Елена Михайловна Иванова, доц., к.т.н.,

КРИТЕРИИ И ОСОБЕННОСТИ СРАВНИТЕЛЬНОЙ ОЦЕНКИ ПРОИЗВОДИТЕЛЬНОСТИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Огромное многообразие различных вычислительных систем (ВС) требует создания критериев и инструментов их сравнительной оценки. Как правило, оцениваются и сравниваются технико-эксплуатационные характеристики ВС, такие как: производительность, отношение производительности к стоимости, ёмкость и тип разных видов памяти, тактовая частота, разрядность процессора и разрядность шин, возможности системы команд, программная совместимость, энергопотребление и др. Сравнительная оценка различных ВС по большинству критериев не представляет никакой сложности, т.к. достоверная информация предоставляется разработчиком или продавцом изделия. Самой большой трудностью, вызывающей наибольшее количество споров и разночтений, является оценка производительности ВС. Даже для одного однопроцессорного компьютера эта характеристика не является величиной постоянной, а тем более для сложной ВС, включающей в свой состав несколько процессоров, ведущих обработку информации с разной скоростью.

Для сравнительной оценки различных ВС применяется множество гочисленные исследования в данной области породили множество различных понятий и характеристик: быстродействие, производительность (пиковая, предельная, номинальная, эксплуатационная, системная, эффективная, максимальная), вычислительная мощность, вычислительная способность (сapability computing), вычислительная ёмкость (capacity computing), рейтинг MIPS, рейтинг FLOPS, время ЦП (CPU time) и др. Каждый из критериев подразумевает свои особенности методологии оценки различных ВС и их классов. Но в итоге все пытаются количественно выразить скорее качественное свойство, которое достигается и за счет применяемых технологий, и за счёт свойств архитектуры ВС.

Самым распространенным критерием производительности ВС сегодня являются две величины пиковая или предельная производительность и реальная производительность, измеряемые в MIPS (или FLOPS) – числе инструкций (или операций), выполняемых системой в единицу времени.

Пиковая или предельная производительность *V*пик, измеренная в MIPS, характеризует скорость работы процессора по выполнению собственных инструкций без учёта времени обращения к оперативной памяти и к устройствам ввода/вывода. Определяется эта величина как среднестатистическое число инструкций (целочисленных операций типа «регистр-регистр»), выполняемых в единицу времени (секунду) без учёта их статистического веса в выбранном классе задач (1):

где  *f* – тактовая частота процессора (Гц=такт/сек),  – средняя длительность инструкции (такт/инстр), *t* и *Т* – длительность выполнения программы, измеренное в секундах и тактах.

Под пиковой производительностью многопроцессорной ВС понимают величину, равную произведению пиковой производительности одного процессора на число *N* таких процессоров в системе (2). При этом предполагается, что все устройства ВС работают в максимально производительном режиме.

Как видно из (1) и (2), на критерий влияют не только характеристики системы ( *f*), но и характеристики программы (). Измерение производительности в MIPS (или рейтинг MIPS) имеет свои достоинства и недостатки (даже опасность) [1].

**Достоинствами** такого способа оценки производительности является то, что эта характеристика понятна пользователю и её легко рассчитать, и как правило, более быстрая машина будет характеризоваться большим рейтингом MIPS. **К недостаткам** методики можно отнести следующее. Рейтинг MIPS зависит от набора команд процессора, что затрудняет сравнение по MIPS компьютеров, имеющих разные системы команд, рейтинг MIPS даже на одном и том же компьютере меняется от программы к программе, рейтинг MIPS для системы, обладающей качественно лучшими характеристиками производительности, может быть количественно меньше по отношению к качественно худшим системам.

Для иллюстрации последнего случая исследователями рассматриваются, например, две системы: первая – в которую входит сопроцессор для обработки чисел с плавающей точкой (FPU), вторая – без него. При отсутствии сопроцессора операции над числами с плавающей точкой (вещественными числами) реализуются с помощью специализированных подпрограмм, включающих команды целочисленной арифметики. Производительность первой системы, безусловно, выше, т.к. для обработки одной вещественной операции в ней используется всего одна команда сопроцессора, взамен подпрограммы, включающей до нескольких десятков целочисленных команд в системах второго типа. Но рейтинг MIPS для первой системы будет меньше, т.к. в программах сокращается количество быстрых целочисленных операций (выполняемых за меньшее число тактов) и увеличивается количество более медленных вещественных операций (выполняемых за большее число тактов). Для таких систем среднее время выполнения команды будет больше, и, как следствие, они имеют более низкий рейтинг MIPS – величину, обратную среднему времени выполнения команды. Подобные странности наблюдаются и при использовании оптимизационных техник для компиляторов, когда в результате оптимизации производительность увеличивается, а рейтинг MIPS уменьшается, т.к. сокращается количество выполняемых в программе команд.

Рассмотрим другую единицу измерения производительности – FLOPS (или рейтинг FLOPS). Идея такого критерия возникла, когда сравнивали различные типы компьютеров для решения научно-технических задач. В подобных алгоритмах велика доля вычислений с действительными числами, для которых используется тип данных с плавающей точкой (float).

Если использовать единицы измерения FLOPS (вместо числа инструкций MIPS), то в приведённых выше формулах размерность [инстр] заменяется на [оп] и рассчитывается новое (3):

Отметим **достоинства** этого метода. Рейтинг FLOPS предназначен для более справедливого сравнения различных машин между собой, т.к. зависит от количества выполненных операций, а не от количества выполненных машинных команд. Многие программисты полагают, что одна и та же программа, реализованная на различных компьютерах, может выполняться за различное количество команд, но всегда за одно и то же количество вещественных операций. **Недостатки** измерения производительности во FLOPS тоже достаточно очевидны.

