WIKIPEDIA
The Free Encyclopedia

# Network congestion

**Network congestion** in data networking and queueing theory is the reduced quality of service that occurs when a network node or link is carrying more data than it can handle. Typical effects include queueing delay, packet loss or the blocking of new connections. A consequence of congestion is that an incremental increase in offered load leads either only to a small increase or even a decrease in network throughput.[1]

Network protocols that use aggressive retransmissions to compensate for packet loss due to congestion can increase congestion, even after the initial load has been reduced to a level that would not normally have induced network congestion. Such networks exhibit two stable states under the same level of load. The stable state with low throughput is known as **congestive collapse**.

Networks use **congestion control** and **congestion avoidance** techniques to try to avoid collapse. These include: exponential backoff in protocols such as CSMA/CA in 802.11 and the similar CSMA/CD in the original Ethernet, window reduction in TCP, and fair queueing in devices such as routers and network switches. Other techniques that address congestion include priority schemes which transmit some packets with higher priority ahead of others and the explicit allocation of network resources to specific flows through the use of admission control.

## Network capacity

Network resources are limited, including router processing time and link throughput. Resource contention may occur on networks in several common circumstances. A wireless LAN is easily filled by a single personal computer.[2] Even on fast computer networks, the backbone can easily be congested by a few servers and client PCs. Denial-of-service attacks by botnets are capable of filling even the largest Internet backbone network links, generating large-scale network congestion. In telephone networks, a mass call event can overwhelm digital telephone circuits, in what can otherwise be defined as a denial-of-service attack.

## Congestive collapse

Congestive collapse (or congestion collapse) is the condition in which congestion prevents or limits useful communication. Congestion collapse generally occurs at choke points in the network, where incoming traffic exceeds outgoing bandwidth. Connection points between a local area network and a wide area network are common choke points. When a network is in this condition, it settles into a stable state where traffic demand is high but little useful throughput is available, during which packet delay and loss occur and quality of service is extremely poor.

Congestive collapse was identified as a possible problem by 1984.[3] It was first observed on the early Internet in October 1986,[4] when the NSFNET phase-I backbone dropped three orders of magnitude from its capacity of 32 kbit/s to 40 bit/s,[5] which continued until end nodes started implementing Van Jacobson and Sally Floyd's congestion control between 1987 and 1988.[6] When more packets were sent than could be handled by intermediate routers, the intermediate routers discarded many packets, expecting the end points of the network to retransmit the information.