

**сообщения
объединенного
института
ядерных
исследований
Дубна**

11-84-710

С.Г.Каданцев

**ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ СИСТЕМЫ
ДИСТАНЦИОННОГО ОБСЛУЖИВАНИЯ
НА БАЗЕ БЭСМ-6
И КОНЦЕНТРАТОРА ТЕРМИНАЛОВ**

1984

Цель работы

Создание системы коллективного пользования на базе БЭСМ-6 и концентратора терминалов ЕС-1010/1-5/ способствовало повышению эффективности разработки программ, обработки информации и использования средств вычислительной техники в ОИЯИ. В связи с этим актуальной стала задача оценки эффективности функционирования системы.

Согласно общепринятому определению, эффективность - качество системы, отражающее степень соответствия ее своему назначению, техническое совершенство и экономическую целесообразность^{/6/}. Количественно эффективность системы оценивается через показатели (или параметры, индексы, атрибуты) эффективности на основе критерия эффективности как функции множества этих показателей. Правильно выбранные критерии оценки эффективности могут успешно использоваться для выбора, усовершенствования и разработки вычислительных систем.

При оценке эффективности систем обычно выбирают показатели, характеризующие их целевое назначение. Это могут быть временные параметры, пропускная способность, экономические показатели и др.^{/7-9/}. Одним из главных критериев эффективности является производительность системы. Методы оценки производительности вычислительных систем наиболее полно освещены в работе^{/10/}.

Для разработчиков вычислительных систем вопрос об оценке производительности возникает в связи с планированием усовершенствований с целью достичь максимального роста производительности системы, проектированием варианта, наилучшим образом обеспечивающего выполнение некоторого круга работ и др. Нам, в частности, интересовали вопросы: каково среднее время ответа и его стандартное отклонение в системе коллективного пользования, как меняются эти показатели при изменении рабочей нагрузки. Получив с некоторой степенью точности ответы на эти вопросы, можно говорить о возможности расширения терминальной сети путем подключения дополнительных терминалов и/или концентраторов.

Введем определения^{/10/}:

Индексом производительности будем называть описатель, используемый для представления всей производительности системы или какого-то из ее аспектов. Некоторые индексы производительности могут иметь



численное значение. Сюда относятся, например, пропускная способность, скорость выполнения команд, время ответа, коэффициент использования оборудования. Другим, таким как: легкость использования системы, мощность набора команд, трудно или невозможно придать количественный вид. Здесь будем рассматривать индексы производительности, принимающие количественную форму.

Рабочей нагрузкой будем называть входные сигналы (программы, команды, данные), поступающие в систему.

Оценка производительности вычислительной системы, по-существу, состоит в получении информации о значениях индексов производительности при заданной рабочей нагрузке. Наиболее очевидный способ оценки производительности - измерение индексов производительности реально существующей системы на реальных задачах, - во-первых, является достаточно трудоемким, а, во-вторых, не дает ответа на вопрос об изменении значений индексов при изменении параметров рабочей нагрузки или самой системы. Более плодотворными являются методы моделирования, основные из которых - имитационный и аналитический. Если поведение модели во времени в основном воспроизводит поведение системы согласно некоторым условиям соответствия между различными элементами модели и системы, мы имеем имитационную модель. Когда решение уравнений, составляющих модель, получено математическими методами, мы говорим, что использован аналитический метод и построена аналитическая модель. Среди аналитических методов наиболее распространенным является метод очередей.

В работе представлены результаты измерений некоторых показателей производительности и результаты моделирования системы коллективного пользования на базе БЭСМ-6 и концентратора терминалов ЕС-1010. Мы не ставили задачу провести измерение и моделирование системы в целом, а ограничились рассмотрением работы БЭСМ-6 по обслуживанию запросов ЕС-1010 в процессе дистанционной пакетной и интерактивной обработки заданий. В частности, здесь не рассматриваются команды редактирования, управления локальными файлами и другие, выполняемые концентратором без обращения к БЭСМ-6 (результаты измерений времени выполнения этих команд представлены в [12]), а также вопросы оценки времени прохождения заданий пользователей в БЭСМ-6 в условиях мультипрограммной обработки.

Модель системы

Рассмотрим модель работы управляющей программы связи (УПС) БЭСМ-6 по организации дистанционной пакетной и интерактивной обработки заданий (рис.1).

