



CHAPTER 9

Sizing Call Center Resources

Last revised on: August 27, 2008

Central to designing a Cisco Unified Contact Center (or any call center) is the proper sizing of its resources. This chapter discusses the tools and methodologies needed to determine the required number of call center agents (based on customer requirements such as call volume and service level desired), the number of Unified IP IVR ports required for various call scenarios (such as call treatment, prompt and collect, queuing, and self-service applications), and the number of voice gateway ports required to carry the traffic volume coming from the PSTN or other TDM source such as PBXs and TDM IVRs.

The methodologies and tools presented in this chapter are based on traffic engineering principles using the Erlang-B and Erlang-C models applied to the various resources in a Unified CCE deployment. Examples are provided to illustrate how resources can be impacted under various call scenarios such as call treatment (prompt and collect) in the Unified IP IVR and agent wrap-up time. These tools and methodologies are intended as building blocks for sizing call center resources and for any telephony applications in general.

What's New in This Chapter

[Table 9-1](#) lists the topics that are new in this chapter or that have changed significantly from previous releases of this document.

Table 9-1 *New or Changed Information Since the Previous Release of This Document*

New or Revised Topic	Described in:
This chapter contains no major updates for this release.	

Call Center Basic Traffic Terminology

It is important to be familiar with, and to be consistent in the use of, common call center terminology. Improper use of these terms in the tools used to size call center resources can lead to inaccurate sizing results.

The terms listed in this section are the most common terms used in the industry for sizing call center resources. There are also other resources available on the internet for defining call center terms.

In addition to the terms listed in this section, the section on the [Cisco Unified CCE Resource Calculators, page 9-7](#), defines the specific terms used for the input and output of the Unified CCE Resource Calculator, the Cisco call center sizing tool.

Also, for more details on various call center terms and concepts discussed in this document, refer to the Unified CCE product documentation available online at

<http://www.cisco.com>

Busy Hour or Busy Interval

A busy interval could be one hour or less (such as 30 minutes or 15 minutes, if sizing is desired for such smaller intervals). The busy interval occurs when the most traffic is offered during this period of the day. The busy hour or interval varies over days, weeks, and months. There are weekly busy hours and seasonal busy hours. There is one busiest hour in the year. Common practice is to design for the average busy hour (the average of the 10 busiest hours in one year). This average is not always applied, however, when staffing is required to accommodate a marketing campaign or a seasonal busy hour such as an annual holiday peak. In a call center, staffing for the maximum number of agent is determined using peak periods, but staffing requirements for the rest of the day are calculated separately for each period (usually every hour) for proper scheduling of agents to answer calls versus scheduling agents for offline activities such as training or coaching. For trunks or IVR ports, in most cases it is not practical to add or remove trunks or ports daily, so these resources are sized for the peak periods. In some retail environments, additional trunks could be added during the peak season and disconnected afterwards.

Busy Hour/Interval Call Attempts (BHCA)

The BHCA is the total number of calls during the peak traffic hour (or interval) that are attempted or received in the call center. For the sake of simplicity, we assume that all calls offered to the voice gateway are received and serviced by the call center resources (agents and Unified IP IVR ports). Calls normally originate from the PSTN, although calls to a call center can also be generated internally, such as by a help-desk application.

Servers

Servers are resources that handle traffic loads or calls. There are many types of servers in a call center, such as PSTN trunks and gateway ports, agents, voicemail ports, and IVR ports.

Talk Time

Talk time is amount of time an agent spends talking to a caller, including the time an agent places a caller on hold and the time spent during consultative conferences.

Wrap-Up Time (After-Call Work Time)

After the call is terminated (the caller finishes talking to an agent and hangs up), the wrap-up time is the time it takes an agent to wrap up the call by performing such tasks as updating a database, recording notes from the call, or any other activity performed until an agent becomes available to answer another call. The Unified CCE term for this concept is *after-call work time*.

Average Handle Time (AHT)

AHT is the mean (or average) call duration during a specified time period. It is a commonly used term that refers to the sum of several types of handle time, such as call treatment time, talk time, and queuing time. In its most common definition, AHT is the sum of agent talk time and agent wrap-up time.

Erlang

Erlang is a measurement of traffic load during the busy hour. The Erlang is based on having 3600 seconds (60 minutes, or 1 hour) of calls on the same circuit, trunk, or port. (One circuit is busy for one hour regardless of the number of calls or how long the average call lasts.) If a contact center receives 30 calls in the busy hour and each call lasts for six minutes, this equates to 180 minutes of traffic in the busy hour, or 3 Erlangs (180 min/60 min). If the contact center receives 100 calls averaging 36 seconds each in the busy hour, then total traffic received is 3600 seconds, or 1 Erlang (3600 sec/3600 sec).

Use the following formula to calculate the Erlang value:

$$\text{Traffic in Erlangs} = (\text{Number of calls in the busy hour} * \text{AHT in sec}) / 3600 \text{ sec}$$

The term is named after the Danish telephone engineer A. K. Erlang, the originator of queuing theory used in traffic engineering.

