© 2002 г. А. П. КАРАВАЕВ
(Российская экономическая школа, Москва)

# ПАРЕТО-ЭФФЕКТИВНОСТЬ РАВНОВЕСИЙ В АКТИВНЫХ СИСТЕМАХ С РАСПРЕДЕЛЕННЫМ КОНТРОЛЕМ

Рассматриваются активные системы с распределенным контролем (один активный элемент и несколько управляющих центров). Исследуются вопросы существования равновесия Нэша игры центров в чистых стратегиях. Показывается, что при любом распределении доходов от деятельности активной системы между несколькими центрами и существовании по крайней мере одного Парето-неэффективного равновесия Нэша любой Парето-эффективный исход может быть реализован как равновесие Нэша. Доказывается, что в системе, состоящей из двух центров, всегда существует Парето-эффективное равновесие Нэша в чистых стратегиях.

## 1. Введение

В литературе (см., например, [1, 2]) достаточно хорошо изучены вопросы функционирования активных систем с одним центром и несколькими подчиненными активными элементами. Однако на практике достаточно часто можно встретиться с активными системами, в которых присутствуют несколько центров или когда высшее руководство может определять количество центров в системе.

При рассмотрении активных систем с несколькими управляющими центрами ситуация кардинально отличается от случая активной системы с одним центром. Основное отличие состоит в наличии игры между центрами, причем в силу их возможных противоположных интересов активному элементу отводится более важная роль: равновесия при наличии нескольких центров характеризуются тем, что активный элемент получает не только компенсацию своих затрат на реализацию действия, но и дополнительную ренту за то, что не выбирает действие, более выгодное какому-либо одному из центров.

При увеличении количества центров и при распределении доходов активной системы между ними множество реализуемых в активной системе с одним центром исходов может быть сильно изменено: с одной стороны, неизвестно, будут ли реализуемы (как равновесия) прежние действия, а, с другой стороны, за счет игры центров друг с другом множество равновесий может расшириться.

Специфика равновесий в модели с несколькими центрами заключается в следующем. Центры должны думать о стратегиях друг друга, и, отклоняясь от своей стратегии, должны понимать, что против них может быть образована коалиция, целью которой является сделать для активного элемента более привлекательным реализацию такого исхода, при котором доходы отклоняющегося центра были бы значительно меньше, чем раньше.

В данной статье подробно исследуются вопросы, связанные с такими равновесиями, и устанавливается, что в равновесии против каждого из центров должны быть образованы одинаково сильные коалиции. Находится тот максимальный уровень "угрозы" противостоящей коалиции, который может существовать в активной системе. Показывается, что существуют активные системы, в которых не существует равновесий без угроз.

При анализе равновесий в активной системе из двух центров находятся условия, которые описывают все множество возможных исходов при заданной угрозе. Показывается, что при увеличении угрозы множество исходов может как уменьшаться, так и увеличиваться.

В статье также исследуются вопросы Парето-эффективности возможных исходов. Доказывается, что в системе с двумя центрами при достаточно широких предположениях Парето-эффективный исход достижим. При существовании какого-либо равновесия Нэша показывается, что существует Парето-эффективное равновесие Нэша (при первоначальном Парето-неэффективном равновесии для любого Парето-эффективного исхода существует такое равновесие), при котором выигрыши каждого из центров и активного элемента не уменьшаются по сравнению с первоначальным равновесием Нэша.

Настоящая статья – продолжение работы [3] и организована следующим образом. В первой части строится модель и описывается порядок функционирования системы. Во второй части вводятся основные определения. В третьей части приводится обзор литературы. В четвертой части приводятся основные результаты: теорема о реализуемости Парето-эффективного исхода в активной системе из двух центров и одного активного элемента, характеризация равновесий с угрозами в общем случае, критерий реализуемости некоторого исхода в виде равновесия Нэша в активной системе с двумя центрами, теорема о реализуемости любого Парето-эффективного равновесия в активной системе из двух центров, в которой реализуемо Парето-неэффективное равновесие.

## 2. Описание модели

Будем рассматривать (см. [1]) двухуровневую активную систему (АС), состоящую из $n > 1$ центров и $m \geq 1$ активных элементов (АЭ).

Обозначим через $H_p(x) : X \to \mathbb{R}$ функцию дохода $p$-го центра, $p = \overline{1,n}$, в зависимости от действия $x \in X$, реализованного всей системой в целом, где $X$ – множество возможных действий системы, $c_i(x_i)$ – функция затрат $i$-го АЭ в зависимости от выбранного им действия $x_i \in X_i$, где $X_i$ – множество возможных действий $i$-го АЭ, $i = \overline{1,m}$.

Действие всей системы задается как некоторая функция $G$ от действий АЭ $X_1 \times \times X_2 \times \ldots \times X_m \to X$ (заметим, что в общем случае можно положить: $X = X_1 \times X_2 \times \times \ldots \times X_m$ и $G(\cdot)$ – тождественная функция).

Задачей каждого из центров является максимизация своей целевой функции

$$H_p(G(x_1, x_2, \ldots, x_m)) - \sum_{i=1}^{m} \sigma_{pi}(x_i)$$

путем выбора вектора функций $(\sigma_{p1}(x_1), \ldots, \sigma_{pm}(x_m))$, где $\sigma_{pi} : X_i \to \mathbb{R}_1^+$ – стимулирования $i$-го АЭ при выборе им действия $x_i$, $p = \overline{1,n}$, $i = \overline{1,m}$.

Примем следующий порядок функционирования АС. Сначала $p$-й центр, зная функции $c_i(x_i)$, $i = \overline{1,m}$, и $H_p(x)$, но не зная, какие функции стимулирования выберут другие центры (т.е. центры выбирают свои функции стимулирования одновременно), задает вектор-функцию стимулирования $\sigma_p(x_1, \ldots, x_m) = (\sigma_{p1}(x_1), \ldots \ldots, \sigma_{pm}(x_m))$.

$i$-й АЭ на основании функций стимулирования $\sigma_{pi}(x_i)$ и функции затрат $c_i(x_i)$ выбирает некоторое действие (исход) $x_i$ из множества своих возможных действий $X_i$. Считаем, что АЭ при известных функциях стимулирования посредством выбора действия $x_i$ максимизирует свою целевую функцию

$$(1) \qquad \sum_{p=1}^{n} \sigma_{pi}(x_i) - c_i(x_i) \to \max_{x_i}, \quad i = \overline{1,m}$$

и каждый из центров знает о правилах поведения всех АЭ.

