УДК 004.75
ББК 32.973.202

ОПТИМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ НА ОСНОВЕ ДВУХУРОВНЕВОГО КРИТЕРИЯ

Козлов С. В.[1], Остриков Ю. П.[2], Суханов А. Л.[3]
(*Академия Федеральной службы охраны Российской Федерации, Орел*)

В статье рассматривается модель оптимального распределения ресурсов в распределенной системе обработки информации на основе двухуровневого векторного критерия оптимизации.

Ключевые слова: модель «ресурс–потребитель», векторная оптимизация, суперконкурентное распределение ресурса, лексикографический максимин.

1. Введение

Одним из направлений развития распределенных систем обработки информации (РСОИ) является совершенствование механизмов управления разнородными ресурсами (процессорами, долговременной и оперативной памятью, хранилищами и базами данных, сетями) с целью достижения высокой производительности и устойчивости распределенных приложений. Современные технологии построения РСОИ (Grid, CORBA, J2EE, РСУБД) позволяют реализовать коллективный разделяемый режим доступа к ресурсам и к связанным с ними услугам в рамках географически распределенной инфраструктуры [5, 7].

[1] *Сергей Викторович Козлов, кандидат технических наук* (*kozlov_sv@mail.ru*).
[2] *Юрий Петрович Остриков, кандидат технических наук, старший научный сотрудник* (*тел.* (*4862*) *54-97-32*).
[3] *Аркадий Леонидович Суханов, кандидат технических наук* (*als64@bk.ru*).

В условиях жестких требований к качеству предоставляемых услуг на первый план выступает необходимость оптимального использования ограниченных ресурсов РСОИ с учетом изменяющихся требований к ним и ненадежности элементов инфраструктуры.

Применяемые в настоящее время в РСОИ методы и алгоритмы балансировки нагрузки, как правило, имеют целью достижение равномерной загрузки ресурсов и снижение перегрузок в РСОИ. При этом индивидуальные требования к качеству обслуживания пользователей либо не учитываются, либо учитываются косвенно с использованием механизма приоритетов в обслуживании, чего явно не достаточно [5]. Более продуктивным целесообразно считать подход, основанный на векторной оптимизации распределения ресурсов РСОИ, где в качестве частных критериев рассматривается максимизация степени достижения требований к качеству обслуживания каждого пользователя [2, 3]. Это требует разработки соответствующих моделей и методов оптимизации.

2. Задача оптимального распределения ресурсов

Представим РСОИ в виде модели «ресурс–потребитель» (РП-модели) $S = \langle R, P, A\rangle$ (рис. 1).

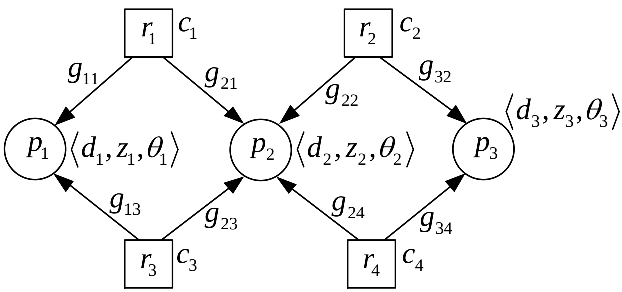

Рис. 1. Модель «ресурс–потребитель»

РП-модель содержит вершины двух типов: $R = \{r_1, \ldots, r_e\}$ – множество поставщиков ресурсов РСОИ, $P = \{p_1, \ldots, p_m\}$ – множество потребителей ресурсов. Будем считать, что все поставщики предоставляют пользователям ресурс одного типа. Чтобы учесть разнотипность ресурса, требуется определить отдельную модель S для каждого типа ресурса.

Каждый поставщик r_k может предоставить потребителям ресурс суммарным объемом не более c_k. Предположим, что весь ресурс поставщика может делиться любыми долями на любое число потребителей. Единицы измерения c_k должны соответствовать типу ресурса (таблица 1). Будем считать вектор $\overline{C} = \{c_1, \ldots, c_e\}$ известным и строго положительным.