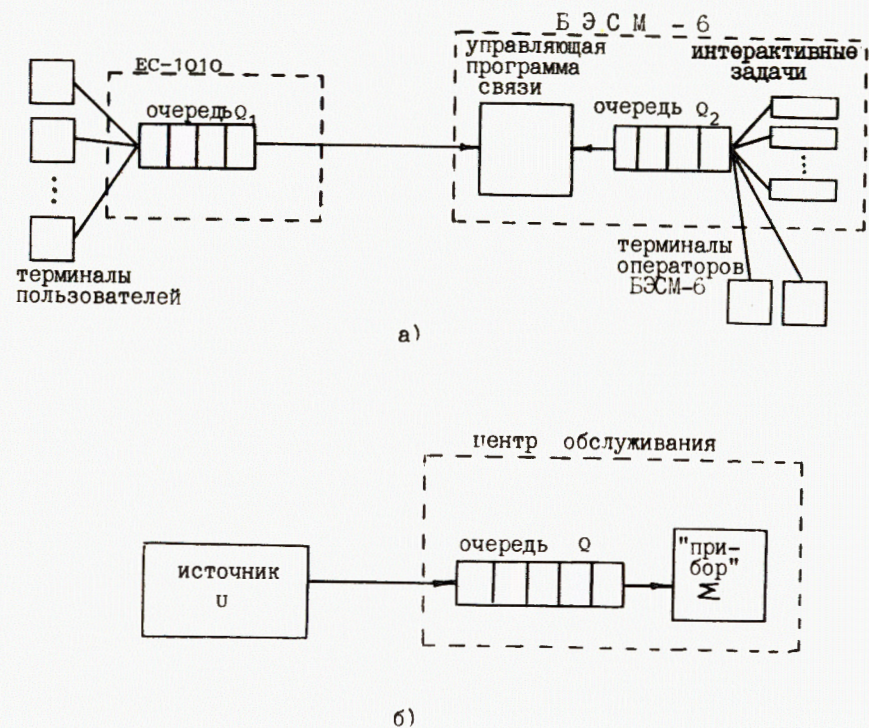


Рис.1. Модель системы дистанционного обслуживания.

УПС обслуживает запросы, поступающие от пользователей с терминалов, от интерактивных задач, связанных с терминалами, и от операторов ЭВМ. Все запросы пользователей можно разделить на две группы:

- (1) локальные: выполняемые на концентраторе без обращения к БЭСМ-6 (редактирование, управление локальными файлами и др.);
- (2) дистанционные: вызывающие обращение к БЭСМ-6.

К дистанционным относятся следующие запросы (в скобках - условное обозначение их типа):

1. Передача заданий на БЭСМ-6 (INPUT).
2. Передача листингов на концентратор (LOCAL).
3. Опрос состояния заданий (QUEUE).

4. Передача сообщений интерактивной задаче по требованию оператора READ (READ).

5. Передача сообщений операторам БЭСМ-6 (MES1).

6. Передача сообщений от интерактивной задачи на терминал по требованию оператора PRINT и выдача приглашения на ввод для оператора READ (PRINT).

7. Передача сообщений от операторов БЭСМ-6 на терминал (MES2).

Дистанционные запросы 1-5 помещаются ОС ЕС-1010 в очередь Q_1 , а запросы 6-7 - ОС БЭСМ-6 в очередь Q_2 (рис. I-a). Так как механизм обработки всех дистанционных запросов управляется единым аппаратом прерываний БЭСМ-6, можно считать, что существует одна очередь запросов Q от источника U к обслуживающему "прибору" Σ (рис. I-б), представляющему собой сложную совокупность программных и аппаратных средств БЭСМ-6, обрабатывающих запросы. В "прибор" входят центральный процессор, внешние запоминающие устройства, канал связи, УПС и другие компоненты ОС.

В нашей системе обработка запросов организована так: когда наступает очередь для обработки запроса, он вместе с необходимыми данными передается по каналу связи на БЭСМ-6 (если это не запрос типа 6 или 7, возникающий в самой БЭСМ-6), обрабатывается УПС, и выходная информация, если нужно, посылается на концентратор. Отметим, что все время с момента поступления запроса до окончания вывода УПС занята обработкой одного запроса. Например, при поступлении команды QUEUE УПС "просматривает" каталоги очередей ввода, вывода, считающихся задач и посылает выходное сообщение на концентратор.

