REPORTE N° 9
GPR

Determinación de Densidad Óptima de Facilidades: El caso del Teléfono de Uso Público (TUP)

Juan Manuel García, Miguel Martínez y Pamela Medina *

Gerencia de Políticas Regulatorias

Organismo Supervisor de Inversión Privada en Telecomunicaciones | OSIPTEL

Resumen

El objetivo de este documento es determinar el número óptimo de personas deben incluirse en el ámbito de cobertura de cada Teléfono de Uso Público (TUP). Como la ubicación de los TUP se debe basar en el bienestar de los usuarios de este servicio, se asume que este debe permitir reducir el tiempo de espera (y la distancia promedio) que enfrentan los usuarios del TUP. Con este fin se desarrolla un modelo de colas que estima el tiempo de espera máximo para utilizar un TUP (y la distancia recorrida hasta el TUP) para distintas número de TUP por cuadra en zonas urbanas. Los resultados de simulación de este modelo permiten inferir la densidad necesaria de TUP (por cada 1000 personas) para obtener una calidad adecuada de servicio.

Palabras clave: Telefonía pública, Telecomunicaciones, Facilidades, Densidad.
http://www.osiptel.gob.pe

---

*. El documento se ha beneficiado de los comentarios de José Gallardo. Los autores agradecen la valiosa asistencia de Eduardo Salazar. Las opiniones vertidas en este documento son de responsabilidad exclusiva delos autores y no reflejan necesariamente la posición del OSIPTEL. Remitir comentarios y sugerencias a: jgarcia@osiptel.gob.pe; mangel@osiptel.gob.pe; pmedina@osiptel.gob.pe

1. Introducción

En el Perú, los servicios de Telecomunicaciones se caracterizan desde hace unos años por un dinamismo sin precedente. Este hecho ha suscitado la atención deestudios especializados que buscan mostrar las causas de este crecimiento, su alcance y, desde luego, comparaciones con el desempeño de dicho sector en otras economías emergentes. Así, es posible encontrar estudios recientes que analizan el desempeño de estos servicios fundamentalmente en base a datos de penetración y cobertura; y en menor medida, con porcentajes de acceso[1]. Estos indicadores permiten mostrar una idea general del desarrollo de tales servicios y el aumento del número de personas que se pueden intercomunicar gracias a ellos.

No obstante, dentro de estos servicios, el servicio de telefonía de uso público (TUP) presenta características peculiares derivadas de su naturaleza pública que necesitan de especial atención. Este servicio tiene entre sus funciones primordiales que es utilizado en casos de emergencia y que el uso regular en zonas donde no existe unalto porcentaje de acceso a través de otros medios de comunicación. En las zonas subatendidas, dicho servicio es de mayor valor ya que el acceso a telefonía por parte de los hogares es bajo (17% en las zonas rurales) y además, se ha encontrado en la literatura que los retornos de la infraestructura de servicios públicos en zonas alejadas o rurales superan a los que se obtiene en zonas con demanda ya atendida. En este sentido, es de vital importancia que estos terminales se encuentren instalados de tal manera que toda la población pueda tener acceso a los mismos sin tener que incurrir en costos muy elevados. Sus costos se verán reflejados en el tiempo incurrido por el usuario esperando en cola para el uso del servicio, así como por el invertido en desplazarse hasta llegar al terminal.

El problema, entonces, es que dada la capacidad limitada que tienen estos terminales para abastecer a toda la población en todo momento del tiempo (solo puede hablar un individuo en cada momento), se suscitan problemas de congestión y formación de colas para el uso del servicio de telefonía pública. Si esta congestión es severa, el servicio no está brindando una calidad suficiente pues el usuario no puede comunicarse. Por otro lado, el tiempo necesario para el desplazamiento al terminalmás cercano también puede ser un desincentivo a la comunicación, y en los casos más extremos, una barrera de acceso al servicio.

[1]. Ver Gallardo, López y González (2007).

En ese sentido, existe un consenso en que un tiempo de espera muy largo para el uso del servicio no es deseable en sí mismo e incluso es inaceptable en situaciones consideradas de emergencia. Por esta razón, el documento busca formular un criterio para la determinar la densidad mínima óptima de facilidades, específicamente de terminales de TUP, de tal manera que éstas permitan que los usuarios puedan hacer uso de este servicio eficientemente.

Con la finalidad de realizar dicho análisis, se ha revisado la literatura relevante sobre ubicación óptima de facilidades inmóviles cuando existen problemas de congestión, basada principalmente en modelos de colas; lo cual permite superar teóricamente los enfoques que se sustentan en comparaciones internacionales sobre la oferta de los mismos. Los supuestos claves del modelo son la aleatoriedad de la demanda y la capacidad limitada de los terminales. En base al modelo desarrollado se realiza una simulación de efectos de distintas distribuciones de TUPs.