Таблица 1. Величины и единицы измерения ресурсов РСОИ

Тип ресурса	Величина	Единица измерения
Сеть передачи данных	Пропускная способность	бит/с; число каналов
Системы распределенных вычислений	Вычислительная мощность	операций/с
Файловые серверы и сетевые хранилища	Объем памяти или дискового пространства	байт
Веб-серверы	Максимальный обслуживаемый поток запросов	запросов/с
Серверы баз данных	Скорость выполнения транзакций	транзакций/с

Для каждого потребителя p_i известна количественная мера его ресурсных требований d_i, которые измеряются в тех же величинах, что и объем поставляемого ресурса c_k. Таким образом, задается известный положительный вектор ресурсных требований $\overline{d} = \{d_1, \ldots, d_m\}$.

Для удовлетворения своих потребностей каждый потребитель может использовать ресурс одного или нескольких поставщиков одновременно, т.е. диверсифицировать поставки. Структура поставок определяется матрицей связности $A = (a_{ik})$, где $a_{ik} = 1$, если i-й потребитель может использовать ресурс k-го поставщика, и $a_{ik} = 0$ в противном случае. Матрица A предполагается известной.

Обозначим g_{ik} величину ресурса, предоставляемого k-м поставщиком i-му потребителю. Матрица $G = (g_{ik})$, $g_{ik} \geq 0$, задает распределение ресурсов РСОИ. Каждое распределение G однозначно определяет вектор обеспеченных ресурсных требований $\bar{z} = \{z_1, \dots, z_m\}$, где $z_i = \sum_{k=1}^{e} g_{ik}$. Очевидно, что для рациональных распределений должно выполняться условие $z_i \leq d_i$, что соответствует принципу разумной достаточности: совокупный объем ресурса, полученный каждым потребителем, не может превышать его потребностей. Кроме того, объем ресурса, отданный каждым поставщиком, не может превышать имеющегося у него объема c_k. Таким образом, допустимое распределение G должно удовлетворять ограничениям

$$(1)\quad \begin{cases} \sum_{k=1}^{e} g_{ik} \leq d_i, & i = 1,2,\dots,m, \\ \sum_{i=1}^{m} g_{ik} \leq c_k, & k = 1,2,\dots,e. \end{cases}$$

Задача оптимального распределения ресурсов заключается в нахождении такого распределения G^*, которое в наиболее полной мере удовлетворяло бы ресурсным требованиям $\bar{d}$. Степень обеспеченности ресурсных требований (о.р.т.) для i-го потребителя будем определять коэффициентом о.р.т.:

$$(2)\quad K_i^{\text{орт}} = \frac{z_i}{d_i}.$$

При этом $0 \leq K_i^{\text{орт}} \leq 1$. Обеспеченность ресурсных требований для всей РСОИ оценивается вектором о.р.т.:

$$(3)\quad \bar{K}_{\text{орт}} = \left(\frac{z_1}{d_1}, \dots, \frac{z_m}{d_m} \right).$$

Таким образом, рассматриваемая задача оптимизации может быть представлена в виде

$$(4)\quad G^* = \arg\max_{G \in \{G\}} \left\{ \bar{K}_{\text{орт}} \right\},$$

где $\{G\}$ – множество всех распределений ресурсов, удовлетворяющих ограничениям (1). Максимальное значение $\overline{K}_{\text{орт}}$ обозначим $\overline{K}^{*}_{\text{орт}}$.

Условия работы РСОИ таковы, что в ней могут происходить флуктуации векторов ресурсов $\overline{c}$ и ресурсных требований $\overline{d}$. При этом вектор $\overline{z}$ отклоняется от вектора $\overline{d}$ при ухудшении ситуации и приближается при её улучшении. Степень такого отклонения оценивается вектором $\overline{K}_{\text{орт}}$. Задачей системы распределения ресурсов является подстройка матрицы G под изменившуюся ситуацию путем решения оптимизационной задачи (4) (рис. 2).