В случае, если максимум (1) достигается в нескольких точках, будем предполагать, что АЭ выбирает действие в соответствии с некоторой функцией $\Psi_i : 2^{X_i} \to X_i$:

$$x_i = \Psi_i \left( \operatorname*{Argmax}_{x_i \in X_i} \left( \sum_{p=1}^{n} \sigma_{pi}(x_i) - c_i(x_i) \right) \right), \quad \text{где} \quad \Psi_i(A) \in A \quad \forall A \subseteq X_i.$$

$p$-й центр будет выбирать такие функции стимулирования, чтобы при рациональном выборе АЭ своих действий доставить максимум своей целевой функции

$$H_p(G(x_1, x_2, \dots, x_m)) - \sum_{i=1}^{m} \sigma_{pi}(x_i), \quad p = \overline{1,n}.$$

Обобщая сказанное, запишем порядок функционирования АС следующим образом:

1) центры выбирают функции стимулирования;

2) АЭ на основании функций стимулирования выбирают действия;

3) в АС одновременно происходят все выплаты.

Необходимо отметить, что в данной модели предполагается, что все участники АС сообщают друг другу только правду и всегда выполняют свои обещания (по выплатам).

Для дальнейшего изложения ограничимся случаем $m = 1$ (в системе присутствует только один АЭ) и введем следующие предположения.

**А1.** $X = X_1$ есть компактное подмножество множества $\mathbb{R}_+^\ell$ ($\ell \geq 1$), а $G(x_1)$ – тождественная функция:

$$G(x) = x.$$

**А2.** Функция затрат неотрицательна и всюду непрерывна.

**А3.** Функции дохода центров неотрицательны и всюду непрерывны.

Вместо предположения **А1** можно потребовать (наряду с **А2** и **А3**) замкнутость множества $X = X_1$ и существование такой точки $x^0$, что

$$\sum_{p=1}^{n} H_p(x) - c(x) < 0 \quad \text{при} \quad x > x^0.$$

Введем множество $M'$ неотрицательных всюду определенных на $X$ функций стимулирования (функций стимулирования, при которых центры не могут ничего забирать у АЭ):

$$M' = \{\sigma(x) : X \to [0, +\infty)\}.$$

Введем также множество $M''$ полунепрерывных сверху функций:

$$M'' = \{\sigma(x) : X \to \mathbb{R} : \quad \forall x_j \in X, x_j \to \overline{x}, \sigma(x_j) \geq \sigma(\overline{x}) \Rightarrow \sigma(x_j) \to \sigma(\overline{x})\}$$

(для многих случаев просто непрерывность является слишком жестким требованием, которое не позволит доказать многие результаты, в то время как полунепрерывности сверху будет достаточно. К тому же часто используемые квазикомпенсаторные функции стимулирования полунепрерывны сверху).

Будем считать, что класс возможных функций стимулирования ограничен пересечением множеств $M'$ и $M''$:

$$M = M' \cap M'',$$

т.е. центры ничего не забирают у АЭ (штрафы запрещены), а только могут предложить ему стимулирование за выбор действия, и функции стимулирования являются полунепрерывными сверху.

Будем считать, что у АЭ есть право участия: если при любом действии его целевая функция меньше нуля (стимулирования заданы), то он может отказаться от участия в АС. В этом случае будем полагать выигрыши центров равными нулю (действие всей системы не определено). Для этого можно предположить наличие особой (возможно, изолированной) точки $\widehat{x}$: $H_p(\widehat{x}) = 0$, $p = \overline{1,n}$, $c(\widehat{x}) = 0$. Выбрав эту точку, АЭ ничего не теряет (его затраты нулевые), и стимулирования в этой точке равны нулю (поскольку центры не могут предложить больше, чем сами получат в этой точке, а функции стимулирования не могут быть отрицательными). Часто в моделях в роли такой выделенной точки $\widehat{x}$ выступает точка $x = 0$.

В дальнейшем вместо функции $\Psi$ выбора АЭ будем указывать один элемент $x^*$ множества $X$, который будет выбираться как наилучший с точки зрения АЭ при прочих равных возможностях: если он входит в множество

$$\operatorname*{Argmax}_{x \in X} \left( \sum_{i=1}^{n} \sigma_i(x) - c(x) \right)$$

оптимальных выборов АЭ, то АЭ выберет именно $x^*$. Этого будет достаточно, поскольку в данной статье мы будем рассматривать только равновесные стратегии (для которых и важен данный $x^*$), а при отклонении от равновесной стратегии, если АЭ может получить ненулевую выгоду, то он, увеличивая стимулирование, может сделать так, что множество лучших для АЭ точек будет тривиальным (состоять из одного элемента).

Заметим, что в данной модели бо́льшее количество центров не означает бо́льший контроль за АЭ (более точное знание исхода), как это бывает в жизни, поскольку мы предположили, что центр (центры) точно осведомлены о действии АЭ.

В дальнейшем в статье будем предполагать, что соответствующие множества Argmax непусты, минимумы и максимумы, которые встретятся, достижимы. Эти требования можно обосновать с помощью непрерывности (полунепрерывности сверху) соответствующих функций затрат и доходов и замкнутости множества $X$.

## 3. Основные определения

Назовем стратегии центров $(\sigma_p(x))_{p=1}^{n}$ *допустимыми*, если они удовлетворяют следующим свойствам (см. [3]):

1) функции стимулирования неотрицательны и принадлежат определенному ранее классу функций стимулирования $M$:

$$\sigma_p(x) \in M \qquad \forall p = \overline{1,n};$$

2) отсутствует "блеф", т.е. центры обещают столько, сколько они реально могут заплатить:

$$H_p(x) \geq \sigma_p(x) \qquad \forall p = \overline{1, n}.$$

Будем говорить, что $((\sigma_p^*(x)), x^*)$ – фиксированные стратегии центров при заданном действии АЭ $x^*$ – образуют *равновесие Нэша*, если:

1) стратегии $\sigma_p^*(x)$ допустимы;

2) АЭ выбирает действие

$$x^* \in \operatorname*{Argmax}_{x \in X} \left( \sum_{p=1}^{n} \sigma_p^*(x) - c(x) \right);$$

3) при одностороннем изменении любым из центров своей стратегии на $\sigma_p(x)$ он не сможет увеличить значение своей целевой функции вне зависимости от предпочтений АЭ:

$$\forall p = \overline{1, n},\ \forall \sigma_p(x) \in M,\ \sigma_p(x) \leq H_p(x),$$

$$\forall x \in \operatorname*{Argmax}_{x \in X} \left( \sum_{i=1, i \neq p}^{n} \sigma_i^*(x) + \sigma_p(x) - c(x) \right) \Rightarrow$$

$$\Rightarrow H_p(x) - \sigma_p(x) \leq H_p(x^*) - \sigma_p^*(x^*).$$

Значение $x^*$ при этом будем называть *исходом* равновесия Нэша, *реализуемым* АС при данных равновесных стратегиях. Не будем различать равновесия Нэша, которые реализуют один и тот же исход и при которых выигрыши (значения целевых функций) центров и АЭ одинаковы.

