Predicting Multipoint Conference Unit (MCU) Port Requirements of Mobile Video
Using the Erlang B Model

One of the primary elements of guesswork for a video network designer today is the sizing prediction for a multipoint conference bridge. As video becomes ubiquitous through the desktop, mobile devices, and conference room, multipoint video conferences will become as commonplace as audio conferencing is today. In the past, multipoint bridges required scheduled setup with facilitated conference management. This allowed resource usage to be "managed" through time shifting and allocation. In that simplistic environment, maximum capacity could be easily defined and bridge sizing could be determined through user complaints when schedule time was not available.

Today, multipoint videoconferencing expects to see an exponential growth as it becomes an everyday tool for everyone within an enterprise expanding beyond the executive team. Bridge design is moving to the ad-hoc multipoint conference where maximum limits are no longer predictable and usage will follow a more random pattern. In an ad-hoc usage pattern, MCU size requirements will more closely resemble audio sizing predictors for trunk and port estimation. Classic "maximum port" planning will tend to result in gross oversizing of MCU and unnecessary expenses. In order to provide an alternative planning tool for the video network designer, we propose that the classic telephony Erlang B predictive model used for years in the telephony world be brought forward for MCU port usage estimation.

In the 1920's, Danish mathematician A.K. Erlang developed models for predicting telephone traffic to allow for more accurate trunk and call service sizing. These models allow the planner to reduce resources from an absolute maximum value, to more reasonable cost implementation while maintaining an accurate predictor for "blocked" or unavailable resource probabilities. The Erlang B and Erlang C predictive models are used today in everything from network equipment and trunk sizing to determining the number of operators necessary in a call center. We have selected the Erlang B model for MCU port prediction, making the assumption that blocked ad-hoc video callers will simply "hang-up" and try again later.

Usage Scenarios

To begin we need to break down usage scenarios based on the most common applications of Multipoint videoconferencing. The scenarios we look at are:

1. **Classroom / Education**

   This scenario typically represents the distance learning environment. There is typically a single classroom designated as the teaching location and multiple locations designated as student locations. Prior to desktop / mobile student capability, this environment was extremely predictable due to a low limited number (1-5) of remote student locations. Now that students may attend the classroom from home or on the road, ports must be available for each remote student as opposed to groups of remote students resulting in a typical port count of 20-30 remote connections. Classes normally last from 1-2 hours.

2. **Coordinated Business Meetings**

   This scenario is typically associated with set scheduled business meetings, i.e. weekly sales team forecast meetings, monthly board meetings, or recurring administrative conferences. Typically these conferences would be scheduled between a few locations (1-5) where participants would gather into conference rooms to work together. Today, participants may also attend from remote locations on the road or from small home offices. Resource requirements in this environment will typically rise to around 10 locations. This type of meeting normally lasts 2-4 hours.

3. **Enterprise Wide Coordinated Conferences**

   This scenario is typically associated with weekly and/or monthly enterprise wide announcements, coordination meetings, or training session. Historically, employees gather in geographically disparate conference rooms and there exists a single presentation room from which a C level company officer makes a presentation, usually followed by a Q&A at the end. Streaming and recording has taken over a lot of this application, however, many organizations continue to use Multipoint video for this function. These conferences are typically ½ day events.

4. **Ad/Hoc Business Meetings**
   This scenario is used for face to face gatherings of company and/or subordinate organization coordination. Historically, these meetings had to be scheduled and facilitated with all participants gathering in a small number of disperse conferences rooms resulting in 1-3 video connections. With mobility, it will become more common for all participants in these meetings to be at individual locations (home, small office, one the road) and will require around 10-15 connections. These conferences typically last no more than 1 hour.

5. **Command and Control**

   This scenario is used for action coordination in an Emergency Response, or coordinated action environment. In the past the commander / coordinator was typically connected with 3-5 remote sub-coordination centers located in the field. Reports were "walked" to the sub-centers and relayed to the command center. In the same way, orders were given to the sub-centers and "walked" to the personnel. Typically these conferences remain operational during the entire event (8 hours or greater). With mobile video this scenario will change to allow the field personnel to drop in and out of the command and control conference directly without the sub-center involvement. One or more conferences will remain operational with remote connections randomly dropping in and out of the conference.

