CALL FLOWS

CUCM call flows are more complex than CME because CUCM is a distributed system that uses remote components, such as gateways to route calls.

Eliminating DNS—DNS isn't recommended because of the delay and vulnerability it introduces. Default CUCM install has host name in the fields for server(s) that phones should register with. Change this to an IP address:

(Admin) System → Server

Also change the names of all the Phone URL Parameters (and secured versions) in

(Admin) System → Enterprise Parameters

Signaling (SCCP/SIP) is only between the phone and CUCM; voice bearer streams using RTP (Real-time Transport Protocol) are only between phones unless CUCM is hosting a voice conference (where all voice streams go to CUCM and a combo of all except your own comes to you).

Centralized Remote Branch Call Flow—signaling from the remote branch goes to the central CUCM server; RTP traffic (voice bearer stream) only goes over the WAN if it needs to.

SRST (Survivable Remote Site Telephony)—When the WAN is lost and the central CUCM server becomes unavailable to the remote site, the remote site router transparently takes over on-site call management using CME and begins sending off-site traffic over the PSTN, with appropriate modifications to the dialed numbers.

At the central site CUCM, the remote phones are deregistered (thus unreachable), so an alternate path over the PSTN will need to be specified centrally:

- The call-routing table has a second option added to provide a PSTN gateway and appropriate digit manipulation to provide the required PSTN dialed digits.
- The Call Forward UnRegistered (CFUR) option is configured on each branch phone to specify the full PSTN number needed to reach the branch phone.

When the WAN is restored, the phones deregister from the SRST router, re-register with CUCM, and everything is back to normal.

Centralized CUCM Deployment

- Supports 2000 locations & 2100 H.323 or MGCP (Media Gateway Control Protocol) devices per cluster (as of v10)
- The only limitation on the number of phones per site is what the SRST router at that location can handle (model & RAM).
- The WAN needs QoS to prioritize and limit delay for both RTP and signaling traffic
- Use CAC (below)

CAC (Call Admission Control) can limit the number of simultaneous calls over the WAN to the QoS-allocated bandwidth to ensure that those calls that are on the WAN are well handled. Can either be CUCM location-based or based on RSVP (Resource Reservation Protocol). With location-based, CUCM tracks all calls' bandwidth going to a location and refuses to connect (reorder tone) new calls when not enough bandwidth remains.

AAR (Automated Alternate Routing) can only be triggered by CAC and avoids dropped calls (default CAC behavior) by sending them over the PSTN instead.

Note the distinction between WAN failure behavior and WAN full behavior.

- Failure—hierarchical dial plan design and CFUR (on the CUCM side) send the call to PSTN
- Full—CAC triggers AAR, which redials the call with a full PSTN number, if configured to do so

Distributed Deployment—Phones still connect directly to each other (across WAN) for voice data; the signaling flow is:

Phone 1 (site A) → CUCM (site A) → CUCM (site B) → Phone 2 (site B)

Signaling Protocol between clusters can be SIP, ICT (Inter-Cluster Trunk), or H.323. Within the cluster, SCCP and SIP are still used. Scalable to thousands of sites when gatekeepers are used. Each of the sites can choose CME or CUCM, based on their size. For PSTN backup during a WAN failure, use a hierarchical dial-plan. For PSTN backup during a WAN overload, use gatekeeper CAC.

Gatekeeper CAC—A router IOS feature that can be configured on one or more gateway routers in the system and provides a centralized service to track available bandwidth between clusters and trigger AAR as necessary. Location-based CAC doesn't work in a distributed deployment because clusters don't tell each other about their WAN utilization.

CALL-ROUTING REQUESTS & DESTINATIONS

Call-Routing Request—When one of the following sources signals CUCM with a string of dialed digits. Possible sources:

- IP Phone—manually dialed number, speed dial, feature button, or softkey. Request is placed using one of the phone's configured lines
- Trunk—Signal inbound from another CUCM cluster, CUCME, or other call agent
- Gateway—Inbound call from PSTN or PBX
- Translation Pattern—original digits are matched to the pattern and immediately transformed to a new dial string, which is resubmitted to digit analysis for routing
- Voicemail Port—if the application attempts a call, transfer, or message notification on behalf of a user’s mailbox

Call-Routing Destinations (that are possible in CUCM)—AKA digit patterns that dialed digits can match against

- DN (Directory Number)—Each phone button can be assigned an on-net extension
- Translation Pattern—When matched, the number is transformed and resubmitted to digit analysis
- Route Pattern—when matched, triggers a call-routing process with a hierarchical set of potential paths
- Hunt Pilot—triggers a customizable call-coverage system
- Call Park Number—a (range of) pattern(s) from which a call can be picked up by any phone
- Meet-Me Number—To start a conference call, a user dials into the "Meet-Me" number, as do all subsequent participants.