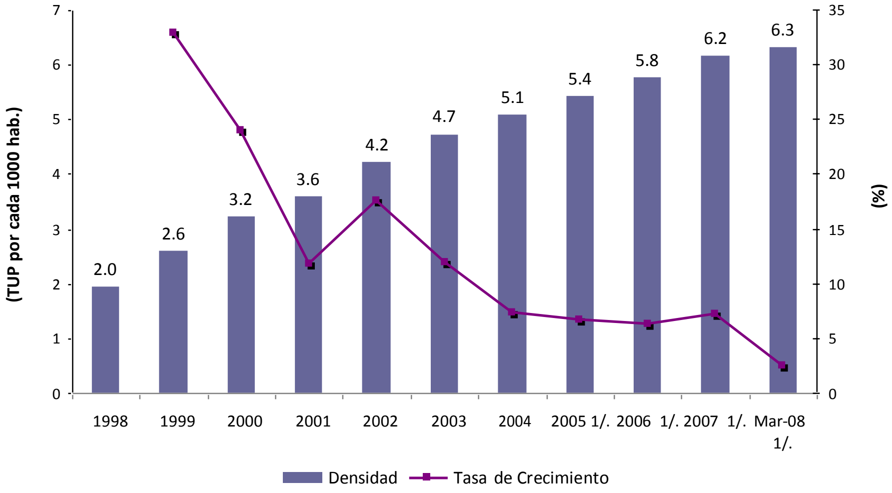
En la sección 2, se discuten las características del servicio de TUP. En la tercera sección, se realiza una descripción del modelo de colas de tipo M/M/1 usado para comprender el problema analizado. En sección 4, se muestra los supuestos y las simulaciones para distintos escenarios en el ámbito urbano. Finalmente, en la sección 5 se concluye y se incorporan las recomendaciones de política pertinentes.

2. Características y Evolución del Servicio TUP en Perú

En el caso del servicio de telefonía pública (TUP), este funciona como una alternativa de acceso a medios de comunicación telefónica para los hogares o zonas con baja cobertura de redes de servicios de telefonía fija o telefonía móvil.

En lo que respecta a la densidad de líneas en servicio a nivel nacional, ésta ha venido creciendo constantemente desde 2 líneas por cada 1000 habitantes en 1998 hasta llegar a triplicarse en una década, con 6.3 líneas por cada 1000 habitantes a marzo del 2008. No obstante, se debe tomar en cuenta que la tasa de crecimiento anual ha sido cada vez más baja, por ejemplo, en 1999 la tasa de crecimiento fue de 33%; mientras que desde el 2004 la tasa de crecimiento ha estado en torno al 7% anual. Es decir, el crecimiento de las líneas en servicio poco a poco se ha retrasado conrespecto al que presenta la población.

Gráfico No. 1
Densidad de TUP y tasa de crecimiento anual.
(1998 - 2008)



Fuente: OSIPTEL.
1/ Se ha calculado el indicador de densidad por departamento con las nuevas proyecciones de población a partir del censo de 2005.

Por otro lado, si se analiza esta información a nivel departamental es claro que, a cifras de marzo 2008, Lima y Callao presentan la mayor densidad de líneas de teléfonos públicos (11.1 líneas por cada 1000 habitantes) seguidos de Tacna, Moquegua y Arequipa con 9.7, 8.2 y 6 líneas por cada 1000 habitantes. Asimismo, en los últimos cuatro años, casi todos los departamentos presentan incrementos en la densidad con las excepciones de Arequipa y Ayacucho. La primera de ellas se encuentra asociada con un descenso en el número absoluto de líneas. Por su parte, Ayacucho muestra este resultado producto de un crecimiento poblacional mayor que elnúmero de líneas en dicho departamento. Por su parte, el crecimiento de la densidad TUP en La Libertad fue prácticamente nulo, posiblemente afectado por un bajo crecimiento en el número de líneas.

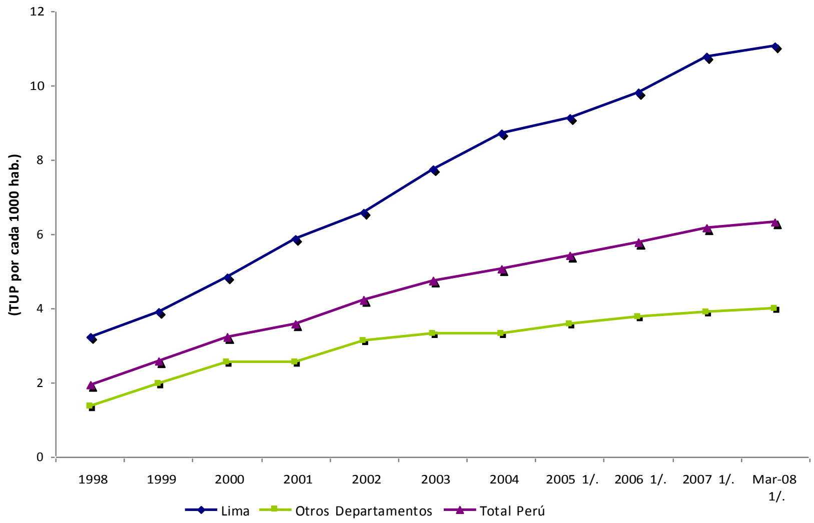
Gráfico No. 2
Densidad de TUP por departamentos.
(2004 - 2008)

Fuente: OSIPTEL.

En cuanto al número de líneas en servicio sí existe una diferencia significativa pues Lima ha registrado un crecimiento muy superior al de otros departamentos con un aumento de 24 432 nuevas líneas desde el 2004 a marzo 2008, comparadas con las 1 762 líneas que se han puesto a disposición del público en Piura, departamento que ocupa el segundo lugar en dicha lista. Hay que resaltar el hecho que casi todos los departamentos crecieron en este periodo excepto Arequipa, como se había mencionado anteriormente, donde el número de líneas disminuyó en 1.2%; mientras que otros lo hicieron a tasas muy bajas como La Libertad, Ayacucho y Junín que crecieron apenas 5, 7 y 8%, respectivamente.

Asimismo, no solo la densidad de Lima es mayor en términos absolutos sino que además ésta viene creciendo más rápidamente que la correspondiente a otros departamentos del Perú. Ello se puede observar en la inclinación más pronunciada que muestra la pendiente de la evolución de la densidad de TUPs en Lima con respecto a la de otros departamentos, tal y como se aprecia en el gráfico siguiente.