6. **Small Group Coordination**

   This scenario is used for face to face gatherings of company and/or subordinate organization coordination. Historically, these meetings were held over the telephone with the addition of WebEx or a similar collaboration product because the scheduling and facilitation of a video conference was overly time consuming and might take longer to carry out than simply solving the issue with the telephone. With mobility, it will become more common for all participants in these meetings to use video from individual locations (home, small office, one the road) and will require around 10-15 connections. These conferences typically last no more than 1 hour.

Impact of Mobile Video

Mobility will change the paradigm of multipoint video conferencing from defined scenarios above to an off-the-shelf capability that every organization expects within reach. Therefore it is probably best to look at the future not as much in terms of defined scenarios, but a combination of scenarios. Therefore resource sizing must take into account not only resources for the primary scenario, but additional resources for secondary and even tertiary activities.

In each of the above scenarios, it is possible to use maximum usage predictors at low volume to provide a minimum level of multipoint video conferencing to a small group of participants. However, as usage expands beyond the boundaries of conference room and enterprise and becomes more ad-hoc, it is not reasonable to purchase enough resources to ensure that all connection requirements are met 100% of the time. A trade-off must therefore be made between over-sizing and reasonable expectation. This is where the Erlang B Model comes into play.

The Erlang B Model

The Erlang B Model assumes a random Poisson distribution of resource requests for a service during a defined “busy” time period. It then uses a request arrival rate estimation, a service time period estimate, and the number of available service delivery elements to predict the probability that a service request gets “blocked” or cannot be serviced. Typically in the telephony world a probability of 0.05 (or 1 of 20) is considered “acceptable” while 0.10 is terrible, and 0.001 is so good that it is considered non-blocking. Because multipoint video conference is still in its infancy with most “customers” we believe that acceptability should be defined as no higher Erlang probability than 0.01 (1 in 100 attempts blocked). This value may be allowed to go higher as ad-hoc becomes the norm.

Erlang B Model inputs consist of:

a) Arrival Rate - Speed at which service requests arrive.
b) Service Period – Time required by the average service request for service.
c) Available Servers – Number of service resources for requests.

Note: Arrival Rate * Service Period becomes Traffic Intensity in “Erlangs”.

The output of the model is:

a) Blockage probability – The probability that a service request will be denied or go unserviced.

Using iterative techniques, the model may be reversed so that the desired Blockage Probability becomes an input and the Available Servers becomes an output as "Required Servers".

Translating the model inputs to a multiport videoconferencing environment looks like this:

a) Arrival Rate = Number of Meetings X Meeting Participants / Busy Time Period (in Minutes)
b) Service Period = Conference Length (In Minutes)
c) Available Servers = Multipoint Conference Ports Available

Output of the model will obviously vary with the appropriate choice of inputs. In differing circumstances it may be appropriate to use maximum expected values for inputs, while in larger capacity environments, average values may be a better suited choice. Below, we have outlined each of the individual scenarios described above and, based on experience, given some example model results.

Scenario Model Examples

1. **Classroom / Education**

<u>Minimum predictor</u>
This is the classic predictor and should be used to determine the minimum number of ports for sizing and recognizing limits of the system.

Minimum Port Count = Maximum Connections in Class x max simultaneous classes

<u>Erlang predictor</u>
This predictor should be used when class sizes and times of day are unknown but the number of classes held during a day is greater than the number of hours in a day.

<u>Inputs</u>
Busy Time = 8 hours
Meetings Per Busy Time = 10
Average Attendees Per Meeting = 20
Average Time Per Meeting = 90 (Minutes)
Desired Blockage = 0.01

<u>Outputs</u>
Arrival Rate =0.42 per minute
Traffic Intensity = 37.5
Ports Required = 50
Calculated Blockage = 0.0087

Notes: This will support up to 2 simultaneous classes out of the 10 used as input. It would be wise to up the output to 60 ports to support 3 classes.

2. **Coordinated Business Meeting**

Minimum predictor

This is the classic predictor and should be used to determine the minimum number of ports for sizing and recognizing limits of the system.

Minimum Port Count = Maximum Connections in Meetings x max simultaneous meetings

Erlang predictor

This predictor should be used when meeting sizes and times of day are unknown but the number of meetings held during a day is anticipated to be high.