Приняв такую модель системы, мы можем воспользоваться теорией очередей.

Введем обозначения^{/II/}. Пусть t_s - время обслуживания запроса, σ_{t_s} - стандартное отклонение времени обслуживания, ω - число запросов, ожидающих обслуживания в данный момент, а q - число запросов в системе, ожидающих обслуживания и обслуживаемых в данный момент. Пусть далее t_w - время ожидания запросом обслуживания, а t_q - время пребывания запроса в системе, т.е. время, затраченное им и на ожидание и на обслуживание. Средние значения и дисперсии величин ω , q , t_w , t_q обозначим соответственно через $E(\omega)$, $E(q)$, $E(t_w)$, $E(t_q)$ и $Var(\omega)$, $Var(q)$, $Var(t_w)$, $Var(t_q)$. Среднее число обслуживаемых сообщений - через $E(n)$.

Пусть ρ - коэффициент использования системы. Тогда благодаря устойчивости состояния

$$\rho = E(n) \cdot E(t_s).$$

Воспользуемся формулой Хинчина-Поллачека:

$$E(\omega) = \frac{\rho^2}{2(1-\rho)} \left\{ 1 + \left[\frac{6t_s}{E(t_s)} \right]^2 \right\}. \quad (1)$$

Она применяется в случае экспоненциального распределения времени поступления запросов при любом распределении времени обслуживания и любой дисциплине управления, лишь бы выбор очередного запроса для обслуживания не зависел от времени обслуживания. Следовательно, она применима в нашем случае, т.к. для большинства диалоговых установок процесс поступления заданий достаточно близок к пуассоновскому процессу^{/IO/}, а дисциплина обслуживания в нашем случае - "первым пришел - первым обслужен".

Из (1) следует:

$$E(q) = \rho + \frac{\rho^2}{2(1-\rho)} \left\{ 1 + \left[\frac{6t_s}{E(t_s)} \right]^2 \right\}, \quad (2)$$

$$E(t_w) = \frac{\rho E(t_s)}{2(1-\rho)} \left\{ 1 + \left[\frac{6t_s}{E(t_s)} \right]^2 \right\},$$

$$E(t_q) = E(t_s) + \frac{\rho E(t_s)}{2(1-\rho)} \left\{ 1 + \left[\frac{6t_s}{E(t_s)} \right]^2 \right\}.$$

Предположив, что время обслуживания запросов следует гамма-распределению, что вполне приемлемо для большинства диалоговых систем^{/II/}, можно воспользоваться формулами для дисперсии q и t_q :

$$Var(q) = \frac{1}{(1-\rho)^2} \left\{ 1 - \frac{\rho}{2} \left[3 - \frac{\rho(10-\rho)}{6} - \frac{3-3\rho+\rho^2}{R} - \frac{\rho(8-5\rho)}{6R^2} \right] \right\},$$

$$Var(t_q) = \left[\frac{E(t_s)}{1-\rho} \right]^2 \left\{ \left[1 - \frac{\rho}{6} (4-\rho) \left(1 - \frac{1}{R} \right) \right] \left(1 + \frac{1}{R} \right) - \left[1 - \frac{\rho}{2} \left(1 - \frac{1}{R} \right) \right]^2 \right\}, \quad \text{где} \quad (3)$$

$$R = \left[\frac{E(t_s)}{6t_s} \right]^2.$$

Результаты измерений и моделирования

Для того, чтобы воспользоваться формулами (2) и (3), необходимо было провести измерение параметров рабочей нагрузки и характеристик времени обслуживания запросов.

Измерения проводились программным способом на протяжении месяца в реально функционирующей системе. Для этого в УПС был добавлен блок, который по каждому дистанционному запросу делал запись в специальный системный файл, расположенный на диске. Эти данные ежедневно переносились на магнитные ленты и затем обрабатывались программой /I3/.

Измеренная средняя интенсивность рабочей нагрузки составила в дневные и вечерние часы около 850 запросов в час, достигая в периоды наибольшей нагрузки 1450 заявок в час. Характеристики времен обслуживания запросов приведены в табл.1.

