прикладные задачи теории графов. — Тбилиси: ВЦ АН ГССР, 1974. — 232 с. [*Burkov, V.N., Gorgidze, I.A., Lovetskii, S.E.* Prikladnye zadachi teorii grafov. — Tbilisi: VTs AN GSSR, 1974. — 232 s. (In Russian)]
13. *Новиков Д.А.* Механизмы функционирования многоуровневых организационных систем. — М.: Фонд «Проблемы управления», 1999. — 161 с. [*Novikov, D.A.* Mekhanizmy funktsionirovaniya mnogourovnevykh organizatsionnykh sistem. — M.: Fond «Problemy upravleniya», 1999. — 161 s. (In Russian)]
14. *Новиков Д.А., Цветков А.В.,* Агрегирование информации в моделях стимулирования // Автоматика и телемеханика. — 2001. — № 4. — С. 120—127. [*Novikov, D.A., Tsvetkov, A.V.,* Aggregation of information in incentive models // Automation and Remote Control. — 2001. — Vol. 62, no. 4. — P. 617—623.]
15. *Myerson, R.* Game Theory: Analysis of Conflict. — London: Harvard University Press, 1997. — 584 p.
16. *Fudenberg, D., Tirole, J.* Game Theory. — Cambridge: MIT Press, 1995. — 579 p.
17. *Fudenberg, D., Maskin, E.* The folk theorem in repeated games with discounting or with incomplete information // Econometrica. — 1986. — Vol. 54 (3). — P. 533—554.
18. *Demeze-Jouatsa, G.* A complete folk theorem for finitely repeated games. — Bielefeld: Center for Mathematical Economics, 2018. — 32 p.
19. *Maschler, M., Solan, E, Zamir, S.* Game Theory. — Cambridge: Cambridge University Press, 2013. — 1008 p.
20. *Горелов М.А., Кононенко А.Ф.* Динамические модели конфликтов. III. Иерархические игры // Автоматика и телемеханика. — 2015. — № 2. — С. 89—106. [*Gorelov, M.A., Kononenko, A.F.* Dynamic models of conflicts. III. Hierarchical games // Automation and Remote Control. — 2015. — Vol. 76, no. 2. — P. 264—277.]
21. *Рохлин Д.Б., Угольницкий Г.А.* Равновесие Штакельберга в динамической модели стимулирования с полной информацией // Автоматика и телемеханика. — 2018. — № 4. — С. 152—166. [*Rokhlin, D.B., Ougolnitsky, G.A.* Stackelberg Equilibrium in a Dynamic Stimulation Model with Complete Information // Automation and Remote Control. — 2018. — Vol. 79, no. 4. — P. 701—712.]
22. *Угольницкий Г.А.* Управление устойчивым развитием активных систем. — Ростов-на-Дону: Изд-во ЮФУ, 2016. — 938 с. [*Ugol'nitskii, G.A.* Upravlenie ustoichivym razvitiem aktivnykh sistem. — Rostov-na-Donu: Izdatel'stvo YuFU, 2016. — 938 s. (In Russian)]
23. *Гермейер Ю.Б.* Игры с непротивоположными интересами. — М.: Наука, 1976. — 327 с. [*Germeier, Yu.B.* Igry s neprotivopolozhnymi interesami. — M.: Nauka, 1976. — 327 s. (In Russian)]
24. *Гермейер Ю.Б., Ерешко Ф.И.* Побочные платежи в играх с фиксированной последовательностью ходов // Журнал вычислительной математики и математической физики. — 1974. — № 14. — С. 1437—1450. [*Germeier, Yu.B., Ereshko, F.I.* Pobochnye platezhi v igrakh s fiksirovannoi posledovatel'nost'yu khodov // Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki. — 1974. — No. 14. — S. 1437—1450. (In Russian)]

Статья представлена к публикации членом редколлегии В.Н. Бурковым.

Поступила в редакцию 11.02.2019, после доработки 26.09.2019.
Принята к публикации 24.10.2019.

Белов Михаил Валентинович — д-р техн. наук, компания ИБС, г. Москва, ✉ mbelov59@mail.ru.

INCENTIVE-COMPATIBLE CONTROL IN DYNAMIC MULTI-AGENT SYSTEMS. Part 1. Contracts in Dynamic System with One Principal and Multiple Agents

M.V. Belov

IBS company, Moscow, Russia
✉ mbelov59@mail.ru

Abstract: The formulation and solution are considered of the problems of coordinated control of multi-element dynamic active systems (AS) with restrictions on the joint activity of their elements in the form of technological networks. An AS is studied consisting of one principal and many agents subordinate to it. It has been proved that for any admissible trajectory of results, a coordinated compensatory incentive system can be constructed that implements (as an equilibrium in dominant strategies) the trajectory of the agents leading to the desired trajectory of results; decomposes the control task by agents and by time periods; provides guaranteed (for all possible far-sighted agents) minimum costs of the governing body of the principal for the implementation of this trajectory of results. It is shown that in such incentive systems, the values of payments depend only on the corresponding values of the cost functions, which, in turn, indirectly take into account the technological functions, network structure and AS structure as a whole. The problem of optimal planning is posed and an algorithm for solving it is indicated.

Keywords: incentive problem, dynamic system with one principal and multiple agents, contract theory.