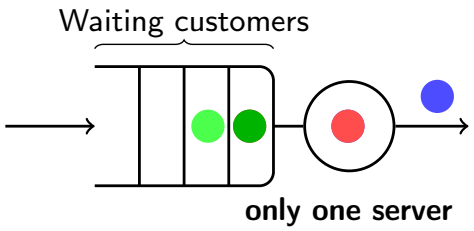
What is a queue?

Servers and waiting buffer

Number of servers

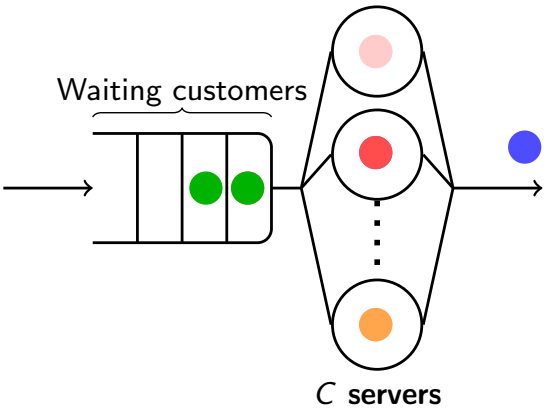
In a single server system, as soon as the server is freed, one customer among those waiting uses it (if the queue is not empty).



Another characteristic of a queuing system is the number of servers. There can be a single server or multiple servers.
In a single-server system, as soon as the server becomes free, one customer among those waiting uses it (assuming that the waiting buffer is not empty). In this example, when the blue customer leaves the system the first customer in the waiting buffer, the red one, starts to be served.

Number of servers

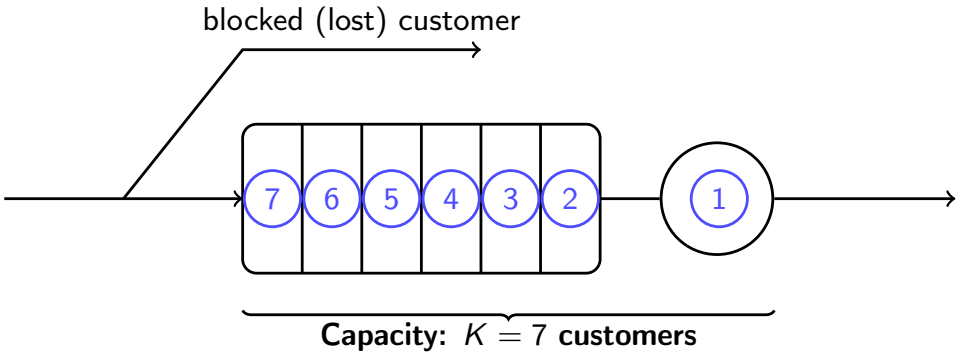
In a system with multiple servers, there cannot be any customers waiting if at least one server is idle. As soon as one server is freed, a customer among those waiting uses it (if the waiting buffer is not empty).
It is generally assumed that servers are identical, so that the distribution of service durations is the same for all servers.



In a system with multiple servers there are several servers which can handle customers simultaneously. In that case there cannot be any customers in the waiting buffer unless all the servers are busy. If there are customers in the waiting buffer one of them starts being served as soon as one of the servers is released. In general it is assumed that the servers are identical and the different service durations are therefore statistically indistinguishable. They follow the same probability law.

Maximum number of customers

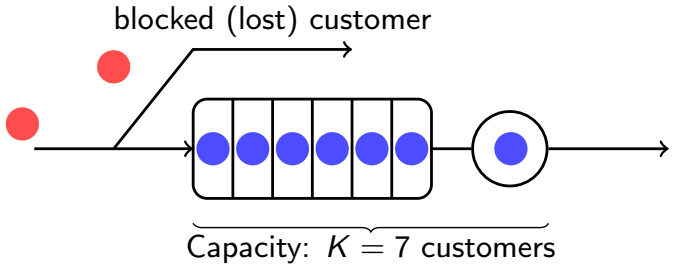
The maximum number of customers in the system is the sum of the number of positions in the waiting buffer and the number of servers. Example:



To end, one has to define the maximum number of customers in the system. This includes the servers and the number of positions in the waiting buffer. In this example the maximum number of customers is 7 since there can be one client being served and up to 6 clients waiting.

Finite buffer

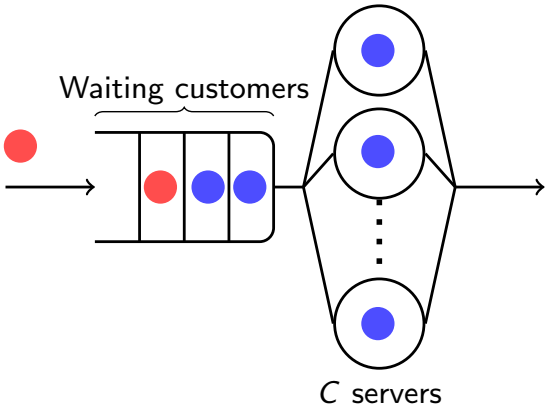
If a customer arrives while the buffer is full, this customer is lost (blocked) ⇒blocking queue



We differentiate between queues with a finite buffer and queues with an infinite buffer.
If the buffer is finite then some customers may be blocked since they may arrive when the buffer is full. These queues are called blocking systems.

Infinite buffer

If a customer arrives and all the servers are busy this customer lines up in the waiting buffer ⇒pure waiting queue



On the contrary in queuing systems with an infinite buffer customers are never blocked. If a new customer arrives and the servers are busy then the customer lines up in the waiting buffer and waits to be served. These queues are pure waiting systems.