5

Big Idea 4: Computer Systems and Networks

"New knowledge is the most valuable commodity on earth. The more we have to work with, the richer we become."

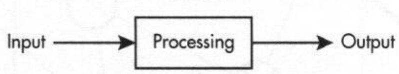
—Kurt Vonnegut Jr.

Chapter Goals

- ➔ Computing device
- ➔ Internet
- ➔ Packets
- ➔ Paths
- ➔ Fault tolerance
- ➔ Redundancy
- ➔ Serial computing
- ➔ Parallel computing
- ➔ Efficiency

Computing Devices

A computing device is a physical artifact that can run a program. Some examples include computers, tablets, servers, routers, and smart sensors. The device must be able to take inputs, process the inputs, and then calculate results based on those inputs.

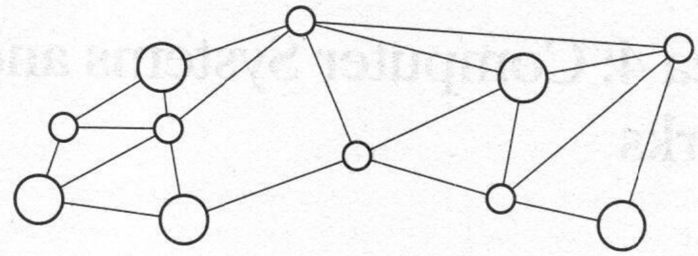


Some examples of a computing input device include but are not limited to a keyboard, mouse, scanner, microphone, and webcam. Some examples of a computing output device include a monitor, printer, and speakers.

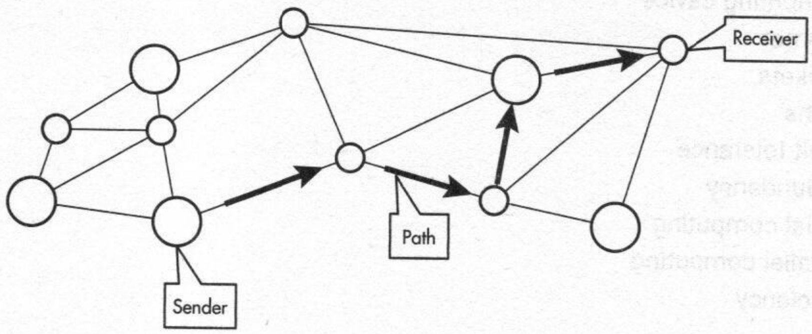
A computer is a computing device, but not all computing devices are computers. A computing device could be a calculator, printer, digital camera, smartphone, computer, or many more.

A computing system is a group of computing devices and programs working together for a common purpose. A type of computing system is a computer network. A computer network is a group of interconnected computing devices capable of sending or receiving data.

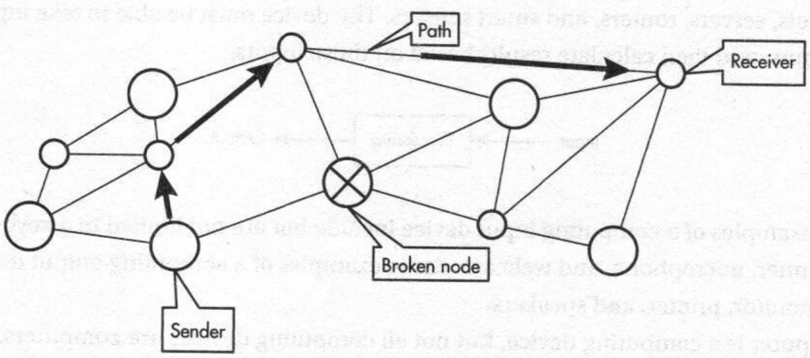
Autonomous Systems of the Internet



Each circle in the diagram above represents a computer system that is connected to another computer system, forming a computer network. The larger circles represent systems that have a higher bandwidth capacity measured in bits per second. The bandwidth of a computer network is the maximum amount of data that can be sent in a fixed amount of time.



A path between two computing devices on a computer network (a sender and receiver) is a sequence of directly connected computing devices that begins at the sender and ends at the receiver. Routing is the process of finding a path from sender to receiver.



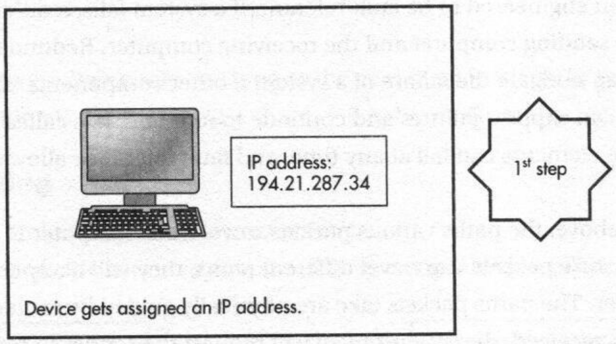
If the path from sender to receiver is broken, the path will be rerouted. This fault-tolerant nature of the internet makes connections between computing devices more reliable.