Будем говорить, что $x^*$ есть *Парето-эффективный для всей АС исход* (социальный Парето-оптимум), если

$$x^* \in \operatorname*{Argmax}_{x \in X} \left( \sum_{p=1}^{n} H_p(x) - c(x) \right).$$

Заметим, что, вообще говоря, из определений не следует, что любой Парето-эффективный исход может быть реализован как равновесие Нэша.

В силу компактности множества $X$ множество

$$\left\{ x \in X : \sum_{p=1}^{n} H_p(x) - c(x) \geq 0 \right\}$$

ограничено (из непрерывности и полунепрерывности сверху функций $c(x)$ и $H_p(x)$ следует замкнутость). В силу предположения о праве участия (существовании точки $\widehat{x}$) АЭ это множество непусто.

## 4. Обзор литературы

В [1] и [2] подробно рассматриваются АС, состоящие из одного центра и одного АЭ. Показывается, что оптимальным стимулированием в данном случае является так называемая квазикомпенсаторная система стимулирования, а именно, такая система стимулирования, при которой центр будет производить выплаты АЭ только в том случае, если он выбрал определенное действие, которое требуется центру. В [4] изучаются АС с распределенным контролем. Случай нескольких АЭ (в том числе, непрерывного распределения элементов) рассмотрен в [5–7].

В [3] рассматривается модель нескольких центров и одного АЭ. Подробно обсуждается игра, которая возникает между центрами. Показывается, в частности, что каждый из $n$ центров может без потери общности использовать так называемые $(n+1)$-пиковые функции стимулирования, т.е. такие функции стимулирования, которые равны нулю всюду, кроме $n+1$ точки. А именно, верна

Теорема 1 ([3]). *Пусть $((\sigma_i(x)), x^*)$ – некоторое равновесие Нэша. Тогда для любого $p$ существует такая $n$-пиковая функция стимулирования $\widetilde{\sigma}_p(x)$, что $((\sigma_i(x))_{i\neq p}, \widetilde{\sigma}_p(x), x^*)$ – равновесие Нэша, при котором выигрыши центров и АЭ такие же, как и в исходном равновесии.*

Применяя данную теорему несколько раз, получаем, что каждый из центров может ограничиться $n$-пиковой функцией стимулирования (пики всех функций стимулирования расположены в $n+1$ точке множества $X$) и, следовательно, при поиске равновесий Нэша можно ограничиться только такими функциями стимулирования. Также в [3] все возможные равновесия классифицируются на равновесия типа "сотрудничество" и "конкуренция". Различие между ними состоит в том, что в первом случае АЭ получает только компенсацию своих затрат на выбор действия $x^*$. При реализации равновесия типа "конкуренция" АЭ получает от центров дополнительно к затратам сумму

$$s = \sum_{p=1}^{n} \sigma_p(x^*) - c(x^*) > 0,$$

которая называется *угрозой*. Название связано с тем, что, несмотря на некоторый "запас" суммы функций стимулирования в точке $x^*$ (АЭ получает больше, чем ему надо для выбора действия $x^*$), ни один из центров не уменьшает стимулирования в этой точке, поскольку другие центры угрожают ему выбором невыгодного для него действия. Величину $s$ будем в дальнейшем называть *угрозой* или *величиной угрозы*.

## 5. Основные результаты

Не все исходы могут быть реализованы как равновесия без угроз ("сотрудничество"), более того, приведем пример, в котором могут быть только равновесия типа "конкуренция".

Рассмотрим ситуацию, приведенную на рис. 1.

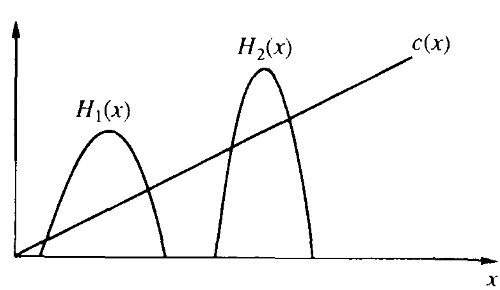


Рис. 1. Пример отсутствия равновесия без угроз.

При любом равновесии $x^*$ и отсутствии угроз $i$-му центру будет выгодно отклониться, если

$$x^* \notin \{x : H_i(x) > 0\},$$

поскольку в этом случае он не получает ничего – $H_i(x^*) = 0$, причем в силу отсутствия угроз достаточно назначить функцию стимулирования

$$\sigma_i(x) = \begin{cases} c(x) + \varepsilon, & x \in \operatorname*{Argmax}_{x \in X}(H_i(x) - c(x)); \\ 0, & \text{иначе}. \end{cases}$$

Заметим, что поскольку

$$\{x : H_1(x) > 0\} \cap \{x : H_2(x) > 0\} = \varnothing,$$

при любом $x^*$ какому-либо из центров обязательно выгодно отклониться. Таким образом, в данном примере существуют только равновесия с угрозой.

Равновесие с угрозой в случае с произвольным числом центров описывает следующая

*Теорема 2. Пусть $n > 1$ и $((\sigma_i(x)), x^*)$ – равновесие Нэша. В том случае, если это равновесие Нэша является равновесием с угрозой (т.е. если $\sum_{i=1}^{n} \sigma_i(x^*) - c(x^*) > > 0$), то для любого $p = \overline{1,n}$ существует такая точка $\widetilde{x} \in X$, что*

$$\sum_{i=1, i \neq p}^{n} \sigma_i(\widetilde{x}) - c(\widetilde{x}) = \sum_{i=1}^{n} \sigma_i(x^*) - c(x^*) = \max_{x \in X} \left( \sum_{i=1}^{n} \sigma_i(x) - c(x) \right). \tag{2}$$

Доказательство теоремы 2 приведено в Приложении.