Gráfico No. 3
Densidad de TUP en Lima y Otros Departamentos.
(1998 - 2008)

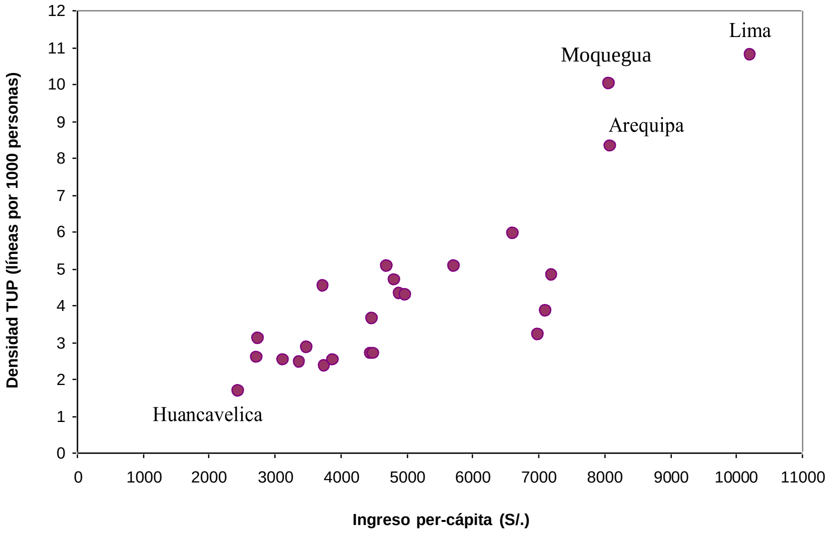


Fuente: OSIPTEL.

Sin embargo, a pesar de dichos incrementos en densidad y número de líneas en servicio, actualmente este servicio no ha logrado aún la cobertura necesaria en muchas regiones del país para poder cumplir con el objetivo de acceso universal y posteriormente, servicio universal.

En general, la densidad de teléfonos de uso público es mayor en los departamentos con mayores niveles de ingreso per cápita, encontrándose ella concentrada en los departamentos de Lima, Moquegua y Arequipa; mientras que el departamento de Huancavelica es el que cuenta con la menor densidad y los menores ingresos per cápita mensuales. Así, existe un vasto sector de la población que necesita con mayor urgencia este tipo de servicio público de telecomunicaciones por tener problemas económicos para adquirir otros medios, pero que tampoco se encuentra cubierta.

Gráfico No. 4
Densidad de TUP según población
y nivel de ingreso per-cápita por departamento (S/.)

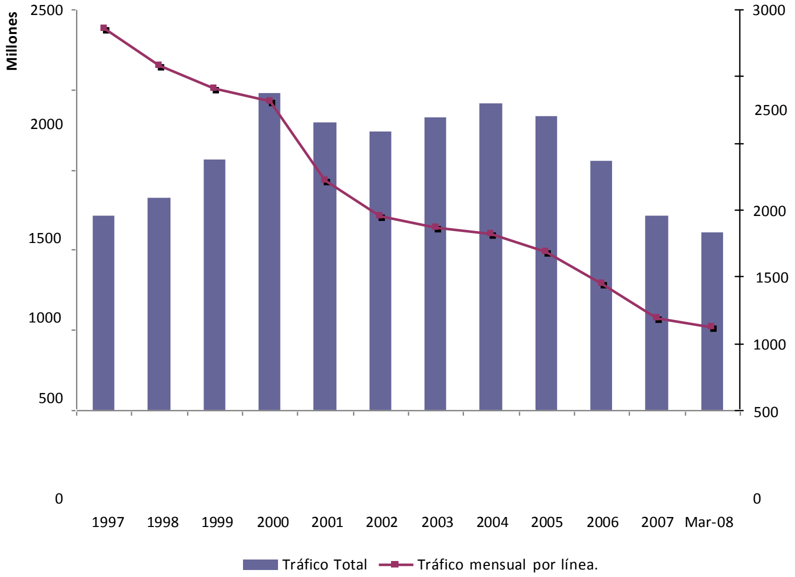


Fuente: Gerencia de Políticas Regulatorias – OSIPTEL.

En lo que respecta al tráfico cursado a través de estos terminales, se evidencia en los últimos años un tema que pone en tela de juicio la sostenibilidad de este servicio. Por ejemplo, si bien el tráfico local por la modalidad de TUP de Telefónica del Perú fue creciente entre 1997 y el 2000 y estable hasta el 2004, desde el año 2005 el tráfico local agregado de Telefónica del Perú ha venido descendiendo sobre todo en el 2007.

La magnitud de este decrecimiento se puede evidenciar mejor en la serie de tráfico mensual por línea, es decir, el tráfico promedio cursado en cada uno de los terminales TUP. Como se observa en el gráfico mostrado a continuación, dicha serie muestra una fuerte caída en los últimos años, posiblemente derivada de la aparición de servicios casi sustitutos como por ejemplo el servicio móvil, los teléfonos instalados en las bodegas, entre otros.

