

## **ИССЛЕДОВАНИЕ ЗАКОНА РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ЗАГРУЖЕННОСТИ ПРОЦЕССОРА ПЕРСОНАЛЬНОГО КОМПЬЮТЕРА**

### **Введение**

Решение задачи увеличения скорости обработки данных является одним из основных критериев при разработке и эксплуатации вычислительных машин (комплексов). Одним из направлений увеличения скорости обработки информации (производительности), является задействование нескольких процессоров на решение одной задачи.

### **Параллельная обработка информации**

Распараллеливание рассматривают на нескольких уровнях

1. Аппаратный(многоконвейерность, многоядерность, многопроцессорность) [1].
2. Уровень операционной системы (MOSIX, openMOSIX) [2].
3. Уровень приложений (использование специальных библиотек при программировании, либо написание специализированных алгоритмов на стандартных языках программирования) [3].
4. GRID вычисления [4].

Уровень 1 - частично закрытая информация.

Уровень 2 - есть проекты с открытым кодом, однако законодателями в этой области являются научные институты, и деятельность отдельного человека может быть только в области использования предложенной технологии.

Уровень 3 - наиболее интересен для отдельных разработчиков

Уровень 4 - здесь уже определились те, кто формирует стандарты, однако ещё далеко до окончательной формулировки этих стандартов.

Созданные программные продукты, реализующие распределённые вычисления, можно разделить на несколько категорий [5]

1. Параллельность на уровне алгоритма.
2. Мигрирование процессов (уровень операционной системы).
3. Разделение данных.

В открытых источниках описаны общие принципы работы систем и некоторые технические подробности. Однако разработчики алгоритмов не афишируют теоретическую базу своих технических решений.

В данной статье рассматривается проблематика для класса задач, которые относятся к третьей категории.

© Б.С. Ризваш, Е.В. Крылов, 2006

## Описание задачи

Имеется локальная вычислительная сеть, где часть времени компьютеры не загружены на 100%. Имеется вычислительная задача, требующая много процессорного времени (более 5 минут 100% загрузки процессора). Требуется оптимально разделить вычисления между имеющимися компьютерами.

Критерием оптимальности является минимизация времени выполнения вычислений.

Ограничение – пользователи, работающие за компьютерами получившими задания, должны иметь возможность выполнять свои задачи.

Для оценки загрузки необходимо провести исследования, и выявить закон распределения вероятностей загрузки.

При отсутствии некоего закона распределения, будет невозможно оценить вероятность загрузки, и соответственно решение о выдаче вычислительного задания не будет учитывать сформулированное ограничение.

### Детализация задачи

Для решения задачи распределения вычислительной нагрузки, необходимо найти способ оценить вычислительные возможности компьютеров (некоторый коэффициент производительности “ $K$ ”), и разделить вычислительную задачу между компьютерами пропорционально полученному коэффициенту. Причём, это распределение должно удовлетворять целевой функции (минимизация времени выполнения задачи), и не нарушать ограничения.

$$K = f(F, L, M), \quad (1)$$

где

$F$  – тактовая частота процессора;

$L$  – загрузка процессора;

$M$  – объём свободной оперативной памяти.

Наибольший интерес представляет загрузка процессора ( $L$ ), как величина быстро изменяющаяся. Процент загрузки процессора влияет не только на “ $K$ ”, но и на решение о выдаче задания.

### Постановка задачи

*Используемые термины.*

Комплекс - совокупность программ, предназначенных для решения поставленной задачи.

Клиент - компьютер, которому выдаётся задание на вычисление.

Сервер – компьютер, распределяющий задания.

*Область исследований.*

Поиск закона распределения.

*Формулировка задачи*

Найти ранее исследованный закон распределения, либо комбинацию таких законов, для описания вероятности загрузки процессора.

*Условия проведения эксперимента.*

Характеристики оборудования.

- Процессор (800 МГц – 2.4 МГц одноядерные, с или без технологии HT).
- Оперативная память (128 Мб – 1 Гб ).
- Жесткие диски (технологии ATA100 – SATA-II).
- Операционная система – Windows2000 или WindowsXP.

### **Планирование эксперимента**

#### **Сбор эмпирических данных**

Наблюдение за загруженностью будет производиться в течении полного рабочего дня, на компьютерах имеющих определённое функциональное предназначение. Планируется выполнить 30 циклов наблюдения по 10 компьютерам. Показания снимаются с помощью службы WMI каждую секунду. Возможные неточности работы данной службы компенсируются тем, что решение про загруженность клиента будет приниматься по данным, полученным с помощью этой же службы. Полученные данные будут разбиваться по часам рабочего времени (с 10 до 11, с 11 до 12 и т.д.).