Таким образом, против каждого из центров остальные центры (или их часть) в точке $\widetilde{x}$ образуют коалицию. Эта коалиция угрожает тем, что при изменении центром своей стратегии из-за наличия угрозы АЭ будет выбирать неблагоприятное для отклонившегося центра равновесие, что в итоге удержит его от изменения своей стратегии. Данное равновесие, не позволяющее ни одному из центров изменить свою стратегию, неустойчиво к изменению стратегий несколькими центрами одновременно (при сговоре), т.е. несколько центров, сговорившись, могут так изменить свои стратегии, что новый исход для них может быть более выгоден, чем старый.

По аналогии с теоремой 2 можно доказать, что для того, чтобы двум центрам было невыгодно отклоняться, необходимо и достаточно, чтобы против каждых двух центров существовала коалиция, которая бы угрожала в случае их отклонения. Разумеется, что в данном случае множество исходов, которые могут быть реализованы, сузится, поскольку равновесие с невозможностью отклонения двух центров является равновесием с невозможностью отклонения одного центра.

Тривиальными следствиями из теоремы 2 являются:

*Следствие 1. В равновесии максимально возможная угроза не может быть больше*

$$\min_{p} \left( \max_{x \in X} \left( \sum_{i=1, i \neq p}^{n} H_i(x) - c(x) \right) \right).$$

Доказательство следствия 1 приведено в Приложении.

*Следствие 2. Пусть $((\sigma_p(x)), x^*)$ – равновесие Нэша. Тогда существует такое конечное множество $Y \subset X$, $|Y| \leq n+1$, $x^* \in Y$, что:*

$$\forall x \in Y \quad \sum_{p=1}^{n} \sigma_p(x) - c(x) = \sum_{p=1}^{n} \sigma_p(x^*) - c(x^*);$$

$$\forall p \in \overline{1,n}\ \exists x \in Y : \ \sigma_p(x) = 0;$$

$$\forall x \in Y\ \exists p \in \overline{1,n} : \ \sigma_p(x) = 0.$$

*Кроме того, набор $((\widetilde{\sigma}_p(x)), x^*)$, где*

$$\widetilde{\sigma}_p(x) = \begin{cases} \sigma(x), & x \in Y; \\ 0, & \text{иначе} \end{cases}$$

*образует равновесие Нэша.*

Доказательство следствия 2 приведено в Приложении.

Следствие 2 говорит о свойствах равновесных $(n+1)$-пиковых функций стимулирования. Множество угроз $Y$ (а именно это множество и является множеством угроз, и если мы обнулим все функции стимулирования, а в точках множества $Y$ оставим их прежними, то получившиеся функции будут равновесными $(n+1)$-пиковыми функциями, причем пики будут в одних и тех же точках у различных функций стимулирования) обладает следующим свойством: в каждой точке множества $Y$ сосредоточена угроза против какого-то центра, причем для любого центра найдется точка из множества $Y$, в которой против него сосредоточена угроза. В точке угрозы против центра с номером $p$ функция стимулирования самого центра, против которого направлена угроза, равна нулю, и размер угрозы равен переплате АЭ в равновесной точке.

Рассмотрим подробнее равновесие Нэша с угрозой $s$ в случае двух центров. Естественно, что каждому из центров выгодно угрожать противнику в той точке, в которой его (противника) выигрыш минимален, т.е. в точках множества $\operatorname{Argmin}\limits_{x: H_i(x) - c(x) \geq s} H_{-i}(x)$, чтобы противник не согласился на реализацию исхода угрозы, где $H_{-i}(x) = \sum\limits_{j \neq i} H_j(x)$. Заметим, что в связи с предположением о непрерывности функций $H_p(x)$ данный минимум достигается (если множество $\{x : H_i(x) - c(x) \geq s\}$ не пусто). Исходя из этого, введем функции

$$f_1(s) = \min_{x: H_2(x) - c(x) \geq s} H_1(x) \quad \text{и} \tag{3}$$
$$f_2(s) = \min_{x: H_1(x) - c(x) \geq s} H_2(x).$$

Введенные функции $f_i(s)$ есть минимальные значения дохода $i$-го центра на подмножестве $X$, где $(-i)$-й центр может обеспечить угрозу $s$. Пусть $x_1^*(s)$ и $x_2^*(s)$ – решения (любые) минимизационных задач (3) для соответственно функций $f_2(s)$ и $f_1(s)$. Тогда выполняется

$$H_i(x_i^*(s)) - c(x_i^*(s)) \geq s \quad \text{и} \qquad H_i(x_{-i}^*(s)) = f_i(s).$$

Определим $a_i = \max\limits_{x \in X}(H_i(x) - c(x))$, а $x_i = x_i^*(a_i)$. Тогда

$$a_i = \max_{x \in X}(H_i(x) - c(x)) = H_i(x_i) - c(x_i),$$

и по следствию 2 в равновесии угроза не может превосходить значения $\min(a_1, a_2)$.

Значение $a_i$ есть максимально возможный выигрыш (значение целевой функции) в игре $i$-го центра с АЭ, если в АС отсутствует другой центр, а $x_i$ – исход, при котором этот выигрыш достигается. Заметим, что функции $f_i(s)$ определены при $s \in [0, a_{-i}]$, являются неубывающими и непрерывными слева.

Далее будем предполагать, что $x_1^*(s)$ и $x_2^*(s)$ ни при каком $s$ не совпадают (во многих случаях, впрочем, это условие можно ослабить).

Пусть мы хотим определить, может ли некоторая точка $x^*$ множества $X$ быть реализована как равновесие Нэша (возможно, с угрозой). Рассмотрим необходимые и достаточные условия для этого в терминах введенных функций.

