Information Communication Technologies

Lecture 4. Improving Computer Performance

Kassymova Aizhan Bakhytzhanovna

PhD, Associate professor

a.kassymova@satbayev.university

Agenda

1 Moore's Law

2 Bottlenecks

3 Throughput and Latency

Moore's Law

- Law can be stated as:

  Number of transistors on a microchip doubles every 18 months.

- Denser the chip => Higher the capacity
- Limitations: Chips must be thick enough for the electrons to pass through.
- Predictions based on Moore's Law
  - Processing power (speed) doubles every 18 months.
  - Storage capacity of RAM increases exponentially.

Moore's Law

- Other observations:
  - Storage capacity of hard disk drives is also increasing exponentially.
  - Cost for consumers to purchase computer parts is decreasing over time.

Parkinson's Law of Data

- **Parkinson's Law of Data:** *Data expands to fill the space available.*
    - As more memory or disk space becomes available, the demand for more memory or disk space increases accordingly.
    - As Parkinson's Law predicts, today's operating systems are much more elaborate and require more memory for their own use.
    - As disk drive capacity increases, people begin using them in new ways (e.g. storing musical recordings, short video clips, and movies).

Bottleneck

- ***Bottleneck*** is a step within a series of steps that takes the longest time to complete.
- Time required to perform a task consisting of several steps may be delayed by the bottleneck step.
- Process time cannot be shortened without speeding up the bottleneck.

Bottleneck

- Typical bottlenecks:
  - Cache
  - RAM
  - I/O
  - Video card (particularly for 3-D gaming)
- To speed up performance of a system:
  - Use profiling tools to measure each section's time taken to complete to determine the bottleneck steps
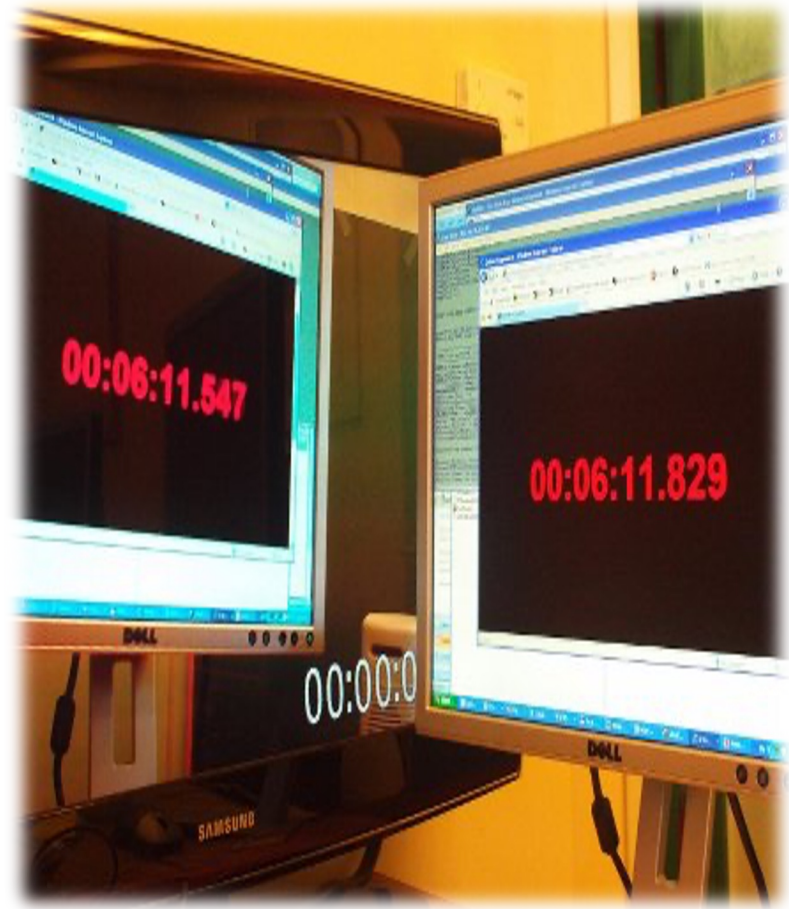 - Improve upon the bottleneck steps

Throughput
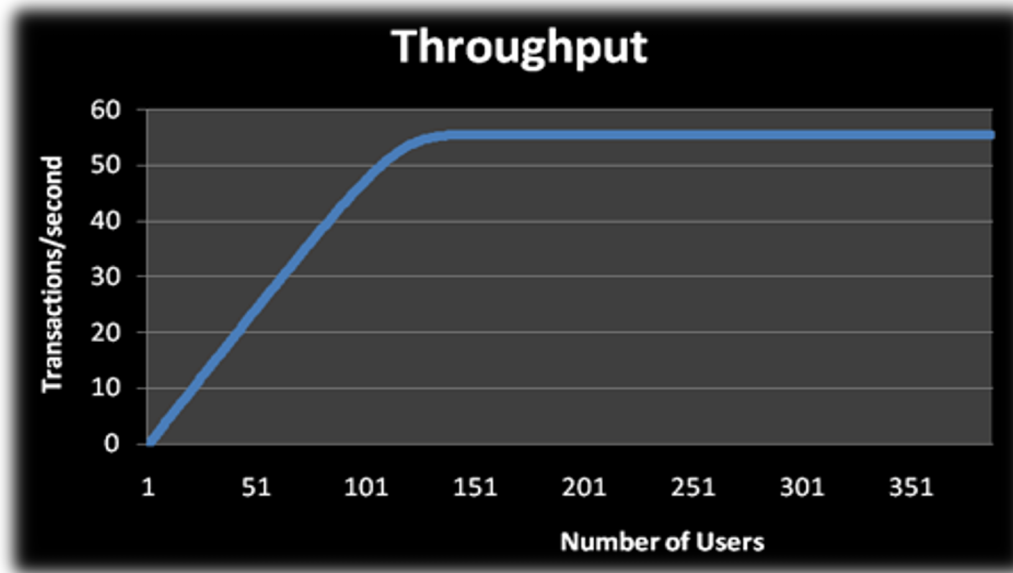
Measure of a computer system's overall performance in sending data through all its components, such as the processor, buses, storage devices. Throughput is more meaningful indicator of system performance than raw clock speed (now measured in gigahertzs) advertised by computer vendors.

Latency

The period of time that one component in a system is spinning its wheels waiting for another component.

Latency, therefore, is wasted time.

For example, in accessing data on a disk, latency is defined as the time it takes to position the proper sector under the read/write head.

End of Lecture 4