1. Эта характеристика предназначена для оценки производительности программ, состоящих из вещественных операций, а значит, плохо подходит для другого класса задач. Рейтинг FLOPS будет изменяется в зависимости от доли целочисленных операций по сравнению с долей операций с плавающей точкой в программе, так и от доли быстрых и медленных операций с плавающей точкой. Например, программа со 100% операций сложения (быстрые операции) будет иметь более высокий рейтинг, чем программа со 100% операций деления (медленные операции). Для программ-компиляторов всегда получим результат, близкий к нулю, т.к. они в основном используют команды обработки строк, логические и команды специальной арифметики и редко содержат вещественных операции.

2. Наборы вещественных операций, как и системы команд различных компьютеров в целом, не всегда совместимы. Исследователи приводят такой наглядный пример. В суперкомпьютерах фирмы Cray Research, чья высокая производительность не подвергается сомнению, рейтинг FLOPS может оказаться низким, для программ, содержащих большую долю операций деления. Вместо одной операции деления (которая в этих ВС отсутствует) используется две операции: вычисления обратной величины числа с плавающей точкой и умножения делимого на обратную величину.

Рассмотрим другую величину оценки производительности ВС - реальную производительность *R*РЕАЛ. Строго говоря, пиковая производительность есть величина теоретическая и недостижимая для реального приложения (в отсутствии конфликтов обращения к памяти при равномерной загрузке всех устройств). Реальная производительность ВС, достигаемая на данной программе (приложении), зависит также от взаимодействия программной модели, в которой реализовано приложение, с архитектурными особенностями машины, на которой это приложение запускается.

Вывести теоретически формулу для расчёта реальной производительности с учётом всех вышеназванных факторов не представляется возможным. В связи с этим многие исследователи [1,2,3] склоняются к использованию в качестве единицы измерения производительности компьютера **время** (в секундах) выполнения реальных программ: компьютер, выполняющий тот же объем работы за меньшее время является более быстрым (меньшее время подразумевает большую производительность). В качестве такой программы используется широко распространённый тест LINPACK [4] с известным количеством операций nLINPACK. Тогда реальную производительность ВС можно оценить, как

Для многопроцессорных систем также имеются параллельные версии LINPACK. Этот тест достаточно прост и понятен, и используется для формирования списка Top500 – пятисот самых мощных компьютеров мира [2]. Однако LINPACK имеет существенный недостаток: программа распараллеливается, поэтому невозможно оценить эффективность работы коммуникационного компонента суперкомпьютера. Используется множество других тестовых программ: Livermore Loops, Dhrystone, NAS Kernel Benchmark Program, ScaLAPACK, NAS Parallel Benchmarks, Perfect Club, ParkBENCH, SLALOM, HINT[4,5,6]. Обилие тестовых программ привело к тому, что в конце 20 века была учреждена корпорация SPEC (Standard Performance Evaluation Corporation) [6]. Основной целью этой организации является разработка, поддержка и публикация стандартизованного набора специально подобранных тестовых программ для оценки производительности новейших поколений высокопроизводительных компьютеров.

Однако окончательного решения проблема оценки производительности ВС пока не получила. По этому поводу часто цитируется доклад и статья Дэвида Бэйли от 1991г. [7] о способах обмана при различных измерениях производительности ВС. За прошедшее время количество тестовых программ, а значит и способов этого своеобразного жульничества только увеличилось.

В заключении хотелось бы отметить, что в будущем во избежание недоразумений, ошибочных оценок, спекуляций следует использовать оценки мощности ВС одновременно по нескольким различным показателям. Это могут быть и выше названные показатели, но также возможно появление новых измеряемых или рассчитываемых величин в качестве таких оценок. Например, в публикуемых списках рейтинга самых мощных компьютеров мира «Top500» приводятся три показателя – *R*PEAK, *R*MAX и эффективность (в %) использования вычислительной мощности системы. *R*PEAK вычисляется как сумма производительности отдельных процессоров, то есть это уже известная теоретическая оценка – предельная производительность, вычисленная во FLOPS (3). Второй показатель *R*MAX (также во FLOPS) – измеренная мощность (наивысший результат, полученный при использовании системы тестов LINPACK), т.е. по формуле (4). Недостатки таких расчётов и измерений мы уже оговорили. А вот третий показатель – эффективность – относительно новый. Он вычисляется, как отношение измеренной производительности к расчётной:

,

и представляется наиболее справедливым, т.к. даёт представление о реальной загрузке разработанной ВС, по сравнению с её гипотетическими возможностями.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Шнитман В. Современные высокопроизводительные компьютеры, информационно-аналитические материалы Центра Информационных Технологий, 1996 г. (http://citforum.ru/hardware/svk/contents.shtml)
2. Top500 – cписок 500 самых быстрых в мире (http://www.top500.org/)
3. Таненбаум Э. Архитектура компьютера. – 5-е изд. – СПб.: Питер, 2011.
4. Тесты производительности процессора (http://www.parallel.ru/computers/benchmarks/perf.html)
5. Сайт Computational Science Alliance / раздел Сравнительная производительность (http://www.csa.ru/CSA/performance1.shtmr)
6. Сайт корпорации SPEC (http://www.spec.org/)
7. David H. Bailey. Twelve Ways to Fool the Masses When Giving Performance Results on Parallel Computers, Ref: Supercomputing Review, Aug. 1991, pg. 54-55.