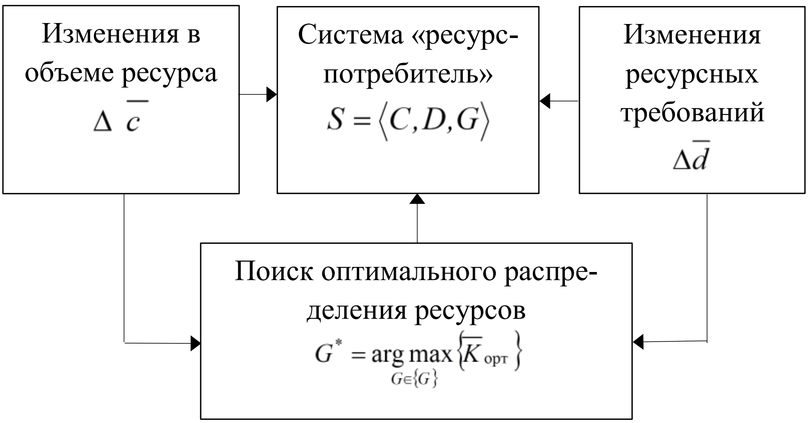

Рис. 2. Схема распределения ресурсов РСОИ

3. Исследование модели оптимального распределения ресурсов

Существование допустимых распределений ресурсов следует из совместности системы линейных неравенств (1). Для доказательства совместности представим систему (1) в обобщенном виде (5).

$$(5) \quad \sum_{j} g_j \le a_i, g_j \ge 0, \quad i = 1,\dots,m, m+1,\dots,m+e.$$

Поскольку все свободные члены a_i неотрицательны (в силу положительности векторов $\bar{c}$ и $\bar{d}$), то неравенства (5) всегда будут выполняться при $g_j = 0$. Следовательно, система (1) имеет по крайней мере одно тривиальное решение. Покажем, что она имеет и нетривиальные решения. Для этого предположим, что все свободные члены a_i строго положительны. Те неравенства, для которых $a_i = 0$, из системы можно исключить, поскольку они дают только тривиальные решения. Пусть $a_{\min} = \min\{a_i\}$. Тогда соответствующее неравенство системы (5) можно преобразовать в равенство (6) путем добавления свободного члена q:

(6) $$\sum_j g_j + q_i = a_{\min}, \ 0 \le q \le a_{\min}.$$

Следовательно, любое из g_j может быть увеличено на величину q, так что все неравенства (5) будут выполняться, причем одно из них будет строгим. Это значит, что система (1) имеет как минимум одно нетривиальное решение. Следовательно, данная система неравенств совместна, а определяемое ей множество допустимых распределений $\{G\}$ не пусто.

Из совместности системы (1) следует, что множество ее решений представляет собой выпуклый ограниченный многогранник в пространстве R^n с осями координат, соответствующими всем g_j [6]. Каждой точке многогранника решений соответствует допустимое распределение G. Многогранник решений ограничен снизу координатными плоскостями, а сверху – гиперплоскостями, описываемыми уравнениями $\sum_j g_j = a_i$. Распределение ресурсов G^*, оптимальное по векторному критерию (4), обеспечивает максимум по всем $\overline{K_i}^{\text{орт}}$ одновременно. Если $\overline{K}^*{}_{\text{орт}} = 1$, то ресурса РСОИ достаточно для удовлетворения требований потребителей, в противном случае выполнить эти требования путем перераспределения ресурсов невозможно. Такая ситуация может возникнуть, например, при выходе из строя элементов РСОИ.