The internet connects devices and networks from all over the world. The internet is a physical network of fiber optics, radio transmitters, and cabling. Devices and networks that make up the internet are connected and communicate using standardized, open communication protocols. A protocol is an agreed-upon set of rules that specify the behavior of a system. These internet protocols, including those for addresses and names, have evolved to allow for the internet to be scalable. The scalability of a system is the capacity for the system to change in size and scale to meet new demands.

Internet Protocol

Internet protocol (IP) is responsible for addressing and routing your online requests. For a device to connect to the internet, it is first assigned an internet protocol address. When the internet was founded in the 1960s, the creators did not predict the need for billions and billions of IP addresses. Currently, we are switching between the fourth and sixth versions of the internet protocol. The fourth version (IPv4) uses 32 bits to store IP addresses. These 32 bits can hold 2^{32} IP addresses. When multiplied out, 2^{32} is actually 4,294,967,296 unique addresses. The newer version, IPv6, uses 128 bits, which can hold 2^{128} IP addresses. This is equivalent to approximately 34,000,000,000,000,000,000,000,000,000,000,000,000 unique addresses.

Switching from IPv4 to IPv6 is an increase in the capacity by $2^{128} - 2^{32} = 2^{96}$ times.



Device gets assigned an IP address.

This IP address is not permanently assigned to a user's device and can change at any time.

Transmission Control Protocol

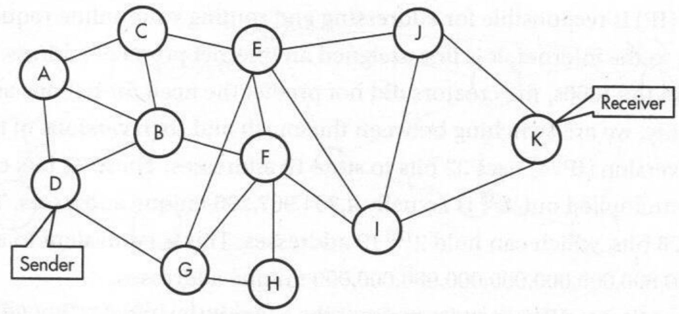
Transmission control protocol (TCP) is a protocol that defines how computers send packets of data to each other. Data traveling in the internet is broken down into small chunks of data called packets. TCP protocols guide the rules on how data are subdivided into packets before transmission.

Trailer End of packet	Data	Header: Sender's IP address Receiver's IP address Packet #

User Datagram Protocol

User datagram protocol (UDP) is a protocol that allows computer applications to send messages without checking for missing packets to save on time needed to retransmit missing packets. UDP is not as reliable as TCP, which does resend packets lost when transmitting.

Fault Tolerance



The internet has been engineered to be fault tolerant. If a system fails, a different path can be chosen between the sending computer and the receiving computer. Redundancy is the inclusion of extra paths that can mitigate the failure of a system if other components fail.

When a system can support failures and continue to function, it is called *fault tolerant*. This is important because elements can fail at any time, and fault tolerance allows users to continue to use the network.

In the diagram above, the paths various packets travel from computer D to computer K will not be the same. Because packets can travel different paths, they will likely arrive at the target computer out of order. The paths packets take are physically guided by routers.

If a packet is not received, the TCP protocol will request the sender to resend the missing packet. The IP addresses of both the sender and receiver are found in the header of the packet. When all packets are received, the packets are put together using the packet numbers found in the header to form the original binary message.

The process from computer D to computer K in the diagram above is called *end-to-end architecture*. This process involves the breaking down and assembling of the packets at each end. What happens to the packets in the middle is hidden from the user in an abstraction.

Difference Between Internet and World Wide Web

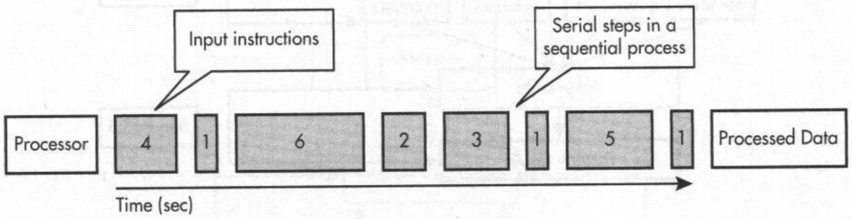
The internet refers to the hardware. It is made up of the computers, cables, routers, and many more components that make up the entire network. It is a global decentralized network connecting millions of computers. The World Wide Web, in contrast, refers to the software used on the internet. HTTP is a protocol used by the World Wide Web to transmit data. The internet allows access to the World Wide Web, which is a system of linked pages, programs, and files.

Efficiency of Solutions

Sequential computing is a computational model in which operations are performed in order one at a time. A sequential solution takes as long as the sum of all of the steps. In contrast, parallel computing involves breaking up a task into smaller, sequential pieces. Then those sequential pieces are all executed at the same time, each on its own processor or on a set of computers that have been networked together. A parallel solution takes at least as long as the longest branch in the program.