Busy Hour Traffic (BHT) in Erlangs

BHT is the traffic load during the busy hour and is calculated as the product of the BHCA and the AHT normalized to one hour:

$$\text{BHT} = (\text{BHCA} * \text{AHT seconds}) / 3600, \text{ or}$$

$$\text{BHT} = (\text{BHCA} * \text{AHT minutes}) / 60$$

For example, if the call center receives 600 calls in the busy hour, averaging 2 minutes each, then the busy hour traffic load is $(600 * 2/60) = 20$ Erlangs.

BHT is typically used in Erlang-B models to calculate resources such as PSTN trunks or self-service IVR ports. Some calculators perform this calculation transparently using the BHCA and AHT for ease of use and convenience.

Grade of Service (Percent Blockage)

This measurement is the probability that a resource or server is busy during the busy hour. All resources might be occupied when a user places a call. In that case, the call is lost or blocked. This blockage typically applies to resources such as voice gateway ports, IVR ports, PBX lines, and trunks. In the case of a voice gateway, grade of service is the percentage of calls that are blocked or that receive busy tone (no trunks available) out of the total BHCA. For example, a grade of service of 0.01 means that 1% of calls in the busy hour would be blocked. A 1% blockage is a typical value to use for PSTN trunks, but different applications might require different grades of service.

Blocked Calls

A blocked call is a call that is not serviced immediately. Callers are considered blocked if they are rerouted to another route or trunk group, if they are delayed and put in a queue, or if they hear a tone (such as a busy tone) or announcement. The nature of the blocked call determines the model used for sizing the particular resources.

Service Level

This term is a standard in the contact center industry, and it refers to the percentage of the offered call volume (received from the voice gateway and other sources) that will be answered within x seconds, where x is a variable. A typical value for a sales call center is 90% of all calls answered in less than 10 seconds (some calls will be delayed in a queue). A support-oriented call center might have a different service level goal, such as 80% of all calls answered within 30 seconds in the busy hour. Your contact center's service level goal determines the number of agents needed, the percentage of calls that will be

queued, the average time calls will spend in queue, and the number of PSTN trunks and Unified IP IVR ports needed. For an additional definition of service level within Unified CCE products, refer to the Unified CCE glossary available online at

<http://www.cisco.com>

Queuing

When agents are busy with other callers or are unavailable (after call wrap-up mode), subsequent callers must be placed in a queue until an agent becomes available. The percentage of calls queued and the average time spent in the queue are determined by the service level desired and by agent staffing. Cisco's Unified CCE solution uses a Unified IP IVR to place callers in queue and play announcements. It can also be used to handle all calls initially (call treatment, prompt and collect – such as DTMF input or account numbers – or any other information gathering) and for self-service applications where the caller is serviced without needing to talk to an agent (such as obtaining a bank account balance, airline arrival/departure times, and so forth). Each of these scenarios requires a different number of Unified IP IVR ports to handle the different applications because each will have a different average handle time and possibly a different call load. The number of trunks or gateway ports needed for each of these applications will also differ accordingly. (See the section on [Sizing Call Center Agents, IVR Ports, and Gateways or Trunks \(Inbound Call Center\)](#), page 9-12, for examples on how to calculate the number of trunks and gateway ports needed.)

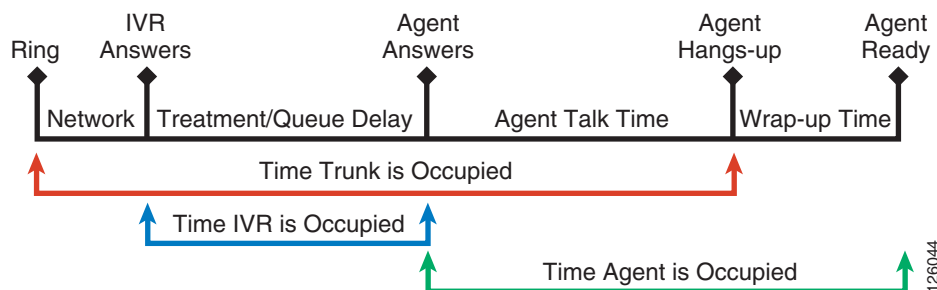
Call Center Resources and the Call Timeline

The focus of this chapter is on sizing the following main resources in a call center:

- Agents
- Gateway ports (PSTN trunks)
- Unified IP IVR ports.

It is helpful first to understand the anatomy of an inbound call center call as it relates to the various resources used and the holding time for each resource. [Figure 9-1](#) shows the main resources used and the occupancy (hold/handle time) for each of these resources.

Figure 9-1 Inbound Call Timeline



Ring delay time (network ring) should be included if calls are not answered immediately. This delay could be a few seconds on average, and it should be added to the trunk average handle time.