Если распределение G таково, что улучшение по одному из $K_i^{\text{орт}}$ невозможно без ухудшения $K_i^{\text{орт}}$ других потребителей, то G – распределение, оптимальное по Парето [4]. Множество распределений, не доминирующих друг над другом по вектору о.р.т., составляет множество Парето-оптимальных распределений. Очевидно, что область Парето-оптимальных распределений представляет собой поверхность выпуклого многогранника решений системы (1), не совпадающую с координатными плоскостями.

Рассмотрим РП-модель в условиях недостаточности ресурсов, когда значение $\overline{K}^{*}_{\text{орт}} = 1$ недостижимо. Такая ситуация может возникнуть в результате повреждений элементов РСОИ или резкого увеличения ресурсных требований. Возникающий при этом дефицит ресурсов приводит к росту конкуренции потребителей. В этом случае становится актуальным распределение ресурсов на основе некоторого компромисса, а свертка векторного критерия (4) должна быть такой, чтобы никто из потребителей не был дискриминирован и при этом были использованы все возможные ресурсы РСОИ, пока они могут пригодиться хотя бы одному потребителю. Очевидно, что компромиссное решение должно принадлежать области Парето.

Одним из компромиссных методов свертки векторного критерия является метод максимина, в основе которого лежит принцип гарантированного результата. Целесообразность его применения в задачах векторной оптимизации с равнозначными частными критериями показана в работах [3, 4]. Исходя из этого исходная оптимизационная задача (4) может быть представлена в виде

(7) $$G^{*} = \arg\max_{G \in \{G\}} \min\left(K_i^{\text{орт}},\ i = 1,\dots,m\right).$$

Максиминный критерий (7) позволяет гарантировать максимальное значение коэффициента о.р.т. того потребителя, который находится в наихудших условиях в смысле уровня обеспеченности его ресурсных требований.

Распределение ресурса G^{*}, отвечающее критерию (7), называется конкурентным [3]. Соответствующее ему значение $K_i^{\text{орт}}$ называется величиной максиминной обеспеченности ресурсных требований. Будем обозначать его как θ_0.

Рассмотрим простейший пример конкурентного распределения ресурса – два потребителя, конкурирующие за разделяемый ресурс единственного поставщика (рис. 3а).

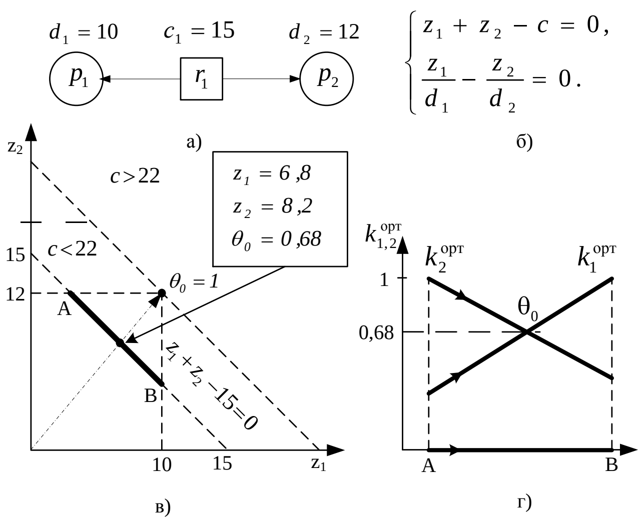

Рис. 3. Конкурентное распределение ресурсов одного поставщика двум потребителям

Очевидно, что максиминный уровень о.р.т. будет обеспечиваться при $K_1^{\text{орт}} = K_2^{\text{орт}}$. Для нахождения максиминного распределения ресурсов необходимо решить систему уравнений, представленную на рис. 3б. Область Парето-оптимальных распределений представлена отрезком AB (рис. 3в). Точка $\theta_0 = 0{,}68$, оптимальная по критерию (7), может быть найдена, если двигаться вдоль отрезка AB, начиная от одного из его концов, до тех пор, пока значения $K_1^{\text{орт}}$ и $K_2^{\text{орт}}$ не сравняются (рис. 3г).