Для каждого часа строится график загруженности процессора и занятости памяти по времени. Для каждого графика загруженности процессора строится ряд распределения частотей загруженности процессора от 0 до 100 % с шагом 1. Для каждого графика занятости памяти строится ряд распределения частотей занятости от 0 до 100% с шагом 1.

Оцениваем характер изменения данных величин и подбираем закон распределения.

### **Результаты проведения экспериментов**

Наиболее характерные из обработанных данных

Похожесть характера рядов распределения частотей позволяет говорить о том, что может существовать закон распределения, который способен описать все эти графики.

Данные получены в результате независимых экспериментов.

### **Подбор аналитической функции**

При оценке загруженности возможны ошибки 2-х видов:

1. выдали задание компьютеру, который не может его выполнить в требуемые сроки (т.е. недооценили загруженность);
2. не выдали задание компьютеру, который может с задачей справиться (т.е. переоценили загруженность).

При данной постановке задачи, более неприемлемой является ошибка 1-го вида. То есть, при оценке загруженности пользуемся пессимистической оценкой (т.е. считаем, что компьютер будет загружен больше, чем на самом деле).

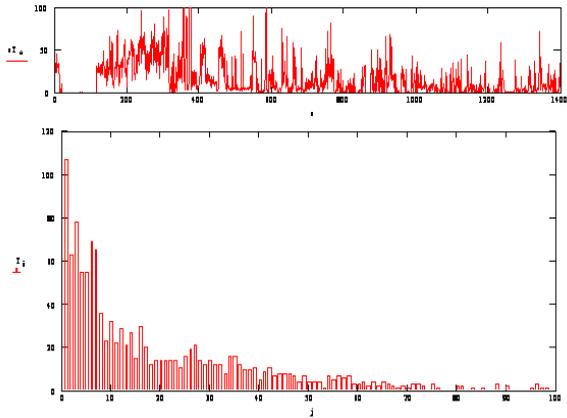


Рис. 1 – (Celeron 2.8 ГГц , 512 Гб) 18ч-19ч (16.03.2006)

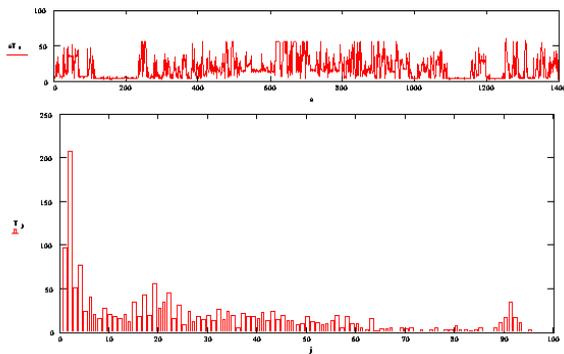


Рис. 2 – (P4 HT , 2.8 ГГц , 1 Гб) 16ч – 17ч (16.03.2006)

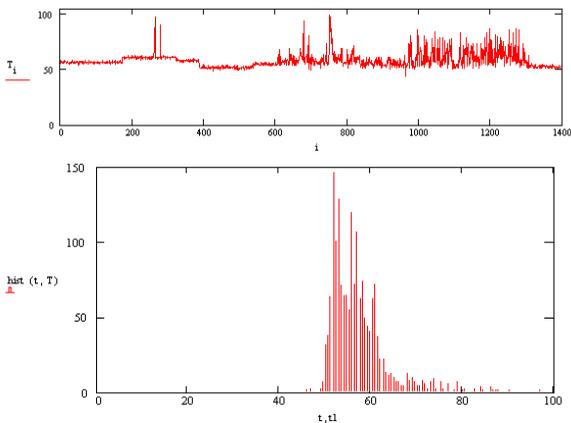


Рис. 3 – (Р4 НТ , 2.8 ГГц , 1 Гб) 13ч – 14ч (15.03.2006)

Так как нас интересует ПЕССИМИСТИЧЕСКАЯ оценка загруженности, то рассматриваем правую от пика загруженности часть ряда распределения частотей.

Соответственно графику, будем искать функцию в виде

$$L(t) = \frac{A1}{A2 \cdot t + A3} \tag{2}$$

где  $A1, A2, A3$  – некоторые коэффициенты, зависящие от ряда распределения частотей.