Измеренное среднее время выполнения запроса $E(t_s)$ составило 1,059 секунды со стандартным отклонением $\sigma_{t_s} = 0,999$ с. Относительно большое среднее время обслуживания имеют запросы INPUT и LOCAL - 4,049 и 3,517 секунд соответственно, что объясняется, с одной стороны, большим числом межмашинных обменов, выполняемых по этим запросам (среднее число обменов для INPUT - 5,038, для LOCAL - 7,725, для остальных запросов - один или два обмена), а с другой стороны - сложностью процедур постановки заданий во входную очередь и выборки листингов из выходной очереди БЭСМ-6. Отметим, что большую часть запросов (более 80%) составляют запросы на обмен между пользователями и интерактивными программами в соответствии с операторами READ и PRINT, редко (0,2%) применяются команды обмена сообщениями между пользователями и операторами ЭВМ.

тип запроса характеристика	QUEUE	READ	PRINT	INPUT	LOCAL	MES1	MES2	смесь запросов
доля запросов	0,156	0,150	0,654	0,023	0,015	0,001	0,001	1,000
$E(t_s)$ (с)	1,440	0,076	1,033	4,049	3,517	0,087	0,016	1,059
$\text{var}(t_s)$ (с ²)	1,157	0,013	0,001	6,875	12,418	0,001	0,001	0,998
σ_{t_s} (с)	1,076	0,114	0,013	2,622	3,524	0,021	0,011	0,999

Табл.1. Характеристики времени обслуживания запросов.

Подставляя входные данные в формулы (2), (3), получим характеристики времени ответа на запросы с учетом очередей. На рис.2 показана зависимость среднего времени ответа (в секундах) от нагрузки (запросов в час).

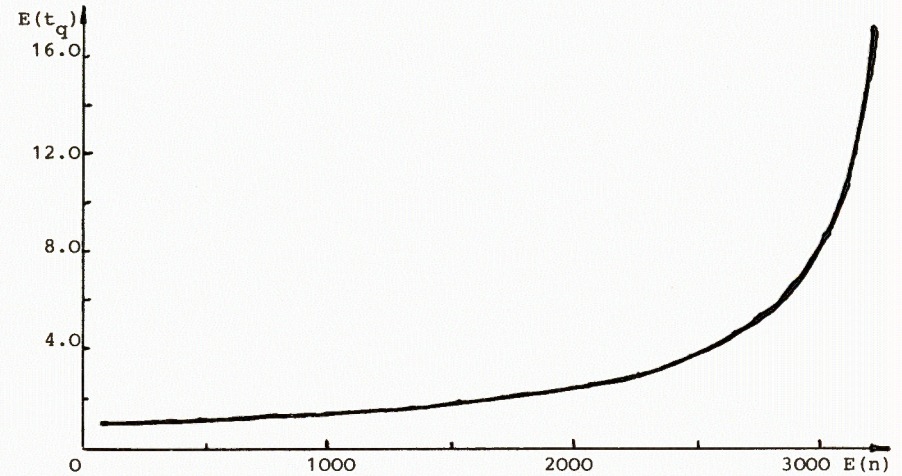


Рис.2. Зависимость среднего времени ответа от рабочей нагрузки.

Значения характеристик времен ответа для средней и пиковой нагрузки приведены в табл.2.

Отсюда видно, что при существующей в системе средней рабочей нагрузке очередь запросов образуется незначительно - ее средняя длина - 0,079 запросов; невелико поэтому приращение среднего времени ответа (1,393 с) по сравнению со средним временем обслуживания запроса (1,059 с). Участок кривой (рис.2) в окрестности точки средней рабочей нагрузки имеет невысокую скорость роста, поэтому при небольшом увеличении нагрузки характеристики времени ответа на запрос увеличиваются незначительно. Резкий рост кривой наблюдается, когда коэффициент использования "оборудования" превышает 0,8. В системе с $\rho > 0,8$ незначительное увеличение нагрузки может привести к резкому спаду производительности или даже заставить ее работать в аварийном режиме /II/.

