

# Glossary for Callcenter Simulator

## Accessibility

The accessibility (on call basis) is the ratio of calls served and calls arrived at the system; in this case the term call covers all types of requests, even e-mails etc. If a client is forwarded in the process of this conversation, this second conversation is counted as a second call. Ideally, the accessibility should be close to 100%. The accessibility (on client basis) is the ratio of clients served and clients arrived at the system. A client is counted as successfully served, if he finally leaves the system successfully served (possibly after more than one unsuccessful call attempts). Therefore the accessibility on client basis is always at least as high as the accessibility on call basis.

$$\text{Accessibility} := \frac{\text{Served calls}}{\text{Arrived calls}} = \frac{\text{Served calls}}{\text{Served calls} + \text{Canceled calls} + \text{Blocked calls}}$$

[Related terms: Call, Canceled clients/calls, Client, Service level and Waiting time]

## ACD (short form for “Automated Call Distributor”)

The ACD is a computer system that manages the incoming calls and the call centers. It carries the list of the available agents and their skill level and does the assignment of the clients to the agents.

## Busy/blocking

An incoming call needed a free phone line is counted as blocked (and will be recorded in the statistics as this) if there are no free phone lines available at the time the call arrives. The caller will get the busy sign in this case. So the number of available lines limits the number of callers who can be in the system at the same time (waiting or talking to an agent). E-mails and mails are not subject to this restriction. Here any number of requests can be present in the system at any time.

[Related terms: Canceled clients/calls and Waiting cancelation]

## Call

A conversation of a client with an agent or the processing of an e-mail by an agent is counted as one call. If a client is forwarded or leaves the queue without being served and is served in a second call attempt only, this is counted in both cases as a second call. So serving a client can consist of more than one calls where some can be unsuccessful (waiting cancelations) and also more than one can be successful (if the client is forwarded to another agent while being in the system). This means in a system where each client needs two call attempts to be served successfully will have accessibility on call basis of 50%, but will have accessibility on client basis of 100%. Therefore the service parameters on client will always be higher than the parameters on call basis. At the same time the number of clients is always lower or equal than the number of calls.

$$\text{Calls} := \text{Fresh calls} + \text{Forwarding} + \text{Retry} + \text{Recall}$$

[Related terms: Client]

## **Cancellation time**

The cancellation time is the waiting time of a caller whose waiting time tolerance is exhausted before he is assigned to an agent.

[Related terms: Canceled clients/calls, Waiting cancellation, Waiting time and Waiting time tolerance]

## **Canceled clients/calls**

Clients or calls that do not advance to an agent are counted as canceled clients or calls.

The reason for such a cancellation can be that there are no free phone lines (on requests needed a phone line) at the time of the arrival (blocking) or that a calling client would need to wait longer than his waiting time tolerance allows (waiting cancellation). A client is only counted as a canceled client, if he finally leaves the system without being served (after possibly several call attempts). If a client cancels waiting during his first attempt but is served successfully on the second attempt, there is no waiting cancellation on call bases recorded but no cancellation on client bases.

[Related terms: Cancellation time and Busy/blocking]

## **Client**

A request which may need several attempts to advance to an agent or which may be forwarded multiple times is counted as one client. This means that in a system where each client needs two attempts to be successfully served, would have 50% accessibility on call basis, but would have 100% accessibility on client basis. Therefore the service parameters recorded on client basis are always higher than the parameters on call basis. At the same time the number of clients will always be lower than or at least as high as the number of calls.

[Related terms: Call]

## **Forwarding**

If a client is served by a single-skill agent, it is possible that it turns out in the course of conversation, that the concerns of the client cannot completely be served by the agent who was supposed to have the right skill for the client. In this case, the client receives a new client type, which describes the now modified concern and is re-queued. It is said that the client gets forwarded. By using some special forwarding factor when calculating the client score, it is possible to prioritize forwarded clients in the queue. This way they do not need to spend the full waiting period like fresh calls.

[Related terms: Skill level]

## **Fresh calls**

Clients arriving at the system for the first time are called fresh calls. If a client is blocked because there are no available free phone lines at the time of his arrival or if he cancels waiting and is starting a second call attempt later, he is called a retryer. The complete number of calls (consisting of the fresh calls and the call retries) is easily measured; but because the retrying clients will result from the simulation process, the number of fresh calls has to be used as input value.

$$\text{Calls} := \text{Fresh calls} + \text{Forwarding} + \text{Retry} + \text{Recall}$$

[Related terms: Retryer]

## **Holding time (or service time)**

The holding time is the time a client is talking to an agent or the time an agent needs to answer an e-mail. After the end of the conversation with the client the agent still needs some time to do some post processing before he announces himself as available again. Additionally the routing of a client to an agent can need a few seconds (this is called the technical free time). Both, the post processing time and the technical free time are not counted as part of the holding time.

$$\text{Work load} := \frac{\text{Active time}}{\text{Complete working time}} = \frac{\text{Technical free time} + \text{Holding time} + \text{Post processing time}}{\text{Idle time} + \text{Technical free time} + \text{Holding time} + \text{Post process. t.}}$$