There are various tools and resources available for sizing the entire Unified CCE system. Using your contact center traffic data and service level requirements as input to the Unified CCE Resource Calculator, the calculator can generate output that you can feed into other tools for sizing the following Unified CCE system components and resources:

- Agents

Size the following components to support the number of agents needed for your Unified CCE system:

- Unified CCE servers — See [Sizing Unified CCE Components and Servers, page 10-1](#)
- Cisco Unified Communications Manager (Unified CM) clusters — Use the Cisco Unified Communications Manager Capacity Tool, available only to Cisco employees and partners at:

<http://www.cisco.com/cgi-bin/CT/PGWCT/ct.cgi>

- IVR ports

Size the following components to support the number of IVR ports needed for your Unified CCE system:

- Unified CCE servers — See [Sizing Unified CCE Components and Servers, page 10-1](#)
- Cisco Unified Communications Manager (Unified CM) clusters — Use the Cisco Unified Communications Manager Capacity Tool, available only to Cisco employees and partners at:

<http://www.cisco.com/cgi-bin/CT/PGWCT/ct.cgi>

- Unified CVP ports and servers — Refer to the *Cisco Unified Customer Voice Portal Solution Reference Network Design (SRND)*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_implementation_design_guides_list.html

- Gateway ports

Size the following components to support the number of gateway ports needed for your Unified CCE system:

- Unified CCE servers — See [Sizing Unified CCE Components and Servers, page 10-1](#)
- Cisco Unified Communications Manager (Unified CM) clusters — Use the Cisco Unified Communications Manager Capacity Tool, available only to Cisco employees and partners at:

<http://www.cisco.com/cgi-bin/CT/PGWCT/ct.cgi>

- Unified CVP ports and servers — Refer to the *Cisco Unified Customer Voice Portal Solution Reference Network Design (SRND)*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_implementation_design_guides_list.html

- Gateways — Refer to the *Cisco Unified Customer Voice Portal Solution Reference Network Design (SRND)*, available at:

http://www.cisco.com/en/US/products/sw/custcosw/ps1006/products_implementation_design_guides_list.html

Erlang Calculators as Design Tools

Many traffic models are available for sizing telephony systems and resources. Choosing the right model depends on three main factors:

- Traffic source characteristics (finite or infinite)
- How lost calls are handled (cleared, held, delayed)
- Call arrival patterns (random, smooth, peaked)

For purposes of this document, there are mainly two traffic models that are commonly used in sizing call center resources, Erlang-B and Erlang-C. There are many other resources on the internet that give detailed explanations of the various models (search using traffic engineering).

Erlang calculators are designed to help answer the following questions:

- How many PSTN trunks do I need?
- How many agents do I need?
- How many IVR ports do I need?

Before you can answer these basic questions, you must have the following minimum set of information that is used as input to these calculators:

- The busy hour call attempts (BHCA)
- Average handle time (AHT) for each of the resources
- Service level (percentage of calls that are answered within x seconds)
- Grade of service, or percent blockage, desired for PSTN trunks and Unified IP IVR ports

The remaining sections of this chapter help explain the differences between the Erlang-B and Erlang-C traffic models in simple terms, and they list which model to use for sizing the specific call center resource (agents, gateway ports, and Unified IP IVR ports). There are various web sites that provide call center sizing tools free of charge (some offer feature-rich versions for purchase), but they all use the two basic traffic models, Erlang-B and Erlang-C. Cisco does not endorse any particular vendor product; it is up to the customer to choose which tool suits their needs. The input required for any of the tools, and the methodology used, are the same regardless of the tool itself.

Cisco has chosen to develop its own telephony sizing tool, called Cisco Unified CCE Resource Calculator. The version discussed here is designed to size call center resources. Basic examples are included later in this chapter to show how to use the Cisco Unified CCE Resource Calculator. Additional examples are also included to show how to use the tool when some, but not all, of the input fields are known or available.

Before discussing the Cisco Unified CCE Resource Calculator, the next two sections present a brief description of the generic Erlang models and the input/output of such tools (available on the internet) to help the reader who does not have access to the Cisco Unified CCE Resource Calculator or who chooses to use other non-Cisco Erlang tools.

Erlang-C

The Erlang-C model is used to size agents in call centers that queue calls before presenting them to agents. This model assumes:

- Call arrival is random.
- If all agents are busy, new calls will be queued and not blocked.

The input parameters required for this model are:

- The number of calls in the busy hour (BHCA) to be answered by agents
- The average talk time and wrap-up time
- The delay or service level desired, expressed as the percentage of calls answered within a specified number of seconds

The output of the Erlang-C model lists the number of agents required, the percentage of calls delayed or queued when no agents are available, and the average queue time for these calls.

Erlang-B

The Erlang-B model is used to size PSTN trunks, gateway ports, or Unified IP IVR ports. It assumes the following:

- Call arrival is random.
- If all trunks/ports are occupied, new calls are lost or blocked (receive busy tone) and not queued.