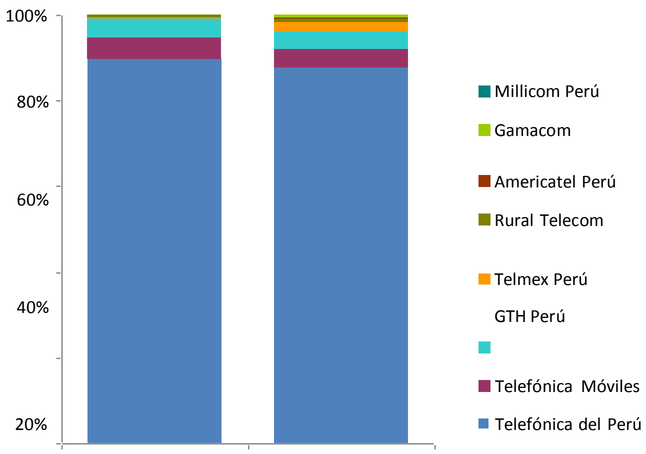
Gráfico No. 5
Tráfico total y mensual por línea a través de TUP
(Minutos)



Fuente: OSIPTEL.

Finalmente, para tener una idea de la estructura del mercado, la empresa que posee el mayor número de líneas es Telefónica del Perú que supera el 80% del total de líneas en el país, seguida por Telefónica Móviles y GTH Perú. Aunque estas tres empresas han visto reducidas su demanda en el mercado, la estructura del mercado no ha cambiado mucho desde el 2004. Un punto importante es que, desde su aparición en este mercado en el 2002, Telmex Perú se encuentra creciendo cada año a tasas muy importantes, especialmente porque sus terminales TUP se triplicaron en el año 2005 con respecto al anterior. Otras empresas que han empezado a aumentar su númerode líneas en servicio son Rural Telecom y Gamacom.

Gráfico No. 6
Participación de operadores en el servicio de TUP
(% de líneas en servicio por empresa)



0%

2004 Mar-08

Fuente: OSIPTEL.

3. Marco teórico

La presente sección se basa en una descripción teórica de los dos problemas que enfrenta la autoridad de política para determinar la densidad óptima de TUP. En primer lugar, se presenta el problema de congestión y el modelamiento del tiempo promedio de espera de un usuario en la cola. Por otro lado, además de tomar en cuenta el problema del tiempo de espera en cola para el uso del servicio, dado que el problema se centra en terminales inmóviles, se tiene que modelar como un costo adicional el tiempo que el usuario utiliza viajando hacia la facilidad por servicio. Luego, se explicala determinación del tiempo promedio de llegada hacia la facilidad. Finalmente, se argumenta cómo ambos costos deben ser combinados de tal manera que se cumplan los parámetros de calidad previamente designados por la autoridad regulatoria.

a. Modelos de Congestión y Determinación del Tiempo de Espera en la Cola

En los modelos teóricos acerca de la ubicación óptima de terminales, cuando existen problemas de congestión el principal objetivo consiste en evitar que un usuario quede fuera del sistema, es decir, que no sea atendido por el mismo dentro de ciertos parámetros fundamentales de calidad. Por ejemplo, en el caso de facilidades que se usen específicamente para casos de emergencia como postas médicas de emergencia o maternidades, el tiempo de espera promedio óptimo debería ser muy cercano a cero; en otros casos, como en el del servicio de TUP o los cajeros automáticos, el tiempo de espera debería ser también reducido. Así, lo que se busca con estos modelos es minimizar la cantidad de usuarios no atendidos bajo dichas condiciones. En el presente documento, se asumirá que dichas condiciones de calidad se refieren a tiempos máximos permitidos por la autoridad.

En ese sentido, el objetivo de esta sección es obtener el tiempo de espera promedio en cola que enfrentará un usuario del sistema al llegar a la facilidad (TUP) en el horario diurno[2]. Este tiempo se puede modelar como un proceso de nacimiento y muerte, el cual deriva, bajo ciertos supuestos, en un modelo de colas tipo M/M/1, el cual se detalla a continuación.

[2]. El análisis solo se realiza en el horario diurno debido a la función probabilística usada para la tasa de llegada al sistema. Con este supuesto, existen problemas con el horario nocturno dado que

Todo el análisis se realiza tomando en cuenta al *sistema*, es decir, a la totalidad de usuarios presentes inicialmente. Del mismo modo, se consideran *usuarios* de una facilidad a aquellas personas que se encuentran situadas en un vecindariodeterminado por número de cuadras. Se asume además que dichos vecindarios son disjuntos.

Estado del Sistema

En primer lugar, el *estado del sistema en el tiempo t* se define como el número de personas que se encuentra en el sistema en el tiempo t. Dicho número toma en cuenta tanto a aquellas personas que usan el servicio como a las que se encuentran en cola esperando para usarlo. Así, en el tiempo t=0, el estado del sistema será el número total de personas presentes inicialmente.

Proceso de arribo

El proceso de arribo es el proceso insumo del modelo. Los arribos son llamados usuarios del sistema. Se asume que en cada momento del tiempo solo es posible un arribo como máximo[3].

Por otro lado, se supone que el tiempo entre arribos es una variable aleatoria continua e independiente; es decir, el valor de arribos entre t-1 y t no depende de ningún otro valor de arribo pasado o futuro. Asimismo, se asume que el tiempo entre arribos es estacionario, lo que quiere decir que la distribución del tiempo entre arribos es independiente del momento particular del día.

En particular, se asume que la distribución del tiempo entre arribos a la facilidad es exponencial con parámetro λ, la cual es comúnmente es usada en este tipo de modelos. Esta distribución presenta como una característica importa la propiedad de falta de memoria; es decir, para determinar la distribución de probabilidad del tiempo necesario hasta un nuevo arribo no interesa cuanto tiempo ha pasado desde el último arribo. Dicho enunciado se puede formalizar como:

$$P(A > t + h \mid A \geq t) = P(A > t)$$

---

3. Dicho supuesto puede ser poco realista en casos como el de bancos o restaurantes.

donde A es una variable aleatoria similar a la descrita anteriormente con distribución exponencial.

