

2.6. ПРОИЗВОДИТЕЛЬНОСТЬ ЭВМ

В данном параграфе дадим понятие о производительности ЭВМ и введем показатели и единицы измерения производительности машин.

2.6.1. Понятие о производительности ЭВМ

Электронные вычислительные машины являются наиболее распространенным, достаточно производительным средством индустрии обработки информации. Неослабевающий интерес к вопросам анализа эффективности функционирования ЭВМ объясняется следующими тремя причинами:

- 1) ЭВМ нашли применение во многих областях человеческой деятельности; число задач, допускающих решение на ЭВМ, постоянно растет; правильный выбор той или иной ЭВМ для конкретной области применения может быть осуществлен на основе анализа их возможностей, на основе численного анализа их показателей эффективности (в частности, производительности);
- 2) ЭВМ в расширенных или минимальных конфигурациях (процессоры с памятью) выступают в качестве массовых функциональных элементов в параллельных вычислительных системах;
- 3) ЭВМ и параллельные ВС составляют основу сложных систем различного назначения и, в частности, более мощных вычислительных средств, таких, например, как вычислительные сети и распределенные ВС.

Ясно, что эффективность функционирования ЭВМ не может быть исчерпывающим образом охарактеризована с помощью одного, пусть даже составного (комплексного, “универсального”) показателя. В самом деле, часто в качестве составного показателя эффективности ЭВМ берут дробь, в числителе которой стоят те количественные характеристики, которые желательно максимизировать, а в знаменателе – те, которые необходимо минимизировать. Общим недостатком составных показателей является то, что малая эффективность по одному показателю всегда может быть скомпенсирована за счет другого (например, малая производительность – за счет низкой стоимости машины). Именно поэтому в поле зрения должен находиться вектор показателей, которые в совокупности позволяют оценить эффективность ЭВМ: производительность и надежность ЭВМ, осуществимость решения задач на машине и технико-экономическую эффективность ее функционирования. При комплексном оценивании поведения ЭВМ могут использоваться и составные показатели, но вместе с другими и обязательно с теми, которые входят в составные показатели.

Некоторые из показателей эффективности ЭВМ мы уже использовали в гл. 1 и в 2.5, например, быстродействие и среднее время безотказной работы (среднюю наработку до отказа). Это позволило достичь завершенности в анализе архитектурных решений, примененных в ЭВМ трех поколений. Однако строгие определения показателей не приводились; расчет был сделан на интуитивное восприятие смысла того или иного показателя эффективности функционирования ЭВМ.

В данном параграфе и последующих трёх мы математически строго определим показатели эффективности работы ЭВМ, выведем расчетные формулы, произведем численный анализ эффективности функционирования ЭВМ.

Общеизвестно, что производительность физического труда и оборудования измеряется объемом работы и продукции производимыми в единицу времени. Измерение производительности умственного труда и производительности такого оборудования как ЭВМ является специфической и сложной проблемой. Ясно, что количественные характеристики для последних измерений связаны с информацией. Здесь предметом нашего рассмотрения будет производительность ЭВМ (Computer Performance).

Под производительностью ЭВМ понимается ее способность обрабатывать информацию. Как правило, когда говорят о производительности, то понимают под этим потенциальную возможность ЭВМ по обработке информации (а не реальную, учитывающую аномальности в работе ЭВМ, например простои из-за отказов, из-за профилактического обслуживания и т.п.). В процессе обработки информации в ЭВМ реализуются те или иные операции из ее набора (или системы) операций. Состав набора операций безусловно, характеризует архитектуру ЭВМ и, следовательно, определяет ее производительность.

Для оценки способности ЭВМ производить обработку информации используют *количественные характеристики или показатели производительности.* Эти показатели безусловно связаны с количеством информации, которое способна обработать ЭВМ в единицу времени, и, следовательно – с временем выполнения операций. Время выполнения операции в общем случае складывается из времени выборки (из памяти ЭВМ в процессор) команды (кода операции и адресов операндов), времени чтения операндов, времени реализации собственно операции и времени занесения результатов в память. При выполнении последовательности команд возможны совмещения во времени выполнения нескольких операций. При этом всегда имеет место зависимость времени выполнения операции от времени обращения к памяти.

2.6.2. Показатели производительности ЭВМ

Распространенным простейшим показателем производительности ЭВМ является *тактовая частота.* Тактовая частота указывает сколько элементарных операций (тактов) может осуществить в единицу времени (секунду) ЭВМ (точнее: ее процессор). Или, говоря иначе, время такта – это время выполнения элементарной операции процессора ЭВМ. Например, время такта в первых ЭВМ измерялось в миллисекундах (в EDSAC оно было равно 4 мс), а в современных машинах оно выражается в наносекундах ($1 \text{ нс} = 10^{-9} \text{ с}$) или даже величинами порядка 10^{-10} с .

