Deriving ErlangC from first principles

by Philip Stubbs, Drakelow Consulting Ltd

philip@drakelowconsulting.com

Contents

1. Introduction

The ErlangC equation is widely used in inbound call centres around the globe. It is deployed within centres that have a number of advisors who can handle every call type, with negative-exponential distributed call arrival and service durations.

The equation calculates the proportion of calls that will be answered within a given time. The calculation inputs are this given time, together with a call volume, an average call handling time, and the number of advisors.

If you would like to calculate the number of advisors required, then you should amend the number of advisors as an input, until the equation result in the desired service level.

The equation is derived using principles from a branch of Operational Research techniques called Queueing Theory. This document includes a full proof of the ErlangC equation.

2. Queueing Theory Basics

n = number of items in system

λ_n = mean arrival rate of new items when *n* items are in system
= expected number of arrivals per unit time when *n*

μ_n = mean service rate when *n* items are in system
= expected number of items completing service per unit time

$P_n(t)$ = Probability that exactly *n* items are in the system at time *t*

s = Number of servers (in parallel channels)

π = Probability that a call arrives to find all servers busy
= Probability that caller has to wait for a server

Note that since there are no services when there is none in the system, then it follows that $\mu_0 = 0$.

Also, if $\lambda_n = \lambda$ is constant for all *n*, and the mean service rate per busy server is a constant μ, then we can define the following...

$1/\lambda$ = Mean interarrival time (sec)

$1/\mu$ = Mean service time (sec)

We can also define ρ as the utilisation factor for the service facility, defined as the expected fraction of the time the servers are busy:

$$\rho = \frac{\lambda}{s\mu} \qquad (1)$$

A system is in steady state if the number of items in the system does not tend to infinity, ie that in the long term, the system can output contents at a faster rate than others can enter.

If $\rho < 1$, then the system is in steady state.

3. State Probabilities

We start by considering a system with random arrivals and random service times.

3.1 Input Behaviour

Given *n* items in system at time *t*, the probability that exactly one arrival will occur during the time interval $(t, t+\delta t)$ is

$$\lambda_n \delta t + o(\delta t) \qquad (2)$$

If we assume that $\lambda_n = \lambda$ (constant) for all *n*, then the system has a random Poisson input. This implies that the number of arrivals in a time interval of length *t* has a Poisson distribution of parameter λt.

$$\Pr(k \text{ arrivals in time interval } t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \qquad (3)$$

The interarrival times have a negative exponential distribution with parameter λ. This verifies $1/\lambda$ as the mean interarrival time.

3.2 Output Behaviour

Given *n* items in system at time t, the probability that exactly one item leaves the system in $(t, t+\delta t)$ is

$$\mu_n \delta t + o(\delta t) \qquad (4)$$

If output is random and mean service rate per busy server is a constant μ, the service time distribution is also negative exponential with parameter μ.

$$\Pr(k \text{ completions of service in time interval } t) = \frac{(\mu t)^k e^{-\mu t}}{k!} \qquad (5)$$

3.3 Input and Output Probabilities

Consider a time period starting at time t with duration δt. Within time period $(t, t+\delta t)$,

P(No. arrivals and departures > 1) = $o(\delta t)$
P(No. arrivals and departures = 0) = $1-\lambda_n \delta t-\mu_n \delta t+o(\delta t)$ (6)

Infinitesimal order $o(\delta t)$ is assumed to be negligible as $\delta t \rightarrow 0$.

The probability $P_n(t+\delta t)$ has three components:

(i) $P_n(t)*$ Pr(zero arrivals and departures in $(t, t+\delta t)$)
$=P_n(t)(1-\lambda_n \delta t-\mu_n \delta t)$

(ii) $P_{n-1}(t)*$ Pr(one arrival and zero departures in $(t, t+\delta t)$)
$=P_{n-1}(t)\lambda_{n-1}\delta t$

(iii) $P_{n+1}(t)*$ Pr(zero arrivals and one departure in $(t, t+\delta t)$)
$=P_{n+1}(t)\mu_{n+1}\delta t$

Combining these components, we get

$$P_n(t+\delta t)=(1-\lambda_n \delta t-\mu_n \delta t)P_n(t)+\lambda_{n-1}\delta t P_{n-1}(t)+\mu_{n+1}\delta t P_{n+1}(t) \quad (7)$$

If δt now tends to zero (preserving t as a constant), the existence of all derivatives $P_n'(t)$ is established and we obtain a system (ξ) of equations as follows.

$$P_n'(t)=\frac{P_n(t+\delta t)-P_n(t)}{\delta t}=-(\lambda_n+\mu_n)P_n(t)+\lambda_{n-1}P_{n-1}(t)+\mu_{n+1}P_{n+1}(t)$$
$$P_0'(t)=-\lambda_0 P_0(t)+\mu_1 P_1(t) \quad (8)$$

A solution to this set of differential equations is available only in the steady state.