De dichos supuestos se deriva que el número de arribos tiene una distribución Poisson con parámetro λt. En el caso de que el tiempo t tenga el valor de un minuto, como ocurre en nuestro caso, el parámetro de la distribución del Poisson es igual a λ, el valor esperado del número de arribos.

Proceso de servicio

Se denomina proceso de servicio a aquel proceso que caracteriza la utilización efectiva de la facilidad. En el caso analizado, el proceso de servicio corresponde a la duración de la llamada.

Se asume que el proceso de servicio para varios usuarios es independiente y se distribuye como una función exponencial con parámetro μ , y por tanto, media $1/\mu$. Como μ se define en función a usuarios por unidad de tiempo se le conoce como tasa de servicio.

Así, la probabilidad de que un usuario complete el servicio entre t y Δt es:

$$\int_0^{\Delta t} \mu e^{-\mu t} dt = 1 - e^{-\mu t} \approx \mu \Delta t$$

De este modo, la probabilidad *p* de que se termine la llamada antes de un periodo de un minuto es:

$$p = 1 - e^{-\mu t} \approx \mu$$

Este proceso también cuenta con la propiedad de falta de memoria, ya que dada la distribución exponencial de duración de llamadas, para determinar la probabilidad de que una llamada acabe en momento no interesa cuando hay empezado la misma.

Especificación del Sistema y procesos de nacimiento-muerte

Según la notación propuesta por Kendall-Lee para modelos de colas, el sistema a estudiar es uno conocido como M/M/1[4]. Dicha clasificación quiere decir que tanto el tiempo entre arribos como el tiempo de servicio se definen como variables aleatorias independientes e idénticamente distribuidas (i.i.d.), las cuales cuentan con distribuciones probabilísticas exponenciales; por ello, la letra M. No se especifican restricciones acerca del ordenamiento del sistema para la atención de nuevos clientes. Se asume que se atienden de acuerdo al orden natural de la cola. Finalmente, se asumirá que solo existe una única facilidad en el modelo, la cual posee a su vez un único servidor.

Dichos modelos de colas pueden ser, a su vez, modelados como procesos de nacimiento-muerte; es decir, procesos estocásticos en tiempo continuo para los cuales el estado del sistema es un entero no negativo.

El proceso de arribos es considerado como el proceso de nacimientos en el sistema, de tal manera que la probabilidad de que ocurra un nacimiento durante el intervalo $[t, t + \Delta t]$ es:

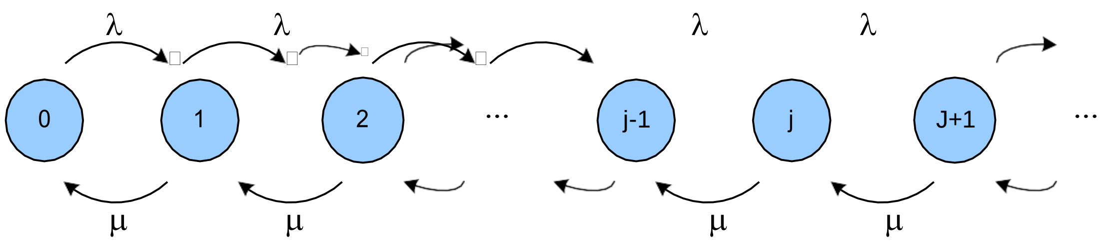
$$\int_0^{\Delta t} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t} \approx \lambda \Delta t$$

Lo que quiere decir que la tasa de nacimiento en el estado j es simplemente la tasa de arribo λ .

Del mismo modo, se determina la tasa de muerte en el tiempo t. Si el estado del sistema es 0, no existe nadie usando el servicio; por tanto, la tasa de muerte es igual a cero ($\mu_0 = 0$). Si el estado en el tiempo es $j \geq 1$, se sabe que exactamente un usuario estará en el servicio. Entonces, la probabilidad de que un usuario complete el servicio entre $[t, t + \Delta t]$ es $\mu \Delta t$. Si tomamos periodos de tiempo de un minuto, entonces la probabilidad de que el usuario complete el servicio en menos de este periodo es simplemente μ . Así, si $j \geq 1$, la tasa de muerte será $\mu_j = \mu$.

[4]. Ver Winston (2004).

Dicho proceso de nacimiento y muerte se puede representar gráficamente como:



Matriz de probabilidades de transición

A partir de los supuestos sobre las probabilidades de que llegue una persona más a la facilidad y de que una persona que se encuentre en ella termine su llamada, se desea calcular la matriz de probabilidades de que el sistema se encuentre en un estado dado estado inicial.

A dichas probabilidades se les conoce como las probabilidades de transición (P_{ij}), las cuales se encuentran condicionadas al estado del sistema en el pasado inmediato. Así, se define $P_{ij}(t)$ como la probabilidad de que j personas se encuentren presentes en el sistema en el tiempo t, dado que en el tiempo 0, i personas se encuentran presentes.

Para obtener la matriz con todas las probabilidades de transición del sistema construyen dos matrices insumo.

a) Se construye la matriz cuadrada D de dimensión N+1, de probabilidades mantenerse o salir del servicio, a partir de la probabilidad de que una llamada termine durante el periodo de observación, es decir en menos de 1 minuto, donde:

- ✓ $D(i,i) = p$: Mantenerse en el estado con *i* personas.
- ✓ $D(i,i) = 1 - p$: Moverse del estado con *i* personas a uno con *i-1* personas.