The input and output for the Erlang B model consists of the following three factors. You need to know any two of these factors, and the model will calculate the third:

- Busy Hour Traffic (BHT), or the number of hours of call traffic (in Erlangs) during the busiest hour of operation. BHT is the product of the number of calls in the busy hour (BHCA) and the average handle time (AHT).
- Grade of Service, or the percentage of calls that are blocked because not enough ports are available
- Ports (lines), or the number of Unified IP IVR or gateway ports

Cisco Unified CCE Resource Calculators

Cisco is continually enhancing the Cisco Unified Communications Resource Calculators, which currently include the following calculators:

- Standard Unified CCE Resource Calculator — Designed to provide outputs for a single call center with a single trunk group. It allows the user to vary the number of agents.
- Advanced Unified CCE Resource Calculator — Includes all the calculations of the Standard calculator and adds the ability to allocate traffic between multiple trunk groups to include calls going to a self-service IVR. It also has inputs for confidence and growth factors.
- Unified IP IVR Self-Service Calculator — A standard Erlang B calculator for determining the number of ports required on a self-service IVR. It has inputs for up to five port groups or separate IVRs.

The latest versions of these calculators and their associated user guides are available at

http://www.cisco.com/web/partners/sell/technology/ipc/integrated-solutions/customer_contact_center.html

The Cisco Unified CCE Resource Calculators are accessible to Cisco internal employees and Cisco partners. These tools are based on industry Erlang traffic models. Other Erlang traffic calculators available on the Web can also be used for sizing various contact center resources.

Figure 9-2 is a snapshot of the current Standard Unified CCE Resource Calculator, followed by a definition of each of the input and output fields, how to use them, and how to interpret them.

Figure 9-2 Cisco Unified CCE Resource Calculator

Advanced Service Technology Group
IPCC Standard Resource Calculator

CISCO SYSTEMS

Project Identification:	Project Title	
Calls per interval (BHCA):	60 min	2000 calls
Service Level Goal (SLG):	90 %	within 30 sec 0m 30s
Avg call talk time:	150 sec	2m 30s
Avg after call work time:	60 sec	1m 0s
Avg handle time (Agent calls):	210 sec	3m 30s
Avg Call treatment Time (IVR):	30 sec	0m 30s
Wait before abandon (Tolerance):	150 sec	2m 30s
Blockage % (PSTN Trunks):	1 %	of calls lost (Busy)
Check to manually enter Agents	<input type="checkbox"/>	Agents

Recommended Agents:	124
Calls completed (BHCC):	1980 calls 20 Blocked calls
Calls answered within SLG:	90% within 30 sec
Calls answered beyond SLG:	10% beyond 30 sec
Queued calls:	33.2% 657 Q Calls 4.7 Erlangs
Calls answered immediately:	66.8% 1322 calls
Avg Queue Time (AQT):	25 sec 0m 25s
Avg Speed to Answer (ASA):	8 sec 0m 8s
Avg call duration:	188 sec 3m 8s
Agents utilization:	93%
Calls exceeding Abandon Tolerance:	0% 0 Calls
PSTN Trunk Utilization:	85%
Voice trunks required:	122 Trunks T1/PRI 5.4 T1/PRI
IVR ports required for queuing:	12 IVR Ports
IVR ports required for call treatment:	26 IVR Ports
Sum of Required IVR Ports:	38

143802

Standard Unified CCE Resource Calculator Input Fields (What You Must Provide)

When using the Cisco Standard Unified CCE Resource Calculator, you must provide the following input data:

Project Identification

A description to identify the project or customer name and the specific scenario for this calculation. It helps to distinguish the different scenarios run (exported and saved) for a project or a customer proposal.

Calls Per Interval (BHCA)

The number of calls attempted during the busiest interval, or busy hour call attempts (BHCA). You can choose the interval to be 60 minutes (busy hour), 30 minutes (half-hour interval), or 15 minutes. This choice of interval length allows the flexibility to calculate staffing requirements more accurately for the busiest periods within one hour, if desired. It can also be used to calculate staffing requirements for any interval of the day (non-busy hour staffing).

Service Level Goal (SLG)

The percentage of calls to be answered within a specified number of seconds (for example, 90% within 30 seconds).

Average Call Talk Time

The average number of seconds a caller will be on-line after an agent answers the call. This value includes time talking and time placed on hold by the agent, until the call is terminated. It does not include time spent in the IVR for call treatment or time in queue.

Average After-Call Work Time

The average agent wrap-up time in seconds after the caller hangs up. This entry assumes that agents are available to answer calls when they are not in wrap-up mode. If seated agents enter into another mode (other than the wrap-up mode) where they are unavailable to answer calls, then this additional time should be included (averaged for all calls) in the after-call work time.

Average Call Treatment Time (IVR)

The average time in seconds a call spends in the IVR before an attempt is made to send the call to an agent. This time includes greetings and announcements as well as time to collect and enter digits (known as prompt and collect, or IVR menuing) to route the call to an agent. It does not include queuing time if no agents are available. (This queuing time is calculated in the output section of the calculator.) The call treatment time should not include calls arriving at the IVR for self-service with no intention to route them to agents. Self-service IVR applications should be sized separately using an Erlang-B calculator.

Wait Before Abandon (Tolerance)

This field is the amount of time in seconds that a contact center manager expects callers to wait in queue (tolerance) for an agent to become available before they abandon the queue (hang up). This value has no effect on any of the output fields except the abandon rate (number of calls abandoned).