Используя выводы, полученные в [4], в общем случае можно утверждать, что если $\theta_0 < 1$, то найдутся два или более потребителей, для которых выполняется соотношение $K_{i1}^{\text{орт}} = K_{i2}^{\text{орт}} = \ldots = K_{in}^{\text{орт}} = \theta_0$. Данное соотношение определяет узкое место («бутылочное горло») всей РП-модели. В узком месте каждый конкурирующий потребитель получит объем ресурса, пропорциональный его требованиям. Такой принцип справедливого распределения ограниченного ресурса соответствует известному принципу выравнивания Гермейера [1, 3].

Несмотря на то что максиминный критерий в наибольшей степени соответствует задачам недискриминирующего распределения ресурса РСИО, он не позволяет, в общем случае, получить Парето-оптимальных распределений, поскольку ориентирован на нижнюю границу о.р.т. [3], при которой всем потребителям обеспечивается θ_0-я часть их требований, тогда как для многих из них коэффициенты о.р.т. могли бы быть улучшены. Таким образом, невозможность обеспечить требования одних потребителей не способствует стремлению удовлетворить требования остальных. Тем не менее сама структура РП-модели ставит ее потребителей в неравные условия, учет которых позволяет выбрать из всех конкурентных распределений максимально обеспечивающее требования всех потребителей без дискриминации какой-либо из них. Такое распределение было предложено в работах [3] и названо суперконкурентным. Суперконкурентное распределение ресурсов определяется следующим образом. Обозначим через $X_0(c, d)$ множество конкурентных распределений потоков

$$X_0(c,d) = Arg\max_{G\in\{G\}}\left\{\min_{i\in M}\frac{z_i}{d_i}\right\}. \tag{8}$$

Пусть $M_0 \in M$ обозначает множество индексов i тех потребителей, которым нельзя увеличить $K_i^{\text{орт}}$ выше, не понизив при этом о.р.т. какого-либо потребителя ниже величины θ_0 (9).

$$M_0 = \left\{ i \in M \left| \frac{z_i}{d_i} = \theta_0, \forall G \in X_0(c,d) \right.\right\}. \tag{9}$$

Если $M_0 \neq M$, то существуют такие распределения ресурсов, которые являются конкурентными не только на нулевом, но и на более высоких уровнях обеспеченности требований $\theta_0, \theta_1, \ldots, \theta_L$. Множества конкурентных распределений более высоких уровней определяются рекурсивно (10).

$$X_{i+1}(c,d) = \underset{G\in X_i(c,d)}{Arg\max}\left\{\min_{i\in M\setminus\bigcup_{j=0}^{i}M_j}\frac{z_i}{d_i}\right\} = \left\{G\in X_i(c,d)\,\middle|\,\min_{i\in M\setminus\bigcup_{j=0}^{i}M_j}\frac{z_i}{d_i} = \theta_{i+1}\right\},$$

(10) $$\theta_{i+1} = \max_{G\in X_i(c,d)}\ \min_{i\in M\setminus\bigcup_{j=0}^{i}M_j \frac{z_i}{d_i}},$$

$$M_{i+1} = \left\{i\in M\setminus\bigcup_{j=0}^{i}M_j\,\middle|\,\frac{z_i}{d_i} = \theta_{i+1},\quad G\in X_{i+1}(c,d)\right\}.$$

Рекурсивное построение повторяется до тех пор, пока все потребители не будут исчерпаны. Очевидно, что количество уровней $L \le m$. Любое конкурентное распределение максимально возможного уровня называется суперконкурентным или лексикографически оптимальным [3]. Ему соответствует вектор $\overline{Z}^L = (z_1^L,\dots,z_m^L)$ с компонентами (11).

Ни один компонент z_i^L не может быть увеличен без уменьшения какого-либо другого компонента. Таким образом, суперконкурентное распределение является Парето-оптимальным. РП-модель с двумя уровнями суперконкурентного распределения ресурсов представлена на рис. 4.