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ p & 1-p & 0 & 0 & \cdots \\ 0 & p & 1-p & 0 & \cdots \\ 0 & 0 & p & 1-p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

b) Se construye la matriz cuadrada A de dimensión N+1, de probabilidades de arribo de un número de personas (empezando desde 0), donde:

- ✓ $A(i,i) = a(0)$: Probabilidad de ningún arribo.
- ✓ $A(i,i+h) = a(h)$: Probabilidad de *h* arribos.
- ✓ $A(i,N+1) = 1 - suma \;\; fila\ i$: Para que las probabilidades sumen 1.

$$A = \begin{bmatrix} a(0) & a(1) & a(2) & a(3) & \cdots & 1-\sum_{i=0}^{2} a(i) \\ 0 & a(0) & a(1) & a(2) & \cdots & 1-\sum_{i=0}^{2} a(i) \\ 0 & 0 & a(0) & a(1) & \cdots & 1-\sum_{i=0}^{2} a(i) \\ 0 & 0 & 0 & a(0) & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \ddots & 1-\sum_{i=0}^{2} a(i) \\ 0 & 0 & 0 & 0 & 0 & \vdots \\ & & & & & 1 \end{bmatrix}$$

c) Se obtiene la matriz cuadrada P de dimensión N+1 para el número de personas en el sistema multiplicando ambas matrices (P=DxA), donde P_{ij} es igual a la probabilidad de que el sistema se mueva al estado j dado que está en el estado i.

$$P = \begin{bmatrix} a(0) & a(1) & a(2) & \cdots & 1-\sum_{i=0}^{2} a(i) \\ pa(0) & pa(1)+(1-p)a(0) & pa(2)+(1-p)a(1) & \cdots & p(1-\sum_{i=0}^{2} a(i))+(1-p)(1-\sum_{i=0}^{1} a(i)) \\ 0 & pa(0) & pa(1)+(1-p)a(0) & \cdots & p(1-\sum_{i=0}^{1} a(i))+(1-p)(1-a(0)) \\ 0 & 0 & pa(0) & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & p(1-a(0))+(1-p).1 \end{bmatrix}$$

Como se observa en la matriz, la primera línea se puede leer como el número de arribos necesarios para pasar a un estado j luego de determinado tiempo, dado que se inicia con el sistema con 0 usuarios. Por otro lado, en las filas siguientes, se parte de estados del sistema con una persona o más; por ello, los caminos por los que se puede llegar a un estado observado en particular se diversifican.

Por ejemplo, si partimos de que existe un individuo en el sistema y se desea mantener el sistema en dicho estado, para el siguiente momento existen dos posibilidades, que el individuo termine la llamada y un nuevo usuario arribe a la facilidad, o que el individuo no termine la llamada y que no llegue otro usuario. Finalmente, como se asume que la facilidad solo puede atender a un usuario a la vez, la probabilidad de que el estado del sistema pase de tener j+2 personas a j debe ser 0, pues no existe dicha posibilidad. De esta manera se rellenan todas las celdas de la matriz P.

Vector de Probabilidades de Estado Estacionario

Las probabilidades de transición, para periodos de tiempo largos, se aproximan a un límite π_j, el cual se define como la probabilidad de tener *j* usuarios en el sistema independientemente del momento inicial *i*.

$$\lim_{t\to\infty} P_{ji}(t) = \pi_i$$

Estas probabilidades se estiman como valores de un *“estado estacionario” del sistema* cuando pasa el tiempo, es decir la probabilidad del estado *j.* Partiendo desde cualquier estado en un momento de equilibrio dicha condición, se puede interpretar como que, dada una distribución de personas entre estados determinada por π , si se aplica esta sobre ella la matriz de transiciones, las proporciones de personas en cada estado deben mantenerse. El algoritmo utilizado es similar a aquel que nos permite encontrar un punto fijo en el sistema, de tal manera que se cumple la siguiente condición.

$$\pi.P = \pi$$

$$[\pi_0 \quad \pi_1 \quad \pi_2 \quad \pi_3 \quad \cdots]\begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ P_{30} & P_{31} & P_{32} & P_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} = [\pi_0 \quad \pi_1 \quad \pi_2 \quad \pi_3 \quad \cdots]$$

Con esta condición, se obtiene π a partir de un sistema de ecuaciones de flujo balanceado o de conservación del sistema, dado que en estado estacionario la tasa a la cual ocurren las transiciones hacia el estado *i* es igual a la tasa de las transiciones hacia fuera del estado *i.* Es decir, el mismo número de usuarios debe salir y entrar al sistema. En caso contrario, se pierde el equilibrio.

Si se analiza una de las filas de dicha igualdad, se tiene que la probabilidad de que el sistema se encuentre en el estado i debe cumplir:

$$\pi_0 P_{0i} + \pi_1 P_{1i} + \pi_2 P_{2i} + \pi_3 P_{3i} + \ldots + \pi_j P_{ji} = \pi_i$$

Dicha expresión se puede leer como la probabilidad total de que el sistema se encuentre en un estado i, donde el lado izquierdo de dicha expresión corresponde a la suma de las probabilidades de que se pase de algún estado inicial al estado i,ponderada por la probabilidad de encontrarse en ese estado inicial.

Reordenando dicho enunciado, se puede mostrar que esta implica el cumplimiento del balance de flujos de usuarios en el servicio.

$$\underbrace{\pi_i(1-P_{ii})}_{\substack{\text{Probabilidad de salir} \\ \text{del estado con i personas}}} = \underbrace{\pi_1 P_{1i} + \pi_2 P_{2i} + \ldots + \pi_{i-1} P_{i-1,i} + \pi_{i+1} P_{i+1,i} + \ldots + \pi_j P_{j,i}}_{\text{Probabilidad de llegar a i desde otros estados}}$$

En la implementación, π se suele obtener mediante un proceso iterativo que parte del valor inicial $\bar{\pi}$, el cual asume que la probabilidad de que el sistema se encuentre en un estado determinado se distribuye uniformemente. Así, se parte del siguiente valor inicial:

$$\bar{\pi} = \left[\frac{1}{N+1} \quad \frac{1}{N+1} \quad \frac{1}{N+1} \quad \cdots \quad \frac{1}{N+1}\right]$$

Si dicho vector no cumple la condición, es reemplazada por el resultado de su multiplicación con la matriz P y se prueba nuevamente si la condición se cumple. Este proceso se realiza numerosas veces de tal manera que se obtenga convergencia a un vector π .