Прежде всего, каждому из центров должно быть невыгодно полностью отказываться от своего стимулирования (предлагать АЭ во всех точках нулевое стимулирование) и соглашаться на то, чтобы реализовывалась угроза противника, т.е.

$$H_1(x_2^*(s)) \le H_1(x^*) - \sigma_1(x^*);$$
$$H_2(x_1^*(s)) \le H_2(x^*) - \sigma_2(x^*).$$

Далее, никакому из центров не должно быть выгодно отклоняться на реализацию своей наилучшей стратегии $x_i$, т.е.

$$H_1(x_1) - c(x_1) - s \le H_1(x^*) - \sigma_1(x^*),$$
$$H_2(x_2) - c(x_2) - s \le H_2(x^*) - \sigma_2(x^*).$$

Объединяя четыре приведенные выше неравенства и заменяя в них $H_i(x_{-i}^*(s))$ на $f_i(s)$, получаем, что

$$\max(f_1(s), a_1 - s) \le H_1(x^*) - \sigma_1(x^*);$$
$$\max(f_2(s), a_2 - s) \le H_2(x^*) - \sigma_2(x^*).$$

Таким образом, для того чтобы существовали равновесные $\sigma_i(x)$, при которых с угрозой $s$ реализовывалось действие $x^*$, необходимо и достаточно, чтобы выполнялись неравенства

$$(4) \qquad \begin{cases} \max(f_1(s), a_1 - s) \le H_1(x^*); \\ \max(f_2(s), a_2 - s) \le H_2(x^*); \\ \max(f_2(s), a_2 - s) + \max(f_1(s), a_1 - s) \le \\ \le H_1(x^*) + H_2(x^*) - c(x^*) - s. \end{cases}$$

Последнее неравенство говорит о следующем. Для того чтобы, например, первому центру не стала выгодным реализация исходов $x_1$ или $x^*(s)$, надо, чтобы при исходе $x^*$ он получал больше. Таким образом, необходимо, чтобы он при $x^*$ получал как минимум $\max(f_2(s), a_1 - s)$. То же самое верно и для второго центра. В совокупности в $x^*$ они получают $H_1(x^*) + H_2(x^*) - c(x^*) - s$. Третье неравенство говорит о том, что центры имеют что делить, причем сумма для дележа равна разности между правой и левой частями неравенства, так как $\sigma_1(x^*) + \sigma_2(x^*) = c(x^*) + s$.

При этом стимулирования в равновесии (и соответственно выплаты центрам) определяются следующим образом:

$$(5) \qquad \begin{aligned} &\sigma_1(x_1^*(s)) = c(x_1^*(s)) + s; \\ &\sigma_2(x_2^*(s)) = c(x_2^*(s)) + s; \\ &\sigma_1(x^*) + \sigma_2(x^*) = c(x^*) + s; \\ &H_1(x^*) - \sigma_1(x^*) \ge \max(f_1(s), a_1 - s); \\ &H_2(x^*) - \sigma_2(x^*) \ge \max(f_2(s), a_2 - s). \end{aligned}$$

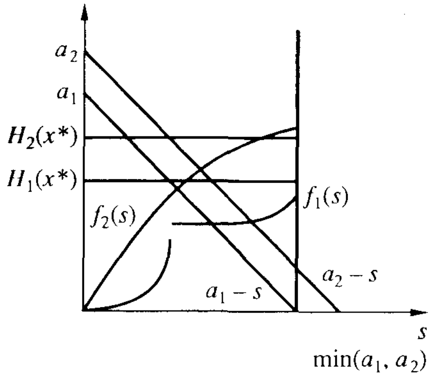


Рис. 2. Анализ равновесий Нэша в АС с двумя центрами.

Существование таких $\sigma_1(x_1^*(s))$, $\sigma_2(x_2^*(s))$, $\sigma_1(x^*)$ и $\sigma_2(x^*)$ обеспечивается системой неравенств (4).

Отличительной от [3] особенностью полученных результатов является исследование вопроса о том, какие исходы реализуемы при данной угрозе и реализуем ли данный исход при заданной угрозе. При этом, как можно видеть, в уравнениях (4) отсутствуют стимулирования, но они могут быть определены (хотя и неоднозначно) с помощью (5). При этом условия упрощаются, и становится возможным их графический анализ (см. рис. 2).

На этом рисунке при фиксированном $x^*$ по оси абсцисс отложена величина угрозы, графики соответствуют функциям, фигурирующим в первых двух неравенствах (4). Жирными линиями отмечен максимум (левые части уравнений), правым частям соответствуют горизонтальные линии. Для того чтобы определить, при каких угрозах может быть получен исход $x^*$, необходимо найти, где соответствующая правой части горизонтальная прямая проходит выше соответствующей жирной линии и взять пересечение полученных множеств для всех трех уравнений. Заметим, что на рисунке не отображены функции, соответствующие третьему уравнению, но это делается аналогично (единственно – вместо горизонтальной прямой будет наклонная прямая, соответствующая $H_1(x^*) + H_2(x^*) - c(x^*) - s$).

При увеличении $s$ нельзя точно сказать, как ведет себя множество реализуемых исходов с угрозой $s$, поскольку в неравенствах (4) при $a_i - s > f_i(s)$ соответствующий максимум уменьшается, а при $a_i - s < f_i(s)$ соответствующий максимум не уменьшается (может увеличиваться). Соответственно если максимумы уменьшаются, то множество возможных $x^*$, удовлетворяющих (4), увеличивается, в противном случае уменьшается. При различных неравенствах (для первого и второго центров) эффекты действуют в различные стороны, и результат их действия (как изменится множество реализуемых с данной угрозой исходов) указать нельзя.

После этих рассуждений естественно задать вопрос: а всегда ли в системе с двумя центрами существует равновесие, исход которого Парето-эффективен? Положительный ответ на данный вопрос дает следующая

*Теорема 3. В АС с двумя центрами и одним АЭ для любого*

$$x^* \in \operatorname*{Argmax}_{x \in X} \left( \sum_{i=1}^{2} H_i(x) - c(x) \right)$$

*существуют такие функции стимулирования $\sigma_i(x)$, $i = 1, 2$, что тройка $(\sigma_1(x), \sigma_2(x), x^*)$ является равновесием Нэша, т.е. любой Парето-эффективный исход реализуем.*

Заметим, что в силу теоремы 1 достаточно предполагать, что функции стимулирования являются двухпиковыми.

Доказательство теоремы 3 приведено в Приложении.

Заметим, что в теореме не только доказано, что в системе из двух элементов всегда реализуем Парето-эффективный исход, но и для большинства случаев перечислены возможные равновесия (с точностью до места расположения угроз, которые, в принципе, тоже можно определить как произвольные точки некоторых множеств), причем выигрыш каждого из центров легко найти.

Однако хотелось бы знать, как связаны выигрыши в равновесии с возможными выигрышами в Парето-эффективном равновесии.