(11) $$z_i^L = \begin{cases}\theta_0 d_i \,|\, i\in M_0,\\ \dots,\\ \theta_L d_i \,|\, i\in M_L.\end{cases}$$

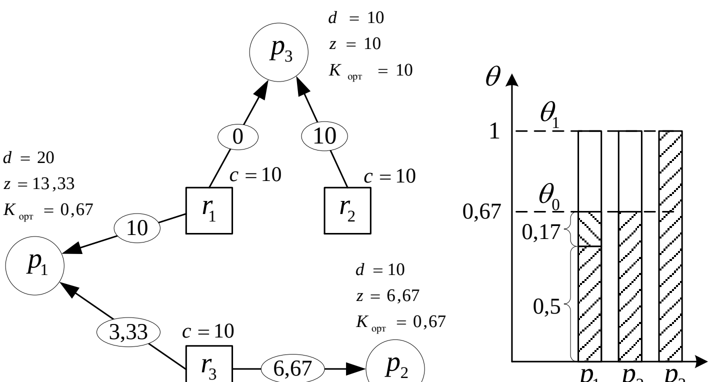

Рис. 4. Суперконкурентное распределение ресурсов при $\theta_0 < 1$

4. Обобщенный двухуровневый критерий оптимального распределения ресурсов

Если $\theta_L = 1$, то РП-модель имеет достаточно ресурса для полного обеспечения ресурсных требований одного или более (возможно, даже всех) потребителей. При этом ресурса может быть даже больше, чем требуется, так что часть его остается нераспределенной. Ни вектор, ни уровни о.р.т. не дают количественной оценки степени избыточности ресурса.

В условиях избытка ресурса для ряда потребителей суперконкурентное распределение может быть не единственным, а значит и решение задачи оптимизации будет неопределенным. Пример такой РП-модели представлен на рис. 4.

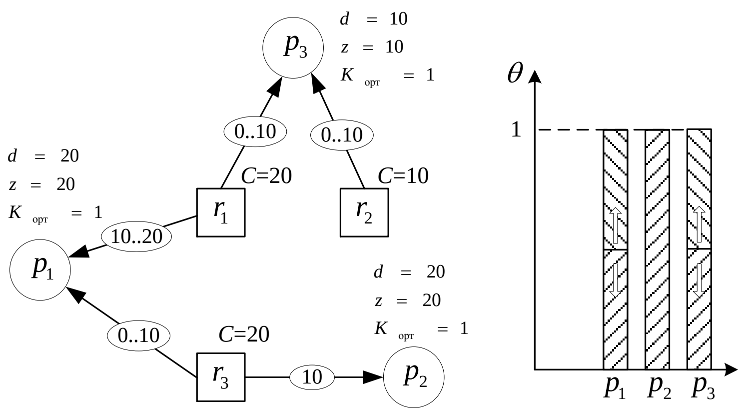

Рис. 4. Суперконкурентное распределение при избытке ресурсов

Решить проблему неоднозначности лексикографического максиминного критерия (10) для случая $\theta_L = 1$ возможно путем введения дополнительного критерия оптимальности, который бы учитывал приоритетность в выборе потребителем поставщиков ресурсов. Таким образом, целевая функция распределения ресурсов должна стремиться не только в максимальной степени удовлетворить все потоковые требования, но и сделать это наиболее предпочтительным для потребителей способом.

Каждому из поставщиков, предоставляющему ресурс i-му потребителю, припишем весовой коэффициент w_{in}. При этом веса должны строго возрастать в соответствии с порядком их предпочтения для потребителя: $W_{i1} > W_{i2} > \ldots > W_{iNi}$. Тогда, используя аддитивную свертку, получим критерий максимума средневзвешенного суммарного ресурса Z_W:

(12) $$Z_w = \max_{G \in \{G^*\}} \sum_{i \in M} \sum_{n \in N_i} w_{in} g_{in} .$$

С учетом критерия предпочтения (12) задача оптимизации распределения ресурсов может быть сформулирована в следующем виде:

(13) $$G^{**} = \arg\max_{G \in X_L(c,d)} \sum_{i \in M} \sum_{l \in L_i} w_{il} g_{il} ,$$

где $X_L(c, d)$ – множество суперконкурентных распределений, которое определяется рекурсивно по формуле (10); веса путей w_{in} задаются априорно исходя из некоторого критерия предпочтения.