Una vez obtenido el vector de probabilidades de estado estacionario, se puede hallar el número promedio de personas en el sistema (L).

$$L = \begin{bmatrix} 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \sum_{i=0}^{3} (i)\pi_i$$

Luego, mediante el *Teorema de Little*[5], el tiempo promedio en el sistema (T) se define como:

$$T = \frac{L}{\lambda} = \frac{\sum_{i=0}^{3} (i)\pi_i}{\lambda}$$

Debe notarse que dicho tiempo de espera T incluye aquel que un individuo usa llamando por teléfono.

---

[5]. Sea un sistema de colas con cualquier distribución de llegadas y servicios, sea L el númerode personas presentes en el sistema en el estado estacionario, T es tiempo medio de espera en el estado estacionario y λ la razón de llegada. El Teorema de Litltle establece lasiguiente relación (Winston; 2004):

$$L = \lambda T$$

Es decir, el número esperado de personas se puede calcular como el producto del tiempo en el sistema por la tasa de llegada.

No obstante, el objetivo es obtener el tiempo promedio de espera de un individuo en la cola (T_c). Por ello, la fórmula se modifica de tal manera que se tiene:

$$T_c = \frac{L_c}{\lambda} = \frac{\sum_{i=1}^{3}(i-1)\pi_i}{\lambda}$$

Donde se utiliza el número de personas promedio esperando en cola, en lugar del número de personas promedio presentes en todo el sistema; es decir, se excluye del promedio a aquellas que se encuentran haciendo uso efectivo del servicio.

b. Determinación del tiempo de recorrido hasta la facilidad

La siguiente fuente de costos que un usuario debe incurrir para llegar a una facilidad es el costo de traslado hacia la misma. En este punto, es importante notar que, al tratarse de terminales de uso público, es deseable que las mismas sean accesibles a la población sin la necesidad de trasladarse grandes distancias a pie o de tener que usar algún medio de transporte para poder hacer uso del servicio.

En este sentido, definimos la distancia promedio que recorre una persona para llegar a un TUP como el número de cuadras entre TUP. Esta distancia genera un tiempo de desplazamiento T_d asumiendo que una persona se desplaza a una velocidad pedestre promedio de 0.97 m/s (3.5 Km/h). Para encontrar dicho tiempo, entonces, se usa la fórmula:

$$d = v.t$$

Donde d representa la distancia; v, la velocidad del usuario y t, el tiempo necesario para recorrer dicha distancia.

c. Determinación del costo de uso del servicio

En base a la determinación de ambos tiempos que necesariamente el usuario de un terminal de TUP incurre para poder acceder al mismo, se construye un indicador final del costo de acceso al servicio, basado en el tiempo total invertido por una persona hasta concretar una llamada.

Dicho costo se mide como la media geométrica de ambos tiempos[6]:

$$T = T_c^{0.5} . T_d^{0.5}$$

Esta media geométrica también se puede entender como una función de desutilidad Cobb-Douglas que pondera de la misma forma a los dos tiempos, es decir, el usuario valora su tiempo de la misma manera, sin importar si lo utiliza esperando en cola o transportándose hasta el terminal de TUP.

d. Ajustes a la tasa de arribo

Durante la explicación del modelo, no se ha tomado en cuenta el posible impacto que la distancia tenga sobre la tasa de arribo. Por ejemplo, mientras un usuario se encuentre más alejado del terminal de TUP más cercano, este se desincentiva de realizar su llamada mediante TUP. Por tanto, conforme el vecindario de análisis sea mayor, menos porcentaje de personas con respecto al total desearán utilizar esta facilidad; así, la tasa de arribo debe tener una relación inversa con el número de personas que usan TUP en el vecindario cubierto por dicho terminal.

Luego, para capturar dicho efecto, se corrige el tiempo con un porcentaje de pérdidas de los arribos, función que toma la forma:

$$prd(d) = e^{(d/m).r} / e^{r}$$

Donde:

d: variable de distancia entre usuario y TUP.

m: parámetro de distancia máxima que un usuario está dispuesto a caminar hasta un TUP.

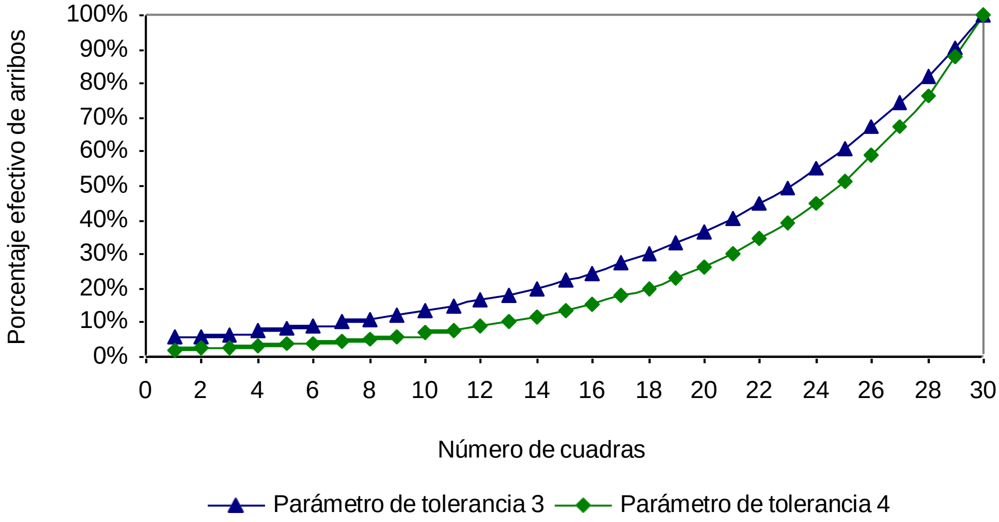
r: parámetro de intolerancia (si es más alto, las pérdidas son más aceleradas).