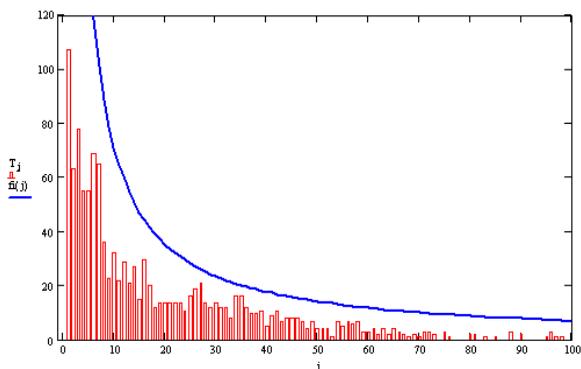


Рис. 4 – Приближение экспериментальных данных эмпирической формулой (соответствует рисунку 1)

Исследования проводились в помощью пакета MathCAD. Формула приведена в формате данного пакета.

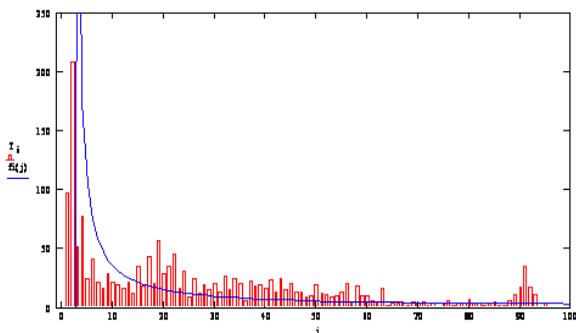


Рис. 5 – Приближение экспериментальных данных эмпирической формулой (соответствует рисунку 2)

$$f_i(x) := \frac{\max(T)}{0.8x - \text{ord}X(T)} \quad (3)$$

Приведённая на графике кривая получена с помощью функции написанной в формате пакета MathCAD. Данная функция возвращает номер интервала из исследуемой области, в котором находится наиболее вероятная загруженность. Так как исследуется интервал от 0 до 100% загруженности, и он разбит на 100 интервалов, то фактически номер возвращаемого интервала является показателем процента загруженности

$$\text{ord}X(T) := \left| \begin{array}{l} j \leftarrow 0 \\ \text{tmp} \leftarrow T_0 \\ \text{for } i \in 0.. \text{rows}(T) - 1 \\ \quad \text{if } \text{tmp} < T_i \\ \quad \quad \left| \begin{array}{l} j \leftarrow i \\ \text{tmp} \leftarrow T_i \end{array} \right. \\ \end{array} \right| j$$

Приведенная полученная эмпирически формула (3), отвечает поставленной задаче.

### Вывод

Применить какой-либо из стандартных законов распределения для описания загруженности процессора не представляется возможным, ввиду наличия в конце исследуемого интервала (около 100% загруженности) некоторого “всплеска” (что особенно хорошо видно на рисунке 2).

Введём новое определение: ШИРИНА ВАРЬИРОВАНИЯ (от пика до 100% загруженности)

Вместо попытки нахождения функции, способной описать полностью данную кривую, было принято решение разбить график на 2 части. На выбранных отрезках можно применить следующие законы распределения.

1. Эмпирическая формула – 80% ШИРИНЫ ВАРЬИРОВАНИЯ.
2. НОРМАЛЬНЫЙ закон распределения с математическим ожиданием в точке 90% ширины варьирования и дисперсией 10% ШИРИНЫ ВАРЬИРОВАНИЯ.

При использовании полученного результата в реальных проектах, для избегания необходимости проверки гипотезы о положении пика загрузки, можно ввести ограничения, которые будут сильнее проверки. К примеру, решении задачи о назначении или не назначении конкретному компьютеру задания, можно ввести условие, что при наличии вероятности более 30 для загруженностей выше 80% не назначать задание, и продолжить сбор информации

В данной статье приведён вариант описания закона загрузки процессора. Можно подобрать более сложную зависимость, но ввиду того, что необходима пессимистическая оценка, то и такая зависимость является подходящей.

Полученный результат может быть использован в задачах принятия решений, касающихся загрузки процессора. Например – в задачах распределенной обработки данных, когда присутствует описанное в начале данной статьи ограничение.

### Литература

1. Multithreading and Hyper-Threading, Technology & Research at Intel. <http://www.intel.com/technology/computing/htt/index.htm>
2. An Introduction to openMosix Clusters. [http://openmosix.sourceforge.net/introduction\\_to\\_openmosix.html](http://openmosix.sourceforge.net/introduction_to_openmosix.html)
3. Распределенные вычисления в Интернете. <http://distributed.ru/?what-is>
4. Концепция грид. <http://gridclub.ru/about>
5. “Основы параллельных вычислений” В.Э. Малышкин ч.1,2. <http://ssga.ru/metodich/paral1/index.html>
6. <http://ssga.ru/metodich/paral2/index.html>