Blockage % (PSTN Trunks)

This field is also known as Grade of Service, or the percentage of calls that will receive busy tone (no trunks available on the gateway) during the busy hour or interval. For example, 1% blockage means that 99% of all calls attempted from the PSTN during the interval will have a trunk port available on the gateway to reach the IVR or an agent.

Check to Manually Enter Agents

After checking this box, the user may manually enter the number of agents. If the number of agents entered is too far from the calculated (recommended) number, the calculator will show an error message. The error will appear any time the number of calls queued reaches 0% or 100%.

Standard Unified CCE Resource Calculator Output Fields (What You Want to Calculate)

The Standard Unified CCE Resource Calculator calculates the following output values based on your input data:

Recommended Agents

The number of seated agents (calculated using Erlang-C) required to staff the call center during the busy hour or busy interval.

Calls Completed (BHCC)

The busy hour call completions (BHCC), or the number of expected calls completed during the busy hour. It is the number of calls attempted minus the number of calls blocked.

Calls Answered Within Target SLG

The percentage of calls that are answered within the set target time entered in the Service Level Goal (SLG) field. This value is the calculated percentage of calls answered immediately if agents are available. It includes a portion of calls queued if no agents are available within the SLG (for example, less than 30 seconds). It does not include all queued calls because some calls will be queued beyond the SLG target.

Calls Answered Beyond SLG

The percentage of calls answered beyond the set target time entered in the Service Level Goal (SLG) field. For example, if the SLG is 90% of calls answered within 30 seconds, the calls answered beyond SLG would be 10%. This value includes a portion of all calls queued, but only the portion queued beyond the SLG target (for example, more than 30 seconds).

Queued Calls

The percentage of all calls queued in the IVR during the busy hour or interval. This value includes calls queued and then answered within the Service Level Goal as well as calls queued beyond the SLG. For example, if the SLG is 90% of calls answered within 30 seconds and queued calls are 25%, then there are 10% of calls queued beyond 30 seconds, and the remaining 15% of calls are queued and answered within 30 seconds (the SLG).

Calls Answered Immediately

The percentage of calls answered immediately by an agent after they receive treatment (if implemented) in the IVR. These calls do not have to wait in queue for an agent. As in the preceding example, if 25% of the calls are queued (including those beyond the target of 30 seconds), then 75% of the calls would be answered immediately.

Average Queue Time (AQT)

The average amount of time in seconds that calls will spend in queue waiting for an agent to become available during the interval. This value does not include any call treatment in the IVR prior to attempting to send the call to an agent.

Average Speed of Answer (ASA)

The average speed of answer for all calls during the interval, including queued calls and calls answered immediately.

Average Call Duration

The total time in seconds that a call remained in the system. This value is the sum of the average talk time, the average IVR delay (call treatment), and the average speed of answer.

Agents Utilization

The percentage of agent time engaged in handling call traffic versus idle time. After-call work time is not included in this calculation.

Calls Exceeding Abandon Tolerance

The percentage (and number) of calls that will abandon their attempt during the interval, based on expected Tolerance time specified in the input. If this output is zero, it means that all queued calls were answered by an agent in less than the specified abandon time (longest queued call time is less than the abandon time).

PSTN Trunk Utilization

The occupancy rate of the PSTN trunks, calculated by dividing the offered load (Erlangs) by the number of trunks.

Voice Trunks Required

The number of PSTN gateway trunks required during the busy interval, based on the number of calls answered by the voice gateway and the average hold time of a trunk during the busy interval. This value includes average time of call treatment in the IVR, queuing in the IVR (if no agents are available), and agent talk time. This calculated number assumes all trunks are grouped in one large group to handle the specified busy hour (or interval) calls. If several smaller trunk groups are used instead, then additional trunks would be required, therefore smaller groups are less efficient.

IVR Ports Required for Queuing

The number of IVR ports required to hold calls in queue while the caller waits for an agent to become available. This value is based on an Erlang-B calculation using the number of queued calls and the average queue time for those calls.

IVR Ports Required for Call Treatment

The number of IVR ports required for calls being treated in the IVR. This value is based on an Erlang-B calculation using the number of calls answered and the average call treatment time (average IVR delay).

Total IVR Ports Requirement

The total number of IVR ports required if the system is configured with separate port groups for queuing and treatment. Pooling the ports for treatment and queuing results in fewer ports for the same amount of traffic than if the traffic is split between two separate IVR port pools or groups. However, Cisco recommends that you configure the number of ports required for queuing in a separate group, with the ability to overflow to other groups if available.

Submit

After entering data in all required input fields, click on the Submit button to compute the output values.

Export

Click on the Export button to save the calculator input and output in the format of comma-separated values (CSV) to a location of your choice on your hard drive. This CSV file could be imported into a Microsoft Excel spreadsheet and formatted for insertion into bid proposals or for presentation to clients or customers. Multiple scenarios could be saved by changing one or more of the input fields and combining all outputs in one Excel spreadsheet by adding appropriate titles to the columns to reflect any changes in the input. This format makes comparing results of multiple scenarios easy to analyze.