[Related terms: Idle time, Post processing time and Technical free time]

## **Idle time**

The idle time is the period of time at which is an agent is logged into the system and is waiting for work, but there is no work. The period of time needed by the ACD to route the client to an agent is reported separately as so-called technical free time and is not counted as part of the idle time.

$$\text{Work load} := \frac{\text{Active time}}{\text{Complete working time}} = \frac{\text{Technical free time} + \text{Holding time} + \text{Post processing time}}{\text{Idle time} + \text{Technical free time} + \text{Holding time} + \text{Post process. t.}}$$

[Related terms: Holding time (or service time), Post processing time and Technical free time]

## **Minimum waiting time**

To differentiate the utilization between multiple call centers even more than what would be possible by differently prioritize them by score values, one can define so-call minimum waiting times. A client can only advance to a call center having a minimum waiting time if he has waited at least as long as the defined minimum waiting time and there was not matching agent at another call center was found during this time. Using a minimum waiting time (in addition or in opposite to using different score values), increases the effect of the load differentiation, but can have the effect that there are agents in idle state in the call center having a minimum waiting while at the same time clients which could be served by these agents would have to wait.

[Related terms: Score]

## **Post processing time**

After a conversation, it may be necessary to log some data before the agent announces himself again as available to the system. At this time the client already has left the system or is already forwarded to another agent, so the post processing time cannot be counted as part of the holding time of the client. But because the agent is not available for incoming calls at the post processing time, this period of time cannot be unconsidered either.

$$\text{Work load} := \frac{\text{Active time}}{\text{Complete working time}} = \frac{\text{Technical free time} + \text{Holding time} + \text{Post processing time}}{\text{Idle time} + \text{Technical free time} + \text{Holding time} + \text{Post process. t.}}$$

[Related terms: Holding time (or service time), Idle time and Technical free time]

## **Productivity**

In the simulation an agent is counted as active, if he is available for incoming calls or just processing a call. In praxis however not always all theoretical active agents are actually active for calls. In this case one can say the productivity is below 100%. (The idle times – in which the agent is active for incoming calls – are not counted for the reduced productivity.) Because for simulation only the real active agents are used but in displaying the statistics often the whole number of agents (including the non-productive agents) is requested, the Callcenter Simulator allows to define some productivity factor. The agents defined in the model editor are multiplied with this factor before being used in simulation.

[Related terms: (Disease-related) surcharge]

## **Recaller**

If a client, although he actually leaves the system successfully served, will later make another call, he is called to be a recaller. He is counted in this case not only as a new call but also as a new client (in opposite to a retry after a waiting time cancelation, in which case only the call counter is incremented, but the client counter not).

## **Residence time**

The residence time the total period of time a client spends in the system. It consists of the waiting time and the holding time. The post processing time after a conversation at which an agent logs the recorded data before he get available again is not part of the residence time. The residence time thus describes the time from a customer perspective.

$$\text{Residence time} := \text{Waiting time} + \text{Holding time}$$

[Related terms: Waiting time]

## **Retryer**

If a client leaves the system without been served successfully (either because already in his dial attempt ended with a busy sign and he was blocked or because his waiting time tolerance was exhausted before he was routed to an agent, and therefore canceled waiting) and later starts a second attempt, he is counted as a retryer.

The counting of the clients is not affected by a retrying client, i.e. if a client needs two attempts to be served successfully, he will be counted as one successful client; however, the number of calls is increased during his retry, i.e. the client who was successful served only in the second attempt is recorded as two calls, one unsuccessful call and one successful call.

[Related terms: Canceled clients/calls and Fresh calls]

## **Score**

Score values can be defined for both, prioritizing clients and prioritizing agents. The client score consists of a basis value for the client type and some additional values for respecting the waiting time of the actual client and his state if he was forwarded or not. The agent score consists of the score of the call center he is working in, of the score of his skill level for the current client type and of his idle time. If on arrival of a new client there are multiple agent who have the right skills to serve him, he is routed to the agent with the highest score value. If an agent gets available and there are multiple matching clients, the client with the highest score gets served next.

[Related terms: Minimum waiting time]

## **Service level**

As service level an amount of seconds is defined, in which a specific part of the clients is to be routed to an agent. This means that for this share of clients the waiting time is no longer than the specified value. Usual values are “80% of the calls in no more than 20 seconds” or even “90% of the calls in no more than 10 seconds”. The service level is reported in different ways. The service level is displayed as the ratio of calls / clients who were served within the specified time, and all served calls / clients (“based on successful calls / clients”), or the ratio of calls / clients who were served within the specified time and all calls / clients including the waiting cancelations (“based on all calls / clients”). Therefore the service level is only information about how many of calls / clients did not had to wait longer than some threshold value. It provides no information about how the distribution of the waiting times of the calls / clients for whom the service level was not reached looks. Therefore the average waiting time is generally the better parameter for describing the behavior of the waiting clients.