Увеличение рабочей нагрузки на небольшую величину $\delta\%$ приводит к увеличению размеров очереди приблизительно на

$$E(t_s) + \frac{E(t_s) \rho (2-\rho)}{2(1-\rho)^2} \left\{ 1 + \left[\frac{\sigma_{t_s}}{E(t_s)} \right]^2 \right\} \delta.$$

При существующей в нашей системе нагрузке ($\rho = 0,25$) это увеличение равно $\frac{16}{9} E(t_s) \delta$ для экспоненциального закона распределения времени обслуживания. Но если коэффициент использования оборудования равен, скажем, 0,9, то увеличение размера очереди равно $100 E(t_s) \delta$, что более чем в 50 раз больше. Аналогично увеличиваются время пребывания в очереди и время ответа.

характеристика нагрузка (сообщений в час)	ρ	$E(\omega)$	$E(q)$	$E(t_w)$	$E(t_q)$	σ_{t_q}
$E(n)=850$	0,250	0,079	0,325	0,334	1,393	1,328
$E(n)=1450$	0,427	0,300	0,727	0,745	1,804	1,736

Табл.2. Характеристики времени ответа для средней и пиковой нагрузок.

Среднее значение и стандартное отклонение времени ответа для отдельных типов запросов можно оценить, используя соотношения.

$$E(t_q^A) = E(t_w) + E(t_s^A), \quad (4)$$

$$\text{Var}(t_q^A) = \text{Var}(t_w) + \text{Var}(t_s^A),$$

где индекс A указывает, что данная характеристика относится к запросу типа A .

На рис.3 показаны средние значения времени ответа для запросов различных типов в зависимости от нагрузки. Значения характеристик времени ответа на эти запросы для средней и пиковой нагрузок приведены в табл.3.

$E(n)=850$

тип запроса характеристика	Q	READ	PRINT	INPUT	LOCAL
$E(t_q)$	1,774	0,410	1,367	4,383	3,851
$\sigma(t_q)$	1,387	0,883	0,876	2,764	3,631

$E(n)=1450$

тип запроса характеристика	Q	READ	PRINT	INPUT	LOCAL
$E(t_q)$	2,185	0,821	1,778	4,794	4,262
$\sigma(t_q)$	1,781	1,424	1,420	2,982	3,799

Табл.3. Характеристики времени ответа на запросы различных типов для средней и пиковой нагрузок.