Sizing Call Center Agents, IVR Ports, and Gateways or Trunks (Inbound Call Center)

The call center examples in this section illustrate how to use the Unified CCE Resource Calculator in various scenarios, along with the impact on required resources for inbound call centers. The first example in this section is a basic call flow, where all incoming calls to the call center are presented to the voice gateway from the PSTN. Calls are routed directly to an agent, if available; otherwise, calls are queued until an agent becomes available.

Basic Call Center Example

This example forms the basis for all subsequent examples in this chapter. After a brief explanation of the output results highlighting the three resources (agents, IVT ports, and PSTN trunks) in this basic example, subsequent examples build upon it by adding different scenarios, such as call treatment and agent wrap-up time, to demonstrate how the various resources are impacted by different call scenarios.

This basic example uses the following input data:

- Total BHCA (60-minute interval) into the call center from the PSTN to the voice gateway = 2,000.
- Desired service level goal (SLG) of 90% of calls answered within 30 seconds.
- Average call talk time (agent talk time) = 150 seconds (2 minutes and 30 seconds).
- No after-call work time (agent wrap-up time = 0 seconds).
- No call treatment (prompt and collect) is implemented initially. All calls will be routed to available agents or will be queued until an agent becomes available.
- Wait time before caller hangs up (tolerance) = 150 seconds (2 minutes and 30 seconds).
- Desired grade of service (percent blockage) for the PSTN trunks on the voice gateway = 1%.

After entering the above data in the input fields, pressing the Submit button at the bottom of the calculator results in the output shown in [Figure 9-3](#).

Figure 9-3 Basic Example

Project Identification:	Basic Example	
Calls per interval (BHCA):	60 min	2000 calls
Service Level Goal (SLG):	90 % within	30 sec 0m 30s
Avg call talk time:	150 sec	2m 30s
Avg after call work time:	0 sec	0m 0s
Avg handle time (Agent calls):	150 sec	2m 30s
Avg Call treatment Time (IVR):	0 sec	0m 0s
Wait before abandon (Tolerance):	150 sec	2m 30s
Blockage % (PSTN Trunks):	1 %	of calls lost (Busy)
Check to manually enter Agents	<input type="checkbox"/>	Agents
Recommended Agents: 90		
Calls completed (BHCC):	1980 calls	20 Blocked calls
Calls answered within SLG:	93% within	30 sec
Calls answered beyond SLG:	7% beyond	30 sec
Queued calls:	31.7% 627 Q Calls	3.6 Erlangs
Calls answered immediately:	68.3% 1352 calls	
Avg Queue Time (AQT):	20 sec	0m 20s
Avg Speed to Answer (ASA):	6 sec	0m 6s
Avg call duration:	156 sec	2m 36s
Agents utilization:	92%	
Calls exceeding Abandon Tolerance:	0%	0 Calls
PSTN Trunk Utilization:	83%	
Voice trunks required :	103 Trunks	T1/PRI 4.6 T1/PRI
IVR ports required for queuing:	10 IVR Ports	
IVR ports required for call treatment:	0 IVR Ports	
Sum of Required IVR Ports:	10	
<input type="button" value="Submit"/> <input type="button" value="Export"/>		

126900

Notice that the output shows 1980 calls received and processed (completed) by the voice gateway, out of the total of 2000 calls attempted from the PSTN. This is because we have requested a provisioning of 1% blockage from our PSTN provider, which results in 20 calls (1%) being blocked by the PSTN (and receiving busy tone) out of the total 2000 calls.

Agents

The result of 90 seated agents is determined by using the Erlang-C function imbedded in the Unified CCE Resource Calculator, and calls will be queued to this resource (agents).

Notice that, with 90 agents, the calculated service level is 93% of calls answered within 30 seconds, which exceeds the desired 90% requested in the input section. Had there been one less agent (89 instead of 90), then the 90% SLG would not have been met.

This result also means that 7% of the calls will be answered beyond the 30 second SLG. In addition, there will be 31.7% of calls queued; some will queue less than 30 seconds and others longer. The average queue time for queued calls is 20 seconds.

If 31.7% of the calls will queue, then 68.3% of the calls will be answered immediately without delay in a queue, as shown in the output in [Figure 9-3](#).

IVR Ports Required for Queuing

In this basic example, the Unified IP IVR is being used as a queue manager to queue calls when no agents are available. The calculator shows the percent and number of calls queued (31.7%, or 627 calls) and the average queue time (20 seconds).

These two outputs from the Erlang-C calculation are then used as inputs for the imbedded Erlang-B function in the calculator to compute the number of IVR ports required for queuing (10 ports in this example).

PSTN Trunks (Voice Gateway Ports)

Similarly the calculator uses Erlang-B to calculate the required number of voice gateway ports (PSTN trunks) based on the call load (answered calls) and the calls that have to queue when no agents are available.

Total trunks required to carry this total traffic load above is 103 trunks.

This calculation does not include trunks that might be needed for call scenarios that require all calls to be treated first in the IVR before they are presented to available agents. That scenario is discussed in the next example.

Call Treatment Example

This example builds upon the basic example in the preceding section. Again, all incoming calls to the call center are presented to the voice gateway from the PSTN, then calls are immediately routed to the Unified IP IVR for call treatment (such as an initial greeting or to gather account information via prompt-and-collect) before they are presented to an agent, if available. If no agents are available, calls are queued until an agent becomes available.