Теорема 4. *Пусть в АС с двумя центрами и одним АЭ имеется $(\sigma_1(x), \sigma_2(x), \widetilde{x})$ – некоторое равновесие Нэша,*

$$(6) \qquad \widetilde{x} \notin \operatorname*{Argmax}_{x \in X} \left( \sum_{i=1}^{2} H_i(x) - c(x) \right).$$

*Тогда для любого*

$$(7) \qquad x^* \in \operatorname*{Argmax}_{x \in X} \left( \sum_{i=1}^{2} H_i(x) - c(x) \right)$$

*существуют такие $\sigma_1^*(x)$ и $\sigma_2^*(x)$, что $(\sigma_1^*(x), \sigma_2^*(x), x^*)$ – равновесие Нэша, причем при переходе от одного равновесия к другому выигрыши центров не уменьшатся (и по крайней мере одного центра увеличатся), а выигрыш АЭ останется прежним.*

Доказательство теоремы 4 приведено в Приложении.

Таким образом, для любого равновесия в системе с двумя центрами существует Парето-эффективное равновесие, в котором выигрыш АЭ не изменится, а выигрыши центров не уменьшатся (при соответствующем дележе строго увеличатся).

И если мы решаем, сколько центров должно быть в системе (два или один), распределяя весь доход от функционирования системы между центрами, то необходимо учитывать следующее:

1) при двух центрах возможен более богатый спектр получающихся равновесий;

2) Парето-эффективный исход может быть реализован в любом случае;

3) при двух центрах возможно такое распределение дохода от деятельности системы между центрами, что при любом равновесии АЭ будет получать ненулевую прибыль (значение его целевой функции будет больше нуля);

4) при двух центрах для любого равновесия существует Парето-эффективное равновесие, которое доминирует (слабо) над первоначальным равновесием.

## 6. Заключение

В теории АС, при большом внимании к АС с одним управляющим центром, уделялось мало внимания системам, в которых присутствует несколько управляющих центров, несмотря на то, что примеры таких систем встречаются достаточно часто. В силу более сложной структуры и анализ таких систем является более сложным, хотя бы из-за бóльшего разнообразия получающихся равновесий.

В данной работе исследованы вопросы равновесия в АС с несколькими управляющими центрами. Показано, что исход, реализуемый в АС с одним центром, достижим в АС с несколькими центрами как равновесие Нэша, если суммарные доходы центров в обеих системах равны (но во втором случае просто делятся на несколько присутствующих в системе центров). Подробно исследована АС с двумя центрами, показано, что в такой системе всегда существует равновесие Нэша, реализующее Парето-эффективный исход.

Перспективным направлением для дальнейших исследований является вопрос о свойствах АС с произвольным числом и реализуемости в ней Парето-эффективных исходов. Важным представляется вопрос об определении на основании данной модели оптимального количества центров при возможности влиять на распределение дохода АС между центрами или при равномерном распределении дохода между центрами.

## *ПРИЛОЖЕНИЕ*

*Доказательство теоремы 2.* В силу теоремы 1 достаточно рассматривать только функции стимулирования с конечным числом пиков. Обозначим подмножество множества $X$, на котором расположены угрозы, через $Y$. Тогда множество $Y$ (заметим, что $x^* \in Y$) является конечным множеством. Оставим во множестве $Y$ только те точки $x$, в которых

$$\sum_{i=1}^{n} \sigma_i(x) - c(x) = \sum_{i=1}^{n} \sigma_i(x^*) - c(x^*)$$

(т.е. те точки, в которых угрозы имеют смысл, или, что то же самое, величина угрозы равна суммарной переплате АЭ). Функции стимулирования, равные нулю всюду, кроме точек множества $Y$, где они не изменили своих значений, по-прежнему являются равновесными, причем выигрыши центров и АЭ не изменятся.

Пусть утверждение теоремы неверно. Тогда существует такое $p$, что во всех точках угроз центр с номером $p$ предлагает АЭ ненулевое стимулирование (в том числе и в равновесной точке $x^*$, иначе в качестве $\widetilde{x}$ в уравнении (2) можно было бы взять $x^*$). Кроме того, в силу наличия угрозы АЭ получает больше, чем ему надо просто для покрытия затрат на выбор $x^*$. Но тогда, уменьшив во всех точках множества $Y$ стимулирование $p$-го центра на сумму

$$\delta = \min\left(\min_{y\in Y}\left[\sigma_p(y), \sum_{i=1}^{n} \sigma_i(x^*) - c(x^*)\right]\right) > 0,$$

мы получим, что реализуется по-прежнему тот же исход, поскольку

$$\sum_{i=1}^{n} \sigma_i(x^*) - c(x^*) - \delta \geq \sum_{i=1}^{n} \sigma_i(x) - c(x) - \delta I_{\{Y\}}(x),$$

где

$$I_{\{Y\}}(x) = \begin{cases} 1, & x \in Y; \\ 0, & x \notin Y. \end{cases}$$

Однако поскольку функции стимулирования других центров не изменились, а целевая функция $p$-го центра увеличилась (так как мы уменьшили его затраты на стимулирование), то мы имели дело не с равновесием. Противоречие.

*Доказательство следствия 1.* В равновесии против каждого из центров должна существовать коалиция (которая ограничена множеством всех центров без данного), способная обеспечить соответствующую угрозу. Но коалиция против $p$-го центра не может обеспечить угрозу больше, чем

$$\max_{x\in X}\left(\sum_{i=1,i\neq p}^{n} H_i(x)-c(x)\right)$$

для любого $p$, о чем и говорит следствие.

*Доказательство следствия 2.* В качестве точек множества $Y$ необходимо взять точки, в которых против $p$-го центра остальные центры образуют коалицию:

$$Y=\bigcup_{p=1}^{n} x_p\cup x^*,\quad x_p\in \operatorname*{Argmax}_{x\in X}\left(\sum_{i\neq p}\sigma_i(x)-c(x)\right)$$

(заметим, что выбранные $x_p$ есть точки, о которых говорится в теореме 2).

Тогда по теореме 2 $\sigma_p(x_p)=0$, $\sum\limits_{i\neq p}\sigma_i(x_p)-c(x_p)=\sum\limits_{i}\sigma_i(x^*)-c(x^*)$. Отклоняться же от своих стратегий центрам невыгодно в силу наличия противостоящих коалиций. Таким образом, следствие доказано.

*Доказательство теоремы 3.* Возьмем $x_1$, $x_2$ и $x^*$ такие, что

$$x_1\in\operatorname*{Argmax}_{x\in X}(H_1(x)-c(x));\ a_1=H_1(x_1)-c(x_1);$$

$$x_2\in\operatorname*{Argmax}_{x\in X}(H_2(x)-c(x));\ a_2=H_2(x_2)-c(x_2);$$

$$a=\sum_{i=1}^{2}H_i(x^*)-c(x^*)$$

(предполагаем, что $x_1\neq x_2$, $x_1\neq x^*$, $x_2\neq x^*$).

