6: Performance Measures

Kusum Lata Dhiman
Assistant Professor
Computer Science & Engineering

Module Topics

Performance measures: Speedup, efficiency, and scalability. Abstract performance metrics (work, critical paths), Amdahl`s Law, abstract vs. Real performance (granularity, scalability)

Performance measures: Speedup, efficiency, and scalability

- Performance measures in High-Performance Computing (HPC) are essential for assessing the effectiveness and efficiency of a computing system or application when dealing with computationally intensive tasks. Speedup, efficiency, and scalability are indeed key performance measures in HPC:

1. Speedup: Speedup measures how much faster a parallel program or algorithm runs on multiple processors or cores compared to running it on a single processor. It is calculated using the formula:

- Speedup = T(1) / T(p)
- where T(1) is the execution time on a single processor, and T(p) is the execution time on p processors. A speedup greater than 1 indicates that the parallelization has resulted in a performance improvement.

Performance measures: Speedup, efficiency, and scalability

2. Efficiency: Efficiency is a measure of how effectively the available computing resources are utilized in a parallel computation. It is often expressed as a percentage and is calculated using the formula:

- Efficiency (%) = (Speedup / p) * 100
- where p is the number of processors or cores used. Efficiency measures how well the parallel algorithm scales with the number of processors. Higher efficiency values indicate better utilization of resources.

Performance measures: Speedup, efficiency, and scalability

- **Scalability:** Scalability assesses how well a parallel program or system can handle an increasing workload or a growing number of processors without a significant degradation in performance. It is typically evaluated in terms of strong scalability and weak scalability:
- **Strong Scalability:** Strong scalability measures how well a program performs as the problem size remains constant while the number of processors increases. It assesses the ability to solve a fixed-size problem faster as more processors are added.
- **Weak Scalability:** Weak scalability evaluates the performance of a program as both the problem size and the number of processors increase proportionally. It assesses the ability to handle larger and more complex problems as more resources are added.

Performance measures: Speedup, efficiency, and scalability

- In HPC, achieving high speedup, efficiency, and scalability is crucial for making the most of the available computational resources. These metrics help researchers and engineers optimize algorithms and parallel computing systems to tackle large-scale scientific simulations, data analysis, and other computationally intensive tasks effectively.

Amdahl`s Law

- Amdahl's Law is a fundamental principle in computer architecture and parallel computing that was formulated by Gene Amdahl in 1967. It addresses the potential speedup that can be achieved by parallelizing a computation or task. The law is often used to understand the limitations of parallel processing and to make informed decisions about how to allocate resources for parallelization.

Amdahl`s Law

- Amdahl's Law is expressed as a mathematical formula:

$$S = \frac{1}{F + \frac{1-F}{N}}$$

Where:

- **S** is the speedup achieved by parallelizing a task.
- **F** is the fraction of the task that cannot be parallelized (the sequential portion).
- **N** is the number of processing units or processors used for parallel execution.

Amdahl`s Law

Key points to understand about Amdahl's Law:

- 1. **Limitation of Parallelization:** Amdahl's Law highlights that the potential speedup of a computation is limited by the sequential portion of the task. No matter how many processors are added, if a significant portion of the task cannot be parallelized (F is close to 1), the overall speedup will be limited.
- **Diminishing Returns:** As the number of processors (N) increases, the potential speedup approaches a maximum value, which is limited by the sequential portion (F). Adding more processors beyond a certain point doesn't yield significant improvements in overall execution time.

Abstract vs. Real performance (granularity, scalability)

1. Abstract Performance:

- Abstract performance metrics focus on high-level, theoretical, or simplified measures of system performance. These metrics provide a general understanding of how a system should perform under ideal or controlled conditions but may not always reflect real-world performance accurately. Some characteristics of abstract performance metrics include:
- **Simplicity:** Abstract performance metrics are often simple to calculate and understand. They provide a quick and straightforward way to assess a system's performance.
- **Theoretical:** These metrics are based on theoretical models or assumptions and may not consider real-world complexities or variations.

Abstract vs. Real performance (granularity, scalability)

- **Idealized Conditions:** Abstract performance metrics typically assume idealized conditions where external factors such as network latency, hardware failures, and other unpredictable variables are ignored.
- **Granularity:** Abstract performance metrics may lack granularity and may not capture performance variations at different levels of system operation.

Abstract vs. Real performance (granularity, scalability)

2. Real Performance:

- Real performance metrics, on the other hand, provide a more accurate and detailed assessment of how a system performs in real-world, production environments. These metrics take into account the complexities and variability of actual system usage. Key characteristics of real performance metrics include:
- **Complexity:** Real performance metrics often involve more complex measurements and analysis. They consider factors such as system load, resource contention, and user behavior.

Abstract vs. Real performance (granularity, scalability)

- **Empirical Data:** Real performance metrics are based on empirical data collected from real-world system usage. They reflect the actual experiences of users and the system's response to varying conditions.
- **Real-world Conditions:** These metrics consider the impact of real-world conditions, including network latency, hardware failures, and concurrent user activities.
- **Granularity:** Real performance metrics offer greater granularity by measuring performance at various levels of the system, from individual components to the system as a whole.

Abstract vs. Real performance (granularity, scalability)

3. Granularity:

- Granularity refers to the level of detail or resolution at which performance metrics are measured and reported. Abstract performance metrics often have a coarse granularity, providing a high-level overview of system performance, while real performance metrics offer finer granularity, allowing for more detailed and nuanced insights.

Abstract vs. Real performance (granularity, scalability)

4. Scalability:

- Scalability is the ability of a system to handle increasing workloads or users while maintaining or improving performance. Abstract scalability assessments may make simplified assumptions about scalability potential, whereas real scalability evaluations involve testing and measuring a system's performance under realistic load conditions. Real scalability assessments are essential for identifying bottlenecks and optimizing system resources to ensure it can grow effectively.

Parul® University

NAAC GRADE A++

https://paruluniversity.ac.in/