The impact of presenting all calls to the Unified IP IVR is that the PSTN trunks are held longer, for the period of the call treatment holding time. More Unified IP IVR ports are also required to carry this extra load, in addition to the ports required for queued calls.

Call treatment (prompt and collect) in this example appears not to impact the number of required agents because the traffic load presented to the agents (number of calls, talk time, and service level) is assumed not to have changed. In reality, adding call treatment such as collecting information input from the callers to identify them to agents using a CTI-Pop screen will reduce the average time a caller spends with an agent, thus saving valuable resources, providing more accurate selection and routing of appropriate agent, and improving customer service.

Using a 15-second call treatment and keeping all other inputs the same, [Figure 9-4](#) shows the number of PSTN trunks (112) and Unified IP IVR ports (16) required in addition to the existing 10 ports for queuing.

Figure 9-4 Call Treatment in IVR

Project Identification: Call Treatment Example	
Calls per interval (BHCA):	60 min 2000 calls
Service Level Goal (SLG):	90 % within 30 sec 0m 30s
Avg call talk time:	150 sec 2m 30s
Avg after call work time:	0 sec 0m 0s
Avg handle time (Agent calls):	150 sec 2m 30s
Avg Call treatment Time (IVR):	15 sec 0m 15s
Wait before abandon (Tolerance):	150 sec 2m 30s
Blockage % (PSTN Trunks):	1 % of calls lost (Busy)
Check to manually enter Agents	<input type="checkbox"/> Agents
Recommended Agents: 90	
Calls completed (BHCC):	1980 calls 20 Blocked calls
Calls answered within SLG:	93% within 30 sec
Calls answered beyond SLG:	7% beyond 30 sec
Queued calls:	31.7% 627 Q Calls 3.6 Erlangs
Calls answered immediately:	68.3% 1352 calls
Avg Queue Time (AQT):	20 sec 0m 20s
Avg Speed to Answer (ASA):	6 sec 0m 6s
Avg call duration:	171 sec 2m 51s
Agents utilization:	92%
Calls exceeding Abandon Tolerance:	0% 0 Calls
PSTN Trunk Utilization:	84%
Voice trunks required :	112 Trunks T1/PRI 5.0 T1/PRI
IVR ports required for queuing:	10 IVR Ports
IVR ports required for call treatment:	16 IVR Ports
Sum of Required IVR Ports:	26
<input type="button" value="Submit"/> <input type="button" value="Export"/>	

126901

After-Call Work Time (Wrap-up Time) Example

Using the previous example, we now add an average of 45 seconds of work time (wrap-up time) after each call. We can then use the Unified CCE Resource Calculator to determine the number of agents required to handle the same traffic load (see Figure 9-5).

After-call work time (wrap-up time) begins after the caller hangs up, so trunk and Unified IP IVR resources are not impacted and should remain the same, assuming all other input remains the same. Assuming the SLG and traffic load also remain the same, additional agents would be required only to service the call load and to compensate for the time agents are in the wrap-up mode.

Figure 9-5 After-Call Work Time

Project Identification: After Call Work Time	
Calls per interval (BHCA):	60 min 2000 calls
Service Level Goal (SLG):	90 % within 30 sec 0m 30s
Avg call talk time:	150 sec 2m 30s
Avg after call work time:	45 sec 0m 45s
Avg handle time (Agent calls):	195 sec 3m 15s
Avg Call treatment Time (IVR):	15 sec 0m 15s
Wait before abandon (Tolerance):	150 sec 2m 30s
Blockage % (PSTN Trunks):	1 % of calls lost (Busy)
Check to manually enter Agents	<input type="checkbox"/> Agents
Recommended Agents: 116	
Calls completed (BHCC):	1980 calls 20 Blocked calls
Calls answered within SLG:	92% within 30 sec
Calls answered beyond SLG:	8% beyond 30 sec
Queued calls:	30.5% 603 Q Calls 3.8 Erlangs
Calls answered immediately:	69.5% 1376 calls
Avg Queue Time (AQT):	22 sec 0m 22s
Avg Speed to Answer (ASA):	7 sec 0m 7s
Avg call duration:	172 sec 2m 52s
Agents utilization:	92%
Calls exceeding Abandon Tolerance:	0% 0 Calls
PSTN Trunk Utilization:	84%
Voice trunks required:	113 Trunks T1/PRI 5.0 T1/PRI
IVR ports required for queuing:	10 IVR Ports
IVR ports required for call treatment:	16 IVR Ports
Sum of Required IVR Ports:	26
<input type="button" value="Submit"/> <input type="button" value="Export"/>	

126902

Note that trunks and IVT ports remained virtually the same, except that there is one additional trunk (113 instead of 112). This slight increase is not due to the wrap-up time, but rather is a side effect of the slight change in the SLG (92% instead of 93%) due to rounding calculations for the required 116 agents due to wrap-up time.