Таким образом, задача (13) – двухкритериальная с приоритетом критерия. Приоритетный критерий – лексикографический максимин вектора о.р.т. (10).

Если полученное по первому критерию множество суперконкурентных распределений содержит более одного варианта, то из них выбирается максимальный относительно средневзвешенного суммарного ресурса Z_w (12).

5. Выводы и направления дальнейших исследований

Представленная в данной работе модель оптимального распределения ресурсов учитывает индивидуальные требования пользователей РСОИ к качеству обслуживания через ресурсные требования услуг и распределенных приложений. Поскольку пространство допустимых распределений по-разному проявляет свои свойства в условиях недостатка и в условиях избытка ресурсов, то целевая функция формируется на основе обобщенного критерия оптимизации, работающего и в тех и в других условиях.

В настоящей работе не рассмотрен ряд вопросов, требующих дополнительного исследования. Во-первых, требуется определить необходимые и достаточные условия существования и единственности решения рассмотренной задачи оптимизации по обобщенному критерию (13). Во-вторых, требуется обобщить свертку (12) для случая линейно упорядоченного множества предпочтений, чтобы сделать ее нечувствительной к выбору числовых значений действительных коэффициентов w_{in}. В-третьих, необходимо исследовать задачу оптимизации для дискретного ресурса.

Литература

1. ГЕРМЕЙЕР Ю.Б. *Введение в теорию исследования операций*. – М.: Наука, 1972. – 384 с.
2. КОЗЛОВ С.В. *Алгоритм оптимизации распределения потоков в телекоммуникационной сети на основе суперконкурентного резервирования ресурсов* // Системы управления и информационные технологии. –2012. – №3(49). – С. 42–46.
3. МАЛАШЕНКО Ю.Е., НОВИКОВА Н.М. *Модели неопределенности в многопользовательских сетях*. – М.: Эдиториал УРСС, 1999. – 160 с.
4. МАШУНИН Ю.К. *Методы и модели векторной оптимизации*. – М.: Наука, 1986. – 142 с.
5. ТАНЕНБАУМ Э. ВАН СТЕЕН М. *Распределенные системы. Принципы и парадигмы*. – СПб.: Питер, 2003. – 877 с.
6. ЧЕРНИКОВ С.Н. *Системы линейных неравенств* // Успехи математических наук. – 1953. – Т. 8, вып. 2(54). – С. 8–68.
7. TALIA D. *The Open Grid Services Architecture* // IEEE Internet Computing. – 2002. – November/December. – P. 124–132.

OPTIMAL RESOURCE ALLOCATION BASED ON TWO-LEVEL CRITERION

Sergey Kozlov, The Academy of Federal Security Guard Service of the Russian Federation, Orel, Cand.Sc. (kozlov_sv@mail.ru).
Uriy Ostrikov, The Academy of Federal Security Guard Service of the Russian Federation, Orel, Cand.Sc., ((4862)54-97-32).
Arkadiy Suhanov, The Academy of Federal Security Guard Service of the Russian Federation, Orel, Cand.Sc. (als64@bk.ru).

Abstract: *We discuss the model of optimal resource allocation in a distributed data processing system. A two-level vector optimization criterion is used for allocation.*

Keywords: resource-consumer model, vector optimization, superconcurrent resource allocation, lexicographical maxmin.

Статья представлена к публикации членом редакционной коллегии С.Г. Куливцом

Поступила в редакцию 04.09.2014.
Опубликована 31.03.2014.