В силу определений и неотрицательности функций $H_1(x)$, $H_2(x)$ и $c(x)$ всегда выполняются неравенства $a_i\geq 0$, $a\geq 0$ и $a\geq a_i$.

Нетрудно видеть, что если первому центру невыгодно отклоняться с $x^*$ на $x_1$ или $x_2$, то ему вообще никуда больше невыгодно отклоняться (так как возможный максимальный выигрыш будет именно в этих точках). То же самое верно и для второго центра. Таким образом, для утверждения теоремы мы можем указать соответствующие стратегии и проверить, что центры не будут изменять свои стратегии так, чтобы в итоге реализовались $x_1$ или $x_2$.

*Случай 1.* $a_1=0$: в одиночку первый центр ничего не может получить.

Тогда при

$$\sigma_1(x)=\begin{cases} y, & x=x^*;\\ 0, & x\neq x^*,\end{cases}\qquad \sigma_2(x)=\begin{cases} c(x^*)-y, & x=x^*;\\ 0, & x\neq x^*,\end{cases}$$

где

$$y\in\left[H_2(x_2)-c(x_2)-(H_2(x^*)-c(x^*)),H_1(x^*)\right],$$

тройка $(\sigma_1(x),\sigma_2(x),x^*)$ есть Парето-эффективное равновесие Нэша (равновесие типа "сотрудничество"). Так как второму центру переплачивать АЭ смысла нет (первый просто не может угрожать), то мы нашли все Парето-эффективные равновесия Нэша, реализующие исход $x^*$.

*Случай 2.* $a_1 = a$: первый центр в одиночку может получить столько же, сколько и оба центра, объединившись вместе.

Тогда при функциях стимулирования

$$\sigma_1(x) = \begin{cases} c(x^*) + H_2(x_2) - c(x_2) - H_2(x^*), & x = x^*; \\ c(x_1) + H_2(x_2) - c(x_2), & x = x_1; \\ 0, & x \notin \{x^*, x_1\}, \end{cases}$$

$$\sigma_2(x) = \begin{cases} H_2(x^*), & x = x^*; \\ H_2(x_2), & x = x_2; \\ 0, & x \notin \{x^*, x_2\} \end{cases}$$

тройка $(\sigma_1(x), \sigma_2(x), x^*)$ есть Парето-эффективное равновесие Нэша (равновесие типа "конкуренция"). Второй центр при этом ничего не получает, а АЭ получает $a_2$.

*Случай 3.* $a_1 + a_2 \leq a$, $a_i < a$: сумма возможных выигрышей первого и второго центров не больше выигрыша центров при объединении.

При функциях стимулирования

$$\sigma_1(x) = \begin{cases} y, & x = x^*; \\ 0, & x \neq x^*, \end{cases} \qquad \sigma_2(x) = \begin{cases} c(x^*) - y, & x = x^*; \\ 0, & x \neq x^*, \end{cases}$$

где $y \in [a_2 + c(x^*) - H_2(x^*), H_1(x^*) - a_1]$, реализуется Парето-оптимальное равновесие Нэша $(\sigma_1(x), \sigma_2(x), x^*)$ (равновесие типа "сотрудничество"), поскольку выполняются неравенства:

$$(H_1(x^*) - a_1) - (a_2 + c(x^*) - H_2(x^*)) = a - a_1 - a_2 \geq 0$$

(множество возможных значений $y$ непусто),

$$H_1(x^*) - \sigma_1(x^*) = H_1(x^*) - y \geq H_1(x^*) - (H_1(x^*) - a_1) = a_1$$

(первый центр может так стимулировать),

$$H_2(x^*) - \sigma_2(x^*) = H_2(x^*) - (c(x^*) - y) \geq$$
$$\geq H_2(x^*) - c(x^*) + a_2 + c(x^*) - H_2(x^*) = a_2$$

(второй центр может так стимулировать),

$$y \geq a_2 + c(x^*) - H_2(x^*) \geq H_2(x^*) - c(x^*) - (H_2(x^*) - c(x^*)) = 0$$

(стимулирование первого центра неотрицательно),

$$c(x^*) - y \geq c(x^*) - (H_1(x^*) - a_1) \geq 0$$

(стимулирование второго центра неотрицательно).

В данном случае переплачивать не имеет смысла, так как по отдельности (при отклонении от этих стратегий) они получают не больше и угрозы не нужны.

*Случай 4.* $a < a_1 + a_2$, $a_i < a$: сумма возможных выигрышей первого и второго центров строго больше выигрыша центров при объединении.

Прежде всего заметим, что

$$a_1 > a - a2 = H_1(x^*) + H_2(x^*) - c(x^*) - a_2 \geq$$
$$\geq H_1(x_2) + H_2(x_2) - c(x_2) - a_2 = H_1(x_2);$$
$$H_1(x^*) = H_1(x^*) + H_2(x^*) - c(x^*) - (H_2(x^*) - c(x^*)) \geq$$
$$\geq H_1(x_2) + H_2(x_2) - c(x_2) - (H_2(x^*) - c(x^*)) = H_1(x_2);$$
$$a - a_1 = H_1(x^*) + H_2(x^*) - c(x^*) - a_1 \geq$$
$$\geq H_2(x_1) + H_2(x_1) - c(x_1) - a_1 = H_2(x_1),$$

таким образом (проведя аналогичные вычисления для второго центра), получаем

$$a_1 > H_1(x_2); \quad a_2 > H_2(x_1);$$
$$H_1(x^*) \geq H_1(x_2); \quad H_2(x^*) \geq H_2(x_1);$$
$$H_2(x_1) \leq a - a_1; \quad H_1(x_2) \leq a - a_2.$$

Для возможности равновесия с исходом $x^*$ и угрозой $s$ должны выполняться неравенства

$$a - s \geq H_2(x_1) + H_1(x_2) \quad \text{и} \quad a - s \geq (a_1 - s) + (a_2 - s),$$

что говорит о том, что угроза должна принадлежать отрезку

$$[a_1 + a_2 - a, a - (H_2(x_1) + H_1(x_2))].$$

Этот отрезок непуст, поскольку

$$(a - (H_2(x_1) + H_1(x_2))) - (a_1 + a_2 - a)$$
$$(a - a_1 - H_2(x_1)) + (a - a_2 - H_1(x_2)) \geq 0.$$

Исходя из сказанного, подберем подходящее значение для $s$. Пусть $s = a_1 + a_2 - -a$. Тогда на основании последнего неравенства системы (4) должно выполняться неравенство

$$a - s \geq \max(H_1(x_2), a_1 - s) + \max(H_2(x_1), a_2 - s).$$

Проверим это:

$$\max(H_1(x_2), a_1 - s) + \max(H_2(x_1), a_2 - s) =$$
$$= \max(H_1(x_2), a - a_2) + \max(H_2(x_1), a - a_1) =$$
$$= a - a_2 + a - a_1 \leq a < a_1 + a_2.$$

Кроме того, такая угроза обоими центрами реализуема, поскольку

$$a_1 - s = a_1 + a - a_1 - a_2 = a - a_2 \geq 0;$$
$$a_1 - s = a_1 + a - a_1 - a_2 = a - a_2 \geq 0,$$

что и завершает доказательство теоремы.