Agent Staffing Considerations

In calculating agent requirements, make the following adjustments to factor in all the activities and situations that make agents unproductive or unavailable:

Agent Shrinkage

Agent shrinkage is a result of any time for which agents are being paid but are not available to handle calls, including activities such as breaks, meetings, training, off-phone work, unplanned absence, non-adherence to schedules, and general unproductive time.

Agent Shrinkage Percentage

This factor will vary and should be calculated for each call center. In most call centers, it ranges from 20% to 35%.

Agents Required

This number is based on Erlang-C results for a specific call load (BHCA) and service level.

Agents Staffed

To calculate this factor, divide the number of agents required from Erlang-C by the productive agent percentage (or 1 minus the shrinkage percentage). For example, if 100 agents are required from Erlang-C and the shrinkage is 25%, then $100/.75$ yields a staffing requirement of 134 agents.

Call Center Design Considerations

Consider the following design factors when sizing call center resources:

- Compute resources required for the various busy intervals (busy hours), such as seasonal busy hours and average daily busy hour. Many businesses compute the average of the 10 busiest hours of the year (excluding seasonal busy hours) as the busy-hour staffing. Retail business call centers will add temporary staff based on seasonal demands such as holiday seasons. Run multiple interval calculations to understand daily staff requirements. Every business has a different call load throughout the day or the week, and agents must be staffed accordingly (using different shifts or staffing levels). Customer Relationship Management (CRM) and historical reporting data help to fine-tune your provisioning computations to maintain or improve service levels.
- When sizing IVR ports and PSTN trunks, it is better to over-provision than to under-provision. The cost of trimming excess capacity (disconnecting PSTN lines) is much cheaper than lost revenue, bad service, or legal risks. Some governmental agencies are required to meet minimum service levels, and outsourced call centers might have to meet specific service level agreements.
- If the call center receives different incoming call loads on multiple trunk groups, additional trunks would be required to carry the same load using one large trunk group. You can use the Erlang-B calculator to size the number of trunks required, following the same methodology as in the [Call Treatment Example, page 9-14](#). Sizing of required trunks must be done for each type of trunk group.
- Consider marketing campaigns that have commercials asking people to call now, which can cause call loads to peak during a short period of time. The Erlang traffic models are not designed for such short peaks (bunched-up calls); however, a good approximation would be to use a shorter busy interval, such as 15 minutes instead of 60 minutes, and to input the expected call load during the busiest 15 minutes to compute required agents and resources. Using our [Basic Call Center Example, page 9-12](#), a load of 2000 calls in 60 minutes (busy interval) requires 90 agents and 103 trunks. We would get exactly the same results if we used an interval of 15 minutes with 500 calls (¼ of the call load). However, if 600 of the calls arrive during a 15-minute interval and the balance of the calls (1400) arrive during the rest of the hour, then 106 agents and 123 trunks would be required instead to answer all 600 calls within the same service level goal. In a sales call center, the potential to capture additional sales and revenue could justify the cost of the additional agents, especially if the marketing campaign commercials are staggered throughout the hour, the day, and the various time zones.
- Consider agent absenteeism, which can cause service levels to go down, thus requiring additional trunks and Unified IP IVR queuing ports because more calls will be waiting in queue longer and fewer calls will be answered immediately.
- Adjust agent staffing based on the agent shrinkage factor (adherence to schedules and staffing factors, as explained in [Agent Staffing Considerations, page 9-16](#)).
- Allow for growth, unforeseen events, and load fluctuations. Increase trunk and IVR capacity to accommodate the impact of these events (real life) compared to Erlang model assumptions. (Assumptions might not match reality.) If the required input is not available, make assumptions for the missing input, run three scenarios (low, medium, and high), and choose the best output result based on risk tolerance and impact to the business (sales, support, internal help desk, industry, business environment, and so forth). Some trade industries publish call center metrics and statistics,

such as those shown in [Table 9-2](#), available from web sites such as <http://www.benchmarkportal.com>. You can use those industry statistics in the absence of any specific data about your call center (no existing CDR records, historical reports, and so forth).

Table 9-2 eBusiness Best Practices for All Industries, 2001¹

Inbound Call Center Statistics	Average	Best Practices
80% calls answered in? (seconds)	36.7	18.3
Average speed of answer (seconds)	34.6	21.2
Average talk time (minutes)	6.1	3.3
Average after-call work time (minutes)	6.6	2.8
Average calls abandoned	5.5%	3.7%
Average time in queue (seconds)	45.3	28.1
Average number of calls closed on first contact	70.5%	86.8%
Average TSR occupancy	75.1%	84.3%
Average time before abandoning (seconds)	66.2	31.2
Average adherence to schedule	86.3%	87.9%
Cost per call	\$9.90	\$7.12
Inbound calls per 8-hour shift	69.0	73.9
Percentage attendance	86.8%	94.7%

1. Special Executive Summary; Principal Investigator, Dr. Jon Anton; Purdue University, Center for Customer-Driven Quality.

Use the output of the Unified CCE Resource Calculator as input for other Cisco configuration and ordering tools that may require as input, among other factors, the number of IVR ports, number of agents, number of trunks, and the associated traffic load (BHCA).