SIP URI—an alphanumeric string representing a call-routing destination. Handled by CUCM v9+. Out of scope, but they look like "sip:1-999-123-4567@voip-provider.example.net".

CALL-ROUTING CONFIGURATION ELEMENTS

Route Patterns—matches a string of dialed digits, perhaps with the help of wildcards. Each is associated with either a route list or a specific gateway. Note that if a gateway has a route pattern directly associated to it, the gateway is "locked in" to that route pattern and can no longer be referenced by a route group.

Route List—an ordered list of Route Groups, the first is the preferred call-routing path. Each call starts at the top of the list, no round robin. The next option is only used when the first is rejected (no more circuits are available, etc.). Cost often drives the preference order.

Route Group—used to group devices with similar signaling characteristics, like a set of WAN IP trunks or PSTN PRI gateways. Within a group, can choose top-down or "circular" (round-robin).

Local Route Groups can simplify dial-plan designs for multi-location systems. A route group is defined in the device pool, and the route list references it instead of site-specific devices. This can decouple the location of a PSTN gateway from the route lists that use it.

Gateways & Trunks—Gateways (routers) are devices that physically terminate circuits to the PSTN, the PBX, or IP WAN circuits to remote clusters. Gateway devices can be controlled with peer-to-peer gateway protocols (H.323 & SIP) or gateway control protocols (MGCP & SCCP).

Configuration Order—Devices, Route Groups, Route Lists, Route Patterns

Call Processing Order—reverse of above. Dialed digits match a route pattern, which points to a Route List, which points to a Route Group, which references Devices.

CALL-ROUTING BEHAVIOR

Digit Analysis—how CUCM matches dialed digits to possible targets for call routing. On trunks, gateways, and basic SIP phones, the entire string is processed at once; with SCCP phones and SIP phones with KPML (KeyPad Markup Language), the digit analysis happens as each digit is dialed.

CUCM will try for the most specific (non-wildcard) match possible, even if that means waiting for extra digits. In other words, "Among all the potential matches, it matches the fewest strings other than the actual dialed string."

T.302 Inter-Digit Timeout Value—Default 15 seconds. How long to wait for the user to dial further digits to achieve a longer, more specific match before assuming the user is done dialing and matching what you've got against some shorter wildcard pattern or treating the call as a failure if nothing matches.

HUNT GROUPS

Hunt Group—a set of DNs that can be reached by dialing a common number, e.g. helpdesk. 3 Components:

- Line Groups—the DNs that will be rung, with a choice of who to ring: top-down, circular, longest idle, or broadcast. Also tells when (or if) to proceed to the next line group in a hunt list
- Hunt List—Top down ordered list of Line groups. If the first group can't cover the call, go to the next
- Hunt Pilot—A call-routing entry (like a route pattern) that matches a dialed string and targets a hunt list, which in turn targets a line group.

CLASS OF CONTROL

Class of Control—The ability to apply calling restrictions to devices. Configured using Partitions & Calling Search Spaces. Examples:

- Preventing certain individuals from placing long-distance calls
- Routing the same called number to different targets at different times of day
- Routing the same called number to different targets at different locations

Partition—A grouping of dialable things with similar reachability. For example, DNs, Route Patterns, Translation Patterns, Voicemail Ports, Meet-Me Conference Numbers. Up to 75 can be created, in addition to the default partition "<none>."

CSS (Calling Search Space)—Top-down ordered list of partitions. Assigned to things that can place calls to limit what dialable things they can reach. Can be applied to a device (IP phone or gateway) or to a line on the IP phone. The default CSS contains only the <none> partition by default.

Interaction of Partitions and Calling Search Spaces—If the target being called doesn't exist in one of the partitions in the css of the caller, the call will fail.

Example (Lobby Phone)

- Create a partition and add all route patterns that don't cost money (don't forget 911)
- Create a css and assign that partition to it
- Assign that css to the lobby phone

Notes:

- When a route pattern is moved from the default partition <none>, it can no longer be accessed from the default css. Calls to matching numbers can't be made until a working css/Partition structure is complete. Best practice = set this up before the system is in use
- Every css includes the default partition <none> at the end of its list of custom partitions. Therefore, any target remaining in the default partition <none> is reachable by every calling device.

Line CSS—You can apply a css to a line on a phone, not just the whole phone.

- If both a device and line css are applied, they are concatenated—line css partitions, followed by device css partitions. In other words a line css will override a device css. (This doesn't matter in our simple example of allowing only local calls; perhaps the more complex example alluded to ("routing the same number to different targets at different times of day") makes this matter
- Best Practice for fewer total css, making the dial-plan more manageable and scalable:
    - The device css should allow full privileges to all patterns, with call routing based on device location (e.g. use the local PSTN gateway for PSTN calls
    - Apply call restrictions using the line css, which can contain route patterns matching and blocking long-distance numbers