[Related terms: Accessibility and Waiting time]

## **Simulation runs / simulated days**

In order to obtain statistically robust results, the day described in the model is simulated several times and the mean values of all these simulation runs are calculated. Usually 100 or more repetitions of the model day are carried out. It is said in the case of 100 repetitions that 100 simulation runs were performed, or that 100 days have been simulated (although only various possible realizations of the same day were simulated).

## **Skill**

A skill is the ability of an agent to serve a particular client type. The skill consists of a holding time, a post processing time control and a score for the particular client type. The union of one or more skills forms a skill level.

[Related terms: Skill level]

## **Skill level**

A skill level is the amount of one or more skills. Each skill is the ability of an agent to serve a particular client type. Each agent has exactly one skill level and therefore can exactly serve the client types, which are defined in the skills of this skill level. If the skill level of an agent consists only of one skill, the agent is called a single-skill agent; if the skill level consists of more than one skill, the agent is called a multi-skill agent. The skill is no information on how good an agent can serve some client type. This can be configured indirectly by using different score values for the individual skills. If a client is routed to a matching single-skill agent by the ACD but has a second questing about some other topic (i.e. changes his client type during the call), the client has to be forwarded to another agent with a different skill level. If the client was routing to a multi-skill agent at first all of his concerns can be handled by this agent.

[Related terms: Forwarding and Skill]

## **(Disease-related) surcharge**

During optimization it is determined how many agents are needed per half-hour interval, to achieve specific service goals. This value can be used as a basis for staff planning. But it is generally known that in large call center systems always a certain proportion of agents will not appear for duty due to illness, therefore it is necessary to a priori slightly correct this planning value upwards, so that the actually available number of agents again would correspond to the desired design value. For this purpose a disease-related surcharge can be defined before starting the optimization. This value is not used during the optimization itself, but the number of agents is reported at the end of the simulation twice: the number of agents calculated from the optimization and the number of agents multiplied with the surcharge factor (which therefore should be  $\geq 100\%$ ).

[Related terms: Productivity]

## **Technical free time**

Between the assignment of a client to an agent by the ACD and the actual start of the conversation due to the system usually a few seconds (typical values are 3-7 seconds) will pass. For the agent his period of time is idle time because he is not working during this time; but the agent has no influence on this effect and there is already an assignment of a client to the agent so this time cannot be counted as idle time; at the same time this period of time also cannot be counted as holding time because the client will experience this time (depending on the system mostly) as waiting time and therefore still can cancel the waiting during the technical free time. Since the technical free time period is at the end of the waiting time of a client, the probability (compared to the first few seconds of his waiting time) is also relatively high that he will cancel waiting just during the technical free time.

$$\text{Work load} := \frac{\text{Active time}}{\text{Complete working time}} = \frac{\text{Technical free time} + \text{Holding time} + \text{Post processing time}}{\text{Idle time} + \text{Technical free time} + \text{Holding time} + \text{Post process. t.}}$$

[Related terms: Holding time (or service time), Idle time and Post processing time]

## **Waiting cancelation**

If the waiting time tolerance of a client is exhausted before he is routed to an agent, then one says he is canceling the waiting process. The second thing that can happen to prevent a client to get to an agent is that there are no free phone lines at the time he arrives and the client therefore gets the busy sign. Calls that are lost due to too little phone lines are recorded separately.

[Related terms: Cancellation time, Canceled clients/calls and Busy/blocking]

## **Waiting time**

The waiting time of a client is the period of time he has to wait before he can talk to an agent (or in case of an e-mail, the time if will have to wait before being processed). The waiting time of a client recorded in the statistics consists of two stages: the actual waiting time and the technical free time when the client is already assigned to an agent but the conversation has not yet been started. On many systems the client cannot distinguish between the waiting time and the technical free time.

$$\text{Residence time} := \text{Waiting time} + \text{Holding time}$$

[Related terms: Cancellation time, Residence time and Technical free time]

## **Waiting time tolerance**

The waiting time tolerance specifies the time that a client is willing to wait before giving up waiting and leaving the queue without being served. Possibly the client will start later a new trial and will be counted as a retryer. For calls, the waiting time tolerance typically is relatively short. But also requests like e-mails do have a waiting time tolerance: If a request like an e-mail stays unanswered for a too long period of time, the client will start a second request using some other medium (e.g. in form of a call).

[Related terms: Cancellation time, Waiting cancelation and Retryer]

## Abbreviations for mathematical terms

Expected value or mean	$E[\cdot]$
Standard deviation	$\text{Std}[\cdot]$
Variance	$\text{Var}[\cdot]$
Coefficient of variation	$\text{CV}[\cdot]$
Mean queue length	$E[N_q]$
Mean number of clients in the system	$E[N]$
Mean waiting time	$E[W]$
Mean residence time	$E[V]$
Cancellation probability	$P(A)$
Arrival rate	$\lambda$ (Lat.: lambda)
Service rate	$\mu$ (Lat.: mu)
Work load	$a$
Utilization	$\rho$ (Lat.: rho)
Service level (20 sec.)	$P(W < 20)$