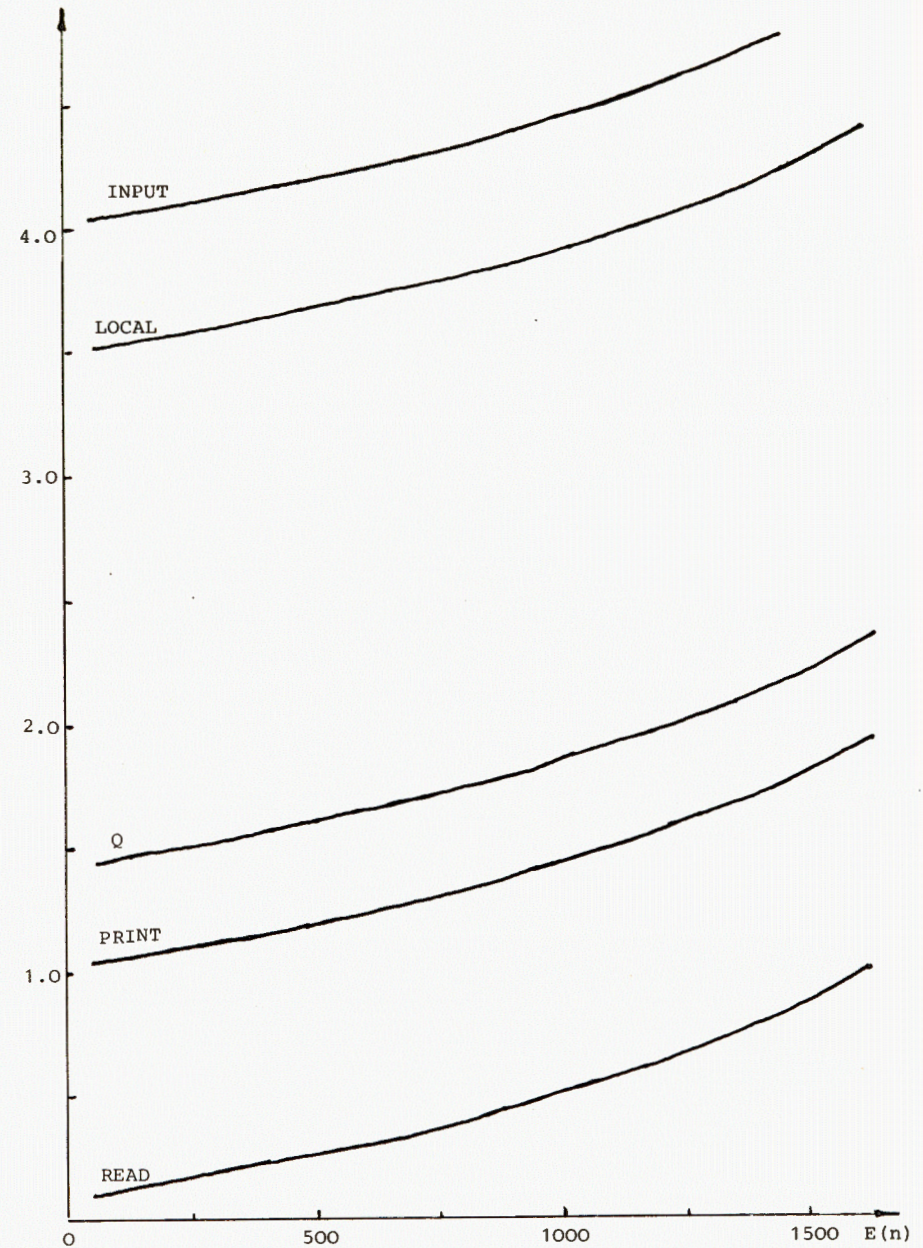


Рис.3. Зависимость средних значений времени ответа от нагрузки для запросов различных типов.

Выводы

Проведенная оценка эффективности системы дистанционного обслуживания пользователей концентратора терминалов показывает, что при существующей рабочей нагрузке система удовлетворительно справляется с выполнением поступающих запросов, при этом среднее время ответа не превосходит 1,4 секунды. Анализ результатов моделирования показывает, что система устойчива по отношению к изменению нагрузки. В частности, удвоение существующей нагрузки, например, путем увеличения числа терминалов или с помощью подключения дополнительной машины-концентратора не вызовет значительного увеличения времени ответа. Действительно, при увеличении рабочей нагрузки до 1700 запросов в час среднее время ответа на дистанционные команды (в соответствии с моделью) будет составлять 2,061 секунды, что вполне приемлемо и соответствует увеличению среднего времени ответа примерно на 48%.

Литература

1. Галактионов В.В., Каданцев С.Г., Шириков В.П. В кн.: Совещание по программированию и математическим методам решения физических задач. Дубна, 1977. ОИЯИ, Д10, II-11264, Дубна, 1978, с.51-58.
2. Галактионов В.В. и др. В кн.: Материалы II Всесоюзного совещания "Диалоговые вычислительные комплексы (ДИАЛОГ-79)". ИФВЭ, Серпухов, 1979, с.66-69.
3. Галактионов В.В. и др. Там же, с.70-74.
4. Галактионов В.В. и др. В кн.: Тезисы докладов Всесоюзной конференции "Диалог человек-ЭВМ". ЛИАП, Ленинград, 1982, с.90-92.
5. Каданцев С.Г. ОИЯИ, II-82-828, Дубна, 1982.
6. Дроздов Е.А., Пятибратов А.П. Основы построения и функционирования вычислительных систем. М., Энергия, 1973.
7. Панфилов И.В., Половко А.М. Вычислительные системы. М., Сов.радио, 1980.
8. Пятибратов А.П. Вычислительные системы с дистанционным доступом. М., Энергия, 1979.
9. Бальбердин В.А. Методы анализа мультипрограммных систем. М., Радио и связь, 1982.
10. Феррари Д. Оценка производительности вычислительных систем. М., Мир, 1981.
11. Мартин Дж. Системный анализ передачи данных. М., Мир, 1975.
12. Галактионов В.В., Микушаускас Р.К. ОИЯИ, P10-82-319, Дубна, 1982.
13. Каданцев С.Г. ОИЯИ, II-84-709, Дубна, 1984.

Рукопись поступила в издательский отдел
2 ноября 1984 года.