*Доказательство теоремы 4.* Из условия (7) следует, что для любого $x$ (и, в частности, для $\widetilde{x}$) выполняется

$$(H_1(x^*) + H_2(x^*) - c(x^*)) - (H_1(x) + H_2(x) - c(x)) \geq 0.$$

Из того, что $(\sigma_1(x), \sigma_2(x), \widetilde{x})$ – равновесие Нэша, следует, что никакому из центров невыгодно переключаться на реализацию $x^*$, т.е.

$$H_i(x^*) - c(x^*) - (\sigma_1(\widetilde{x}) + \sigma_2(\widetilde{x}) - c(\widetilde{x})) \leq H_i(\tilde{x}) - \sigma_i(\tilde{x}), \text{ или}$$
$$H_i(\widetilde{x}) - H_i(x^*) + c(x^*) + \sigma_{-i}(\widetilde{x}) - c(\widetilde{x}) \geq 0.$$

Для доказательства теоремы достаточно показать, что мы можем так изменить $\sigma_1(x)$ и $\sigma_1(x)$ в точке $x^*$, что новая система реализуема, является равновесием Нэша и выигрыши центров при этом не уменьшатся (АЭ должен получить не меньше, так как иначе ему будет невыгодно выбирать $x^*$).

Прежде всего, заметим, что по сравнению с $\widetilde{x}$ появляется дополнительная сумма для дележа в размере

$$d = (H_1(x^*) + H_2(x^*) - c(x^*)) - (H_1(\widetilde{x}) + H_2(\widetilde{x}) - c(\widetilde{x})) > 0.$$

Ее центры и будут делить друг с другом.

Зададим новые функции стимулирования центров как

$$\sigma_1^*(x) = \begin{cases} H_1(x^*) - H_1(\widetilde{x}) + \sigma_1(\widetilde{x}) - y, & x = x^*; \\ \sigma_1(x), & x \neq x^*, \end{cases}$$

$$\sigma_2^*(x) = \begin{cases} \sigma_2(\widetilde{x}) - c(\widetilde{x}) + c(x^*) - H_1(x^*) + H_1(\widetilde{x}) + y, & x = x^*; \\ \sigma_2(x), & x \neq x^*, \end{cases}$$

где переменная $y \in Y = [0, d]$. Заметим, что в силу выбора точки $x^*$ множество $Y$ непусто ($d > 0$). Вторая функция стимулирования подбиралась из того условия, что АЭ должен получить столько же, сколько и раньше. Легко проверить, что при всех допустимых значениях $y$ выполняется $H_i(x^*) - \sigma_i^*(x^*) \geq H_i(\widetilde{x}) - \sigma_i^*(\widetilde{x})$ (из чего, в частности, следует неотрицательность левой части неравенства и, как следствие, неравенство $H_i(x^*) \geq \sigma_i^*(x^*)$), т.е. центрам невыгодно отклоняться для того, чтобы АЭ выбирал $\widetilde{x}$. Но тогда в силу определения $\sigma_i(x)$ и того, что $\sigma_i(x)$ и $\sigma_i^*(x)$ совпадают всюду, кроме точки $x^*$, центрам вообще никуда невыгодно отклоняться.

Теперь остается только проверить, что необходимое стимулирование $\sigma_i^*(x^*)$ не меньше нуля, т.е. функции стимулирования являются допустимыми. Заметим, что суммарный выигрыш центров в точке $x^*$ не меньше, чем в точке $\widetilde{x}$. Найдем стимулирование первого центра при максимальном $y$ (в этом случае само стимулирование минимально):

$$\begin{aligned} \sigma_1^*(x^*) &= H_1(x^*) - H_2(\widetilde{x}) + \sigma_1^*(\widetilde{x}) - \\ &-(H_1(x^*) + H_2(x^*) - c(x^*) - (H_1(\widetilde{x}) + H_2(\widetilde{x}) - c(\widetilde{x}))) = \\ &= H_2(\widetilde{x}) - H_2(x^*) + c(x^*) + \sigma_1(\widetilde{x}) - c(\widetilde{x}) \geq 0, \end{aligned}$$

так как центрам невыгодно самостоятельно отклоняться для реализации $x^*$. Аналогично доказывается для второго центра.

Для полноты необходимо заметить, что в случае отсутствия угроз доказательство остается таким же. Также необходимо заметить, что ни один из центров не мог угрожать другому точкой $x^*$, так как второму центру тогда было бы выгодно переключиться именно на реализацию этого исхода и $(\sigma_1(x), \sigma_2(x), \widetilde{x})$ не было бы равновесием Нэша.

## СПИСОК ЛИТЕРАТУРЫ


1. *Бурков В.Н., Кондратьев В.В.* Механизмы функционирования организационных систем. М.: Наука, 1981.
2. *Новиков Д.А., Петраков С.Н.* Курс теории активных систем. М.: СИНТЕГ, 1999.
3. *Губко М.В., Караваев А.П.* Согласование интересов в матричных структурах управления // АиТ. 2001. № 10. С. 132–147.
4. *Новиков Д.А., Цветков А.В.* Механизмы функционирования организационных систем с распределенным контролем. М.: ИПУ РАН, 2001.
5. *Bernard Salanie.* The Economics of Contracts: A Primer. The MIT Press, 2000.
6. *Mas-Colell A., Whinston M., Jerry R. Green.* Microeconomic Theory. Oxford University Press, 1995.
7. *Maskin E., Tirole J.* The Principal-Agent Relationship with an Informed Principal, I: The Case of Private Values // Econometrica. 1990. V. 58 (2). P. 379–409.



*Статья представлена к публикации членом редколлегии Д.А. Новиковым.*

Поступила в редакцию 28.05.2002