3.4 Steady State Probabilities

If steady state exists for the system (ie $\rho=\dfrac{\lambda}{s\mu}<1$),

$P_n(t)\rightarrow P_n$ as $t\rightarrow\infty$
ie $P_n'(t)\rightarrow 0$ as $t\rightarrow\infty$

$$-(\lambda_n+\mu_n)P_n+\lambda_{n-1}P_{n-1}+\mu_{n+1}P_{n+1}=0 \quad (9)$$

and $\lambda_0 P_0+\mu_1 P_1=0$

As a result,

$$P_n = P_0 \left(\frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^{n} \mu_i} \right) \tag{10}$$

Also, since $\sum_{n=0}^{\infty} P_n = 1 \quad \therefore P_0 + \sum_{n=1}^{\infty} P_n = 1$

Therefore
$$P_0 + P_0 \sum_{n=1}^{\infty} \left(\frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^{n} \mu_i} \right) = 1$$

and as a result,

$$P_0 = \left\{ 1 + \sum_{n=1}^{\infty} \left(\frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^{n} \mu_i} \right) \right\}^{-1} \tag{11}$$

4. Calculating Probability of Wait

In this section, we seek to determine the probability that a caller waits, in terms of arrival, service and number of servers. Using the equations derived in the previous section, we can be more specific about our system. The first stage is to find an expression for the number of calls in our system.

4.1 Probability that *n* in system

Since $\prod_{i=0}^{n-1} \lambda_i = \lambda^n$ for all n

and $\prod_{i=1}^{n} \mu_i = \mu.2\mu.3\mu.....n\mu = n!\mu^n$ where $n = 1, 2, \ldots, s-1$

$\prod_{i=1}^{n} \mu_i = \mu.2\mu.3\mu...s\mu.s\mu...s\mu$ where $n \geq s$

then $P_n = \frac{\lambda^n}{\mu^n n!} P_0$ where $0 \leq n < s$

(12)

and $$P_n = \frac{\lambda^n}{s!s^{n-s}\mu^n}P_0$$ where $n \geq s$

4.2 Probability that zero in system

Since we know that $\sum_{n=0}^{\infty} P_n = 1$

We can say that $\sum_{n=0}^{s-1} P_n + \sum_{n=s}^{\infty} P_n = 1$

Substituting in, we get

$$\sum_{n=0}^{s-1} \frac{\lambda^n}{\mu^n n!}P_0 + \sum_{n=s}^{\infty} \frac{\lambda^n}{s!s^{n-s}\mu^n}P_0 = 1$$

$$\sum_{n=0}^{s-1} \frac{\lambda^n}{\mu^n n!}P_0 + \frac{s^s}{s!}\sum_{n=s}^{\infty} \frac{\lambda^n}{s^n\mu^n}P_0 = 1$$

$$P_0 = \frac{1}{\left\{\sum_{n=0}^{s-1} \frac{\lambda^n}{\mu^n n!} + \frac{s^s}{s!}\sum_{n=s}^{\infty}\left(\frac{\lambda}{s\mu}\right)^n\right\}}$$

and since $\sum_{n=s}^{\infty} X^n = \frac{X^s}{1-X}$ then

$$P_0 = \frac{1}{\left\{\sum_{n=0}^{s-1} \frac{\lambda^n}{\mu^n n!} + \frac{s^s}{s!}\frac{\left(\frac{\lambda}{s\mu}\right)^s}{\left(1-\frac{\lambda}{s\mu}\right)}\right\}}$$

Finally, (13)

$$P_0 = \frac{1}{\left\{\sum_{n=0}^{s-1} \frac{\lambda^n}{\mu^n n!} + \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!\left(1-\frac{\lambda}{s\mu}\right)}\right\}}$$

4.3 Probability that call has to wait

The probability that a call has to wait for service is the probability that the system is full when that call arrives.

This probability (π) can be found by summing P_n on the occasions that $n \ge s$.

$$\pi = \sum_{n=s}^{\infty} P_n \qquad = \sum_{n=s}^{\infty} \frac{P_0 \lambda^n}{s!\, s^{n-s} \mu^n}$$

$$= \frac{P_0 s^s}{s!} \sum_{n=s}^{\infty} \frac{\lambda^n}{s^n \mu^n}$$

$$= \frac{P_0 s^s}{s!} \frac{\left(\frac{\lambda}{s\mu}\right)^s}{\left(1 - \frac{\lambda}{s\mu}\right)}$$

$$= \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!} \frac{P_0}{\left(1 - \frac{\lambda}{s\mu}\right)} \tag{14}$$

Substituting equation for P_0,

$$\pi = \frac{\dfrac{\left(\frac{\lambda}{\mu}\right)^s}{s!\left(1 - \frac{\lambda}{s\mu}\right)}}{\left\{ \sum_{n=0}^{s-1} \frac{\lambda^n}{\mu^n n!} + \dfrac{\left(\frac{\lambda}{\mu}\right)^s}{s!\left(1 - \frac{\lambda}{s\mu}\right)} \right\}} \tag{15}$$

<u>5. Distribution Function of Waiting Time</u>

We have found the probability that incoming calls must wait to be served. We must now turn our attention how the expected waiting time is distributed.

In this section, we seek to determine how probability π is distributed over time.

Let W = waiting time before being served

$P(W>t)=$ = probability that a call entering the system will wait for a time greater than t.