Sequential Computing

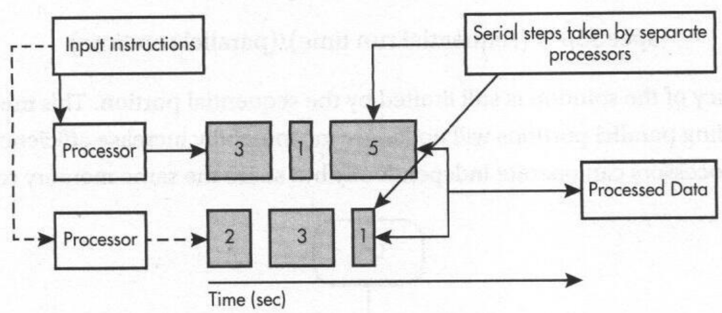
- A problem is broken into discrete instructions.
- These instructions are executed one by one by a single computing device having a single central processing unit (CPU).



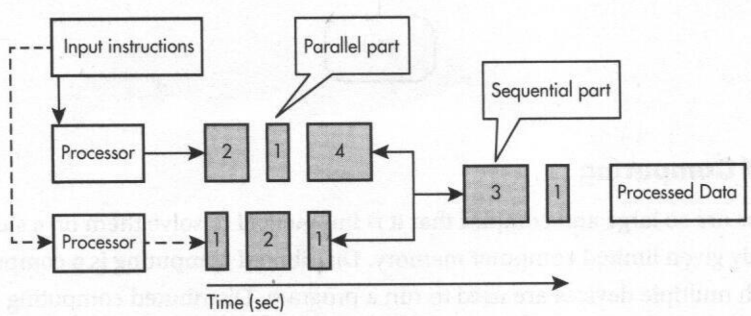
A sequential computing solution takes as long as the sum of all its steps. In the above example, the total processing time is $4 + 1 + 6 + 2 + 3 + 1 + 5 + 1 = 23$ seconds.

Parallel Computing

- A problem is broken into discrete instructions.
- These instructions are executed concurrently by using multiple CPUs.

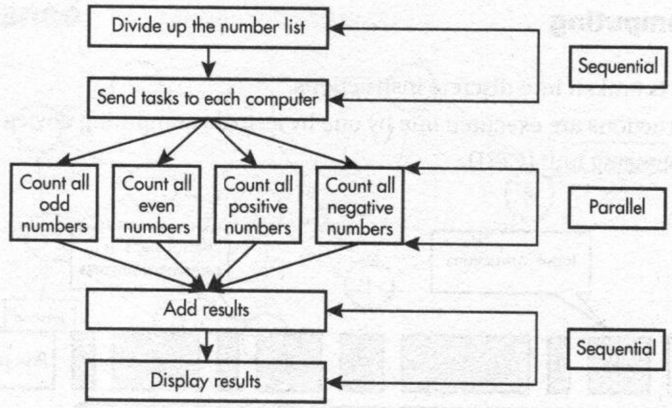


A parallel computing solution takes as long as the longest of the tasks done in parallel. In the above example, the total processing time is 9 seconds.



Parallel computing can consist of a parallel portion and a sequential portion. A parallel computing solution takes as long as its sequential tasks plus the longest of its parallel tasks. Most modern computers are parallel in architecture with multiple processors.

In the above example, the total processing time is $2 + 1 + 4 + 3 + 1 = 11$ seconds.



Why Is Parallel Computing Used?

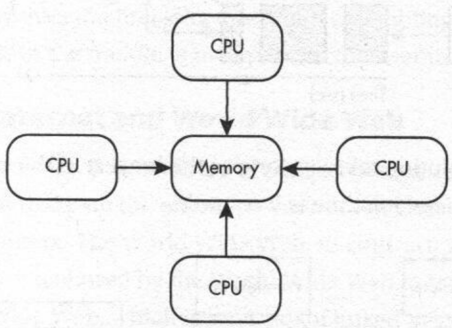
- In the real world, many things happen at the same time in different places concurrently.
- Parallel computing is needed for real-world simulations and modeling.

Some examples of complex computer simulations that benefit from parallel computing include weather forecasting, flight simulators, car crash modeling, seismic surveying, and so on. The "speedup" of a parallel solution is measured in the time to complete the task sequentially divided by the time to complete the task when done in parallel:

Speedup = (sequential run time)/(parallel run time)

The efficiency of the solution is still limited by the sequential portion. This means that at some point, adding parallel portions will no longer meaningfully increase efficiency.

Multiple processors can operate independently but share the same memory resources.



Distributed Computing

Many problems are so large and complex that it is impractical to solve them on a single computer, especially given limited computer memory. Distributed computing is a computational model in which multiple devices are used to run a program. Distributed computing allows problems to be solved that could not be solved on a single computer because of either the processing time or storage needs involved. Much larger problems can be solved more quickly using distributed computing than using a single computer.

Parallel computing uses a single computer with multiple processors. Distributed computing uses multiple computing devices to process those tasks.