СООБЩЕНИЯ, КРАТКИЕ СООБЩЕНИЯ, ПРЕПРИНТЫ И СБОРНИКИ ТРУДОВ КОНФЕРЕНЦИЙ, ИЗДАВАЕМЫЕ ОБЪЕДИНЕННЫМ ИНСТИТУТОМ ЯДЕРНЫХ ИССЛЕДОВАНИЙ, ЯВЛЯЮТСЯ ОФИЦИАЛЬНЫМИ ПУБЛИКАЦИЯМИ.

Ссылки на СООБЩЕНИЯ и ПРЕПРИНТЫ ОИЯИ должны содержать следующие элементы:

- фамилии и инициалы авторов,
- сокращенное название Института /ОИЯИ/ и индекс публикации,
- место издания /Дубна/,
- год издания,
- номер страницы /при необходимости/.

Пример:

1. *Переушкин В.Н. и др. ОИЯИ, P2-84-649, Дубна, 1984.*

Ссылки на конкретную СТАТЬЮ, помещенную в сборнике, должны содержать:

- фамилии и инициалы авторов,
- заглавие сборника, перед которым приводятся сокращенные слова: "В кн."
- сокращенное название Института /ОИЯИ/ и индекс издания,
- место издания /Дубна/,
- год издания,
- номер страницы.

Пример:

Колпаков И.Ф. В кн. XI Международный симпозиум по ядерной электронике, ОИЯИ, Д13-84-53, Дубна, 1984, с.26.

Савин И.А., Смирнов Г.И. В сб. "Краткие сообщения ОИЯИ", № 2-84, Дубна, 1984, с.3.

Принимается подписка на препринты и сообщения Объединенного института ядерных исследований.

Установлена следующая стоимость подписки на 12 месяцев на издания ОИЯИ, включая пересылку, по отдельным тематическим категориям:

ИНДЕКС	ТЕМАТИКА	Цена подписки на год
1.	Экспериментальная физика высоких энергий	10 р. 80 коп.
2.	Теоретическая физика высоких энергий	17 р. 80 коп.
3.	Экспериментальная нейтронная физика	4 р. 80 коп.
4.	Теоретическая физика низких энергий	8 р. 80 коп.
5.	Математика	4 р. 80 коп.
6.	Ядерная спектроскопия и радиохимия	4 р. 80 коп.
7.	Физика тяжелых ионов	2 р. 85 коп.
8.	Криогеника	2 р. 85 коп.
9.	Ускорители	7 р. 80 коп.
10.	Автоматизация обработки экспериментальных данных	7 р. 80 коп.
11.	Вычислительная математика и техника	6 р. 80 коп.
12.	Химия	1 р. 70 коп.
13.	Техника физического эксперимента	8 р. 80 коп.
14.	Исследования твердых тел и жидкостей ядерными методами	1 р. 70 коп.
15.	Экспериментальная физика ядерных реакций при низких энергиях	1 р. 50 коп.
16.	Дозиметрия и физика защиты	1 р. 90 коп.
17.	Теория конденсированного состояния	6 р. 80 коп.
18.	Использование результатов и методов фундаментальных физических исследований в смежных областях науки и техники	2 р. 35 коп.
19.	Биофизика	1 р. 20 коп.

Подписка может быть оформлена с любого месяца текущего года.

По всем вопросам оформления подписки следует обращаться в издательский отдел ОИЯИ по адресу: 101000 Москва, Главпочтамт, п/я 79.

Каданцев С.Г.

11-84-710

Оценка производительности системы дистанционного обслуживания на базе БЭСМ-6 и концентратора терминалов

Проведена оценка производительности системы дистанционного обслуживания на базе БЭСМ-6 и концентратора терминалов. Представлены результаты измерений отдельных показателей производительности и моделирования работы системы по обслуживанию дистанционных запросов, поступающих на БЭСМ-6 от пользователей концентратора терминалов.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1984

Перевод О.С.Виноградовой

Kadantsev S.G.

11-84-710

Performance Evaluation of the Remote Job Entry Based on the BESM-6 Computer and Terminal Concentrator

Performance evaluation of the interactive computer system based on the BESM-6 terminal concentrator is performed. Some results of performance index measurements and the system simulation are given.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1984