$P_n(W>t)=$ = this probability that this inequality is true given that call has found system in state n.

Summing the probabilities for all n,

$$P(W>t)=\sum_{n=0}^{\infty} p_n P_n(W>t)$$

We know that

$$P_n(W>t)=0 \qquad \text{when } n<s \text{ and } t\geq 0$$

Therefore

$$P(W>t)=\sum_{k=n}^{\infty} p_k P_k(W>t) \qquad (16)$$

Assume that $v=n-s$, when $n>s$, and where v = number of calls waiting.

The problem we have is to find the probability of the inequality $(W>t)$ when all the lines are occupied.

Under these circumstances, our call starts a conversation after the $(v+1)$th freeing of a line. The required probability is therefore the probability that during time t after the occurrence of our call there will occur not more than v freeings of a line.

Let's say that $q_r(t)$ is the probability that during this time there occur precisely r freeings $(0\leq r\leq v)$.

$$P_n(W>t)=\sum_{r=0}^{n-s} q_r(t) \qquad \text{where } n\geq s \qquad (17)$$

While the call waits for service, the freeings of the servers are exponentially distributed, a stream with parameter $s\mu$.

The probability that no freeing will occur during time t from the moment when all lines are occupied is

$$=(e^{-\mu t})^s=e^{-s\mu t}$$

We can define $q_r(t)$ as the probability that during time t there will occur r events of this stream.

$$q_r(t) = e^{-s\mu t}\frac{(s\mu t)^r}{r!} \qquad \text{where } 0 \le r \le v \qquad (18)$$

so that $$P_n(W > t) = \sum_{r=0}^{n-s} e^{-s\mu t}\frac{(s\mu t)^r}{r!} \qquad \text{where } n \ge s \qquad (19)$$

Examining (16) and substituting in (19),

$$P(W > t) = \sum_{n=s}^{\infty} p_n \sum_{r=0}^{n-s} e^{-s\mu t}\frac{(s\mu t)^r}{r!}$$

Substituting in equation (12), where $n = s$,

$$= e^{-s\mu t}\sum_{n=s}^{\infty}\left(\frac{\lambda}{s\mu}\right)^{n-s} p_s \sum_{r=0}^{n-s}\frac{(s\mu t)^r}{r!}$$

By multiplying out the summations, we get:

$$= p_s e^{-s\mu t}\sum_{r=0}^{\infty}\frac{(s\mu t)^r}{r!}\sum_{n=s+r}^{\infty}\left(\frac{\lambda}{s\mu}\right)^{n-s}$$

$$= p_s e^{-s\mu t}\sum_{r=0}^{\infty}\frac{(s\mu t)^r\left(\frac{\lambda}{s\mu}\right)^r}{r!}\sum_{n=s+r}^{\infty}\left(\frac{\lambda}{s\mu}\right)^{n-s-r}$$

$$= p_s e^{-s\mu t}\sum_{r=0}^{\infty}\frac{(\lambda t)^r}{r!}\sum_{n=s+r}^{\infty}\left(\frac{\lambda}{s\mu}\right)^{n-s-r}$$

$$= \frac{p_s e^{-s\mu t}}{\left(1-\frac{\lambda}{s\mu}\right)}\sum_{r=0}^{\infty}\frac{(\lambda t)^r}{r!}$$

$$= \frac{p_s}{\left(1-\frac{\lambda}{s\mu}\right)} e^{-(s\mu-\lambda)t} \qquad (20)$$

We know from (12) that $$P_n = \frac{\lambda^n}{s!\,s^{n-s}\mu^n}P_0$$

Therefore for $n = s$, $$P_s = \frac{\lambda^s}{\mu^s s!}P_0 \qquad (21)$$

We know from (14) that $\pi = \frac{\left(\frac{\lambda}{\mu}\right)^s}{s!} \frac{P_0}{\left(1 - \frac{\lambda}{s\mu}\right)}$

Combining (14) and (21) gives us $\pi = \frac{P_s}{\left(1 - \frac{\lambda}{s\mu}\right)}$ (22)

And from (20), $P(W > t) = \pi e^{-(s\mu - \lambda)t}$ where $t \geq 0$

which is the solution to our problem. Further, we can see that the probability of waiting,

$P(W > 0) = \pi$ as we would expect.

6. Summary

We have relationships that estimate the probability that a call has to wait, and a probability the call will wait for a certain number of seconds. These probabilities can be calculated from the number of servers, the arrival rate, the service rate and the service time.

6.1 Calculating number of servers

Starting with one server, the service level is calculated. The service level is then calculated time and time again with an ever increasing number of servers. The number of servers that just overachieves the service level can be selected as the number of agents required for that half hour.

6.2 ErlangC assumptions

The ErlangC model makes the following assumptions, so make sure you are aware of these before using the model:

(1) It is assumes that there are enough telephone lines for people to wait on.

(2) It is assumed that people will queue indefinitely – in practice this will be fine just as long as abandonment rates are low.

(3) The resulting advisor count does not include allowances for absences, so make sure you increase the number to account for within day absences as well as full-day absences.