Inputs

Busy Time = 6 hours (Assume 10:00 a.m. to 4:00 p.m.)
Meetings Per Busy Time = 100 (100 different meetings during a day)
Average Attendees Per Meeting = 10
Average Time Per Meeting = 60 (Minutes)
Desired Blockage = 0.01

Outputs

Arrival Rate =2.78 per minute
Traffic Intensity = 166.67
Ports Required = 187
Calculated Blockage = 0.0094

Notes: This will support up to 18 simultaneous meetings out of the 100 used as input.

3. **Enterprise Wide Coordinated Conferences**

Minimum predictor

This is the classic predictor and should be used to determine the minimum number of ports for sizing and recognizing limits of the system.

Minimum Port Count = Maximum Connections in Meetings x max simultaneous meetings

Erlang predictor

This predictor should be used when times of day are unknown but the number of conferences held during a day is limited.

Inputs

Busy Time = 6 hours (Assume 10:00 a.m. to 4:00 p.m.)
Meetings Per Busy Time = 2 (2 different meetings during a day)
Average Attendees Per Meeting = 30
Average Time Per Meeting = 240 (Minutes) or ½ day
Desired Blockage = 0.01

<u>Outputs</u>
Arrival Rate =0.167 per minute
Traffic Intensity = 40
Ports Required = 53
Calculated Blockage = 0.0082

Notes: This will not support the overlap of two full meetings. It is probably better to use the Minimum port equation in this case.

4. **Ad/Hoc Business Meeting**

<u>Minimum predictor</u>
This is the classic predictor and should be used to determine the minimum number of ports for sizing and recognizing limits of the system.

Minimum Port Count = Maximum Connections in Meetings x max simultaneous meetings

<u>Erlang predictor</u>
This predictor should be used when meeting sizes and times of day are unknown but the number of meetings held during a day is anticipated to be high.

<u>Inputs</u>
Busy Time = 6 hours (Assume 10:00 a.m. to 4:00 p.m.)
Meetings Per Busy Time = 150 (150 different meetings during a day)
Average Attendees Per Meeting = 10
Average Time Per Meeting = 60 (Minutes)
Desired Blockage = 0.01

<u>Outputs</u>
Arrival Rate =4.17 per minute
Traffic Intensity = 250
Ports Required = 273
Calculated Blockage = 0.0090

Notes: This will support 27 simultaneous meetings.

5. **Command and Control**

Minimum predictor
This is the classic predictor and should be used to determine the minimum number of ports for sizing and recognizing limits of the system.
Minimum Port Count = Maximum Connections in Meetings x max simultaneous meetings

Erlang predictor
This predictor should be used when meeting sizes are unknown but the number of meetings held during a day is known.

Inputs
Busy Time = 8 hours
Meetings Per Busy Time = 5 (5 different meetings during a day)
Average Attendees Per Meeting = 15
Average Time Per Meeting = 480 (Minutes) one day
Desired Blockage = 0.01

Outputs
Arrival Rate =0.156 per minute
Traffic Intensity = 75
Ports Required = 91
Calculated Blockage = 0.0087

Notes: This scenario will support all 5 meetings with all 15 participants, but it calculates overkill due to the long service time. It is probably better to use the Minimum Port Equation.

6. **Small Group Coordination**

Minimum predictor
This is the classic predictor and should be used to determine the minimum number of ports for sizing and recognizing limits of the system.

Minimum Port Count = Maximum Connections in Meetings x max simultaneous meetings

Erlang predictor
This predictor should be used when meeting sizes are unknown and the number of meetings held during a day is known.

Inputs
Busy Time = 6 hours (Assume 10:00 a.m. to 4:00 p.m.)
Meetings Per Busy Time = 150

Average Attendees Per Meeting = 5
Average Time Per Meeting = 60 (Minutes)
Desired Blockage = 0.01

Outputs
Arrival Rate =2.083 per minute
Traffic Intensity = 75
Ports Required = 144
Calculated Blockage = 0.0088

New Model for Video Capacity Planning

Gartner forecasts 340% growth for video soft clients, defined as mobile tablets, smartphones, and desktop UC clients. This will this change many aspect of video bridging; scheduled vs adhoc, infrequent large events vs frequent small groups; but most important will be scalability of ports to meet the hyper-demand growth predicted. Taking cues from the telephony Erlang model, we are able to help clients plan for the next evolution in video infrastructure.

Please find our **Port Capacity Planning** model on our website, which will allow you to input your own meeting and service data in order to determine your ultimate port requirements.