Процессоры с одинаковой архитектурой (системой команд и логической структурой) могут иметь разную тактовую частоту. Процессор с более высокой тактовой частотой обладает большей производительностью (и, следовательно, более дорогой). С другой стороны, в различных моделях семейства процессоров одни и те же операции (например, сложения или умножения) выполняются за разное число тактов (за счет совмещения во времени выполнения нескольких элементарных операций). Чем “старше” модель, тем меньше (как правило) требуется тактов для выполнения одних и тех же операций, следовательно, тем выше ее производительность и цена (даже при одинаковых тактовых частотах).

В качестве простейших показателей производительности ЭВМ используются также числа однотипных операций, выполняемых в единицу времени (над операндами с одинаковой разрядностью). Для оценки производительности ЭВМ получили распространение числа операций, например, сложения с фиксированной запятой, сложения с плавающей запятой, умножения и деления, выполняемых в секунду. Так как в общем случае длительности операций даже при одних и тех же операндах зависят от их типов, то для характеристики производительности ЭВМ целесообразно использовать средние числа операций, выполняемых в единицу времени. Ниже введем следующие показатели: номинальное быстродействие, быстродействие по Гибсону, среднее и среднее эффективное быстродействие ЭВМ.

Номинальным (или пиковым, или техническим) быстродействием (Nominal Speed или Peak Speed) ω' ЭВМ назовем среднее число операций, выполняемых процессором (при равновероятном их выборе) в единицу времени (секунду) при работе только с оперативной памятью. (Заметим, что в состав такой памяти включаются собственно

оперативная память и сверхоперативная – регистры общего назначения или кэш-память процессора).

Пусть $\{\kappa_1, \kappa_2, \dots, \kappa_j, \dots, \kappa_k\}$ – часть операций из их полного набора (системы), которые не требуют обращений процессора к внешней памяти и устройствам ввода-вывода информации; τ_j – время (в секундах) выполнения операции κ_j ; вероятность выбора любой операции κ_j , $j \in \{1, 2, \dots, k\}$, есть постоянная вида $1/k$. Тогда, по определению, математическое ожидание времени выполнения операции в ЭВМ будет

равно $k^{-1} \sum_{j=1}^k \tau_j$; следовательно, номинальное быстродействие

$$\omega' = k \left(\sum_{j=1}^k \tau_j \right)^{-1}. \quad (2.5)$$

Очевидно, что при реализации на ЭВМ реальных программ решения задач имеет место неравновероятное использование тех или иных операций. Пусть ρ_j , $j \in \{1, 2, \dots, k\}$, – вероятность выбора операции κ_j (вероятность спроса на κ_j или вес операции κ_j). Тогда средним временем выполнения операции ЭВМ и *быстродействием ЭВМ по Гибсону* будут величины:

$$\sum_{j=1}^k \rho_j \tau_j; \quad \omega^0 = \left(\sum_{j=1}^k \rho_j \tau_j \right)^{-1}. \quad (2.6)$$

Распределение вероятностей $\{\rho_j\}$, $j = \overline{1, k}$, $\sum_{j=1}^k \rho_j = 1$, или набор весов ρ_j зависят от характера задач. Существует несколько наборов весовых коэффициентов или, как говорят, несколько “смесей Гибсона”, которые отражают статистику задач, решаемых на ЭВМ.

На практике достаточно часто используют модификации показателей (2.5) и (2.6), в которых в подмножество $\{\kappa_j\}$, $j = \overline{1, k}$, включаются только операции с фиксированной запятой (см. 1.1.5).

Из (2.6) следует, что даже при работе с оперативной памятью процессор ЭВМ будет выполнять в единицу времени различное число операций при решении задач различных типов. Кроме того, при решении задачи на ЭВМ требуются, в общем случае, затраты машинного времени на ввод программы и данных, обращение к внешней памяти, работу операционной системы, вывод результатов и т.п. Из сказанного видно, что в общем случае среднее число операций, связанных с решением задач и выполняемых процессором в единицу времени, будет отличаться и от номинального быстродействия машины, и от быстродействия ЭВМ по Гибсону. Поэтому для характеристики производительности ЭВМ при решении задач целесообразно ввести дополнительные показатели.

Пусть $\{I_1, I_2, \dots, I_i, \dots, I_L\}$ – набор типовых задач (эталонных тестов или программ оценки производительности, Benchmarks). Пусть также V_i – число операций, непосредственно входящих в программу решения задачи I_i ; t_i – время (в секундах) решения задачи I_i (в t_i входят время, расходуемое ЭВМ собственно на счет, и дополнительные затраты машинного времени при решении I_i); быстродействием ЭВМ

при решении типовой задачи I_i , $i \in \{1, 2, \dots, L\}$, назовем $\omega_i = v_i / t_i$. Величина $1/\omega_i$ является средним временем выполнения одной операции при решении задачи типа i , $i \in \{1, 2, \dots, L\}$. Пусть далее $\{\pi_1, \pi_2, \dots, \pi_i, \dots, \pi_L\}$ – распределение вероятностей спроса на типовые задачи I_i , $i = \overline{1, L}$, $\sum_{i=1}^L \pi_i = 1$, тогда $\sum_{i=1}^L \pi_i / \omega_i$ будет средним временем выполнения одной операции при решении набора типовых задач. Средним *быстродействием ЭВМ* назовем величину

$$\omega = \left(\sum_{i=1}^L \pi_i / \omega_i \right)^{-1}. \quad (2.7)$$

Очевидна технология применения показателя ω для оценки производительности совместимых машин или ЭВМ в пределах одного семейства. Но формулу (2.7) можно использовать и для анализа производительности ЭВМ различных семейств. В самом деле, если машины имеют “языковую” совместимость, то все тесты I_i , $i = \overline{1, L}$, могут быть представлены программами на одном и том же языке (например, на языке высокого уровня ФОРТРАН). Тогда, используя для различных ЭВМ свои значения t_i и v_i и ω_i , можно легко рассчитать величину ω , см. (2.7).

Существует несколько тестовых наборов $\{I_i\}$, $i = \overline{1, L}$, среди которых достаточно популярен LINPAC.

Как уже отмечалось выше, при сравнении производительности (2.7) несовместимых ЭВМ можно “пропускать” наборы эталонных тестов и оценивать время их решения. Однако в этом случае следует учитывать то, что для каждого теста должны быть произведены свои программы на языках команд анализируемых ЭВМ. Можно поступить иначе: ввести “стандартизацию” операций и форматов операндов. В качестве стандартной может быть взята операция, через которую могут быть выражены все остальные операции в каждой из сравниваемых машин. Стандартным форматом операндов может служить одна из структурных единиц информации, например, байт. В этих условиях в качестве показателя производительности ЭВМ целесообразно использовать среднее эффективное быстродействие. Эффективным быстродействием ЭВМ при решении типовой задачи I_i , $i \in \{1, 2, \dots, L\}$, назовем величину $\omega_i^* = v_i^* / t_i$, где v_i^* – число стандартных операций, с помощью которых можно интерпретировать программу решения задачи I_i ; t_i – время решения задачи I_i на ЭВМ (включающее время реализации собственно программы и накладные расходы времени). Средним *эффективным быстродействием ЭВМ* будет

$$\omega^* = \left(\sum_{i=1}^L \pi_i / \omega_i^* \right)^{-1}. \quad (2.8)$$

Введенные показатели производительности выражают значение потенциально возможной производительности ЭВМ, причем номинальное быстродействие (2.5) и быстродействие по Гибсону (2.6) информируют о потенциально возможной производительности при условии, что ЭВМ исправна, а среднее (2.7) и среднее эффективное (2.8) быстродействия – при условии, что машина исправна и занята решением задач (“полезной работой”). Ясно, что показатели (2.7) и (2.8) зависят от набора эталонных тестов.

При техническом описании ЭВМ, нередко ограничиваются оценкой простейших показателей производительности, таких как тактовая частота, время выполнения операций, числа операций сложения с фиксированной или плавающей запятой, выполняемых в секунду, и т.п. Показатель производительности по Гибсону был популярен при анализе возможностей ЭВМ третьего поколения (например, семейств IBM и ЕС ЭВМ).

Показатель типа (2.7) используется и при анализе производительности параллельных вычислительных систем (ВС). В этом случае берутся наборы параллельных эталонных тестов, например, таких как NAS Parallel Banchmark EP, NAS Parallel Banchmark FFT, NAS Parallel Banchmark CG.

2.6.3. Единицы измерения производительности ЭВМ

Для измерения производительности современных ЭВМ используются крупные единицы, которые на несколько порядков отличаются от базовых: герц и операция в секунду.

Для измерения тактовой частоты ЭВМ используют МегаГц или МГц (MegaHz или MHz), здесь Мега = М = 10^6 , а также ГигаГц (GigaHz), 1 ГигаГц = 10^9 Гц.

Для оценки номинального быстродействия (2.5) и быстродействия ЭВМ по Гибсону (2.6) в случае, когда учитываются только операции с фиксированной запятой, применяются следующие единицы измерения:

- MIPS (Million of Instructions Per Second), равная 10^6 опер./с;
- GIPS, соответствующая 10^9 опер./с.

Измерение производительности на тестовых наборах задач осуществляется при помощи следующих единиц:

- 1 FLOPS (FLoating-point Operations Per Second), 1 операция с плавающей запятой в секунду;
- 1 MegaFLOPS = 1 MFLOPS = 10^6 опер./с = 1 миллион операций с плавающей запятой в секунду;
- 1 GigaFLOPS = 1 GFLOPS = 10^9 опер./с = 1 миллиард опер./с;
- 1 TeraFLOPS = 1 TFLOPS = 10^{12} опер./с = 1 триллион опер./с;
- 1 PetaFLOPS = 1 PFLOPS = 10^{15} опер./с = 1 квадриллион опер./с.