Así, se utiliza un factor de ajuste para obtener la tasa de arribo efectiva:

$$\lambda_e(d) = \lambda(1 - \mathrm{prd(d)})$$

[6]. Se utiliza esta especificación y no el promedio aritmético porque este último es muy sensible a cambios en alguno de los tiempos.

Por ejemplo, si asume que la distancia máxima entre TUP es de 30 cuadras (es decir, las personas caminan en promedio alrededor de 15 cuadras para llegar a un TUP), se tiene las siguientes funciones de pérdida de arribos (se asumen parámetros de intolerancia de 3 y 4):



4. Simulación

En esta sección se presenta la simulación del modelo en zonas urbanas, es decir, centros poblados con más de 400 viviendas contiguas[7]. Asimismo, dentro de estas zonas, se realiza la distinción entre los ámbitos de Lima Metropolitana y el Resto Urbano por las diferencias que presentan en cuanto a uso y acceso a TUP.

En cada uno de ellos, se debe plantear un nivel de calidad exigido por la autoridad de política. En este caso, la calidad se medirá como el valor máximo del tiempo esperado para hacer uso del TUP. Luego, los criterios empleados son: en Lima Metropolitana se permite que las personas esperen aproximadamente hasta un minuto y en el resto urbano dos minutos. Esto se debe al menor costo de oportunidad del tiempo, el cual se refleja en remuneraciones por hora mucho menores que en Lima Metropolitana. Además, en el resto urbano las personas se encuentran dispuestas a esperar más tiempo pues existen menos servicios de telecomunicaciones que se puedan presentar como sustitutos del TUP (baja densidad de TUP, menor existencia de bodegueros, chalequeros y móviles).

[7]. De acuerdo a la definición normalmente utilizada por el INEI (por ejemplo, en la Encuesta Nacional de Hogares)

En esta sección se asume que la duración de llamadas sigue una distribución Poison y que los arribos a los terminales siguen una distribución exponencial. Los valores de los parámetros se obtienen de la información provista por las operadoras sobre llamadas según duración para el año 2007.[8]

El primer paso consiste en estimar los tiempos de espera en cola y desplazamiento para distintos números de cuadras entre TUP (y, por tanto, diferente número de usuarios del servicio) en los dos escenarios.

a. Escenario: Lima Metropolitana

Este escenario cuenta con número mayor de personas por cuadra que el Resto Urbano pues se asume un mayor número de personas por metro cuadrado de viviendadebido, por ejemplo, a mayor presencia de edificios, quintas, entre otros.

Por otro lado, la probabilidad que la llamada termine en un minuto es de 0.4. Dicha probabilidad corresponde a la tasa de muerte detallada en la sección anterior. Este número se obtiene a partir de la distribución de llamadas a diciembre 2007, pues se encuentra que el 40% de las llamadas terminaron dentro del primer minuto. Dicho dato corresponde al total nacional, pero, a falta de información detallada por ámbito geográfico, es usado en ambos escenarios.

Se supone, además, que el ratio de arribo promedio al TUP de 0.12. Dicha tasa es moderadamente mayor al valor de arribo promedio efectivo (alrededor de 0.1), debido a que este representa solo a las personas que realizaron efectivamente una llamada y no a las que intentaron llamar y sus llamadas no se concretó.

Con dichos parámetros, se obtiene el tiempo promedio de espera en la cola. De la misma manera, se obtiene el tiempo promedio de traslado hacia la facilidad de acuerdo al número de cuadras que se supone se encuentran alrededor del terminal.

[8]. Se considera los datos para el horario diurno o normal a fin de hacer realista la simulación (lunes a sábado de 7 a.m. a 10:59 p.m.)

Gráfico No. XX:
Simulación y valores óptimos de densidad de TUP – Lima Metropolitana

No de cuadras	Densidad de telefonía pública			Tiempo de espera (minutos)	Tiempo hasta llegar al TUP	Media geométrica del tiempo
1 cuadra	**7.41**	**5.56**	**6.48**	**0.83**	**0.72**	**0.77**
2 cuadras	3.7	2.8	3.2	0.83	1.44	**1.09**
3 cuadras	2.5	1.9	2.2	0.82	2.16	**1.34**
4 cuadras	1.9	1.4	1.6	0.82	2.89	**1.54**
5 cuadras	1.5	1.1	1.3	0.81	3.61	**1.71**
6 cuadras	1.2	0.9	1.1	0.81	4.33	**1.87**
7 cuadras	1.1	0.8	0.9	0.80	5.05	**2.01**
8 cuadras	0.9	0.7	0.8	0.87	5.77	**2.24**
9 cuadras	0.8	0.6	0.7	7.85	6.49	**7.14**
10 cuadras	0.7	0.6	0.6	50.55	7.21	**19.10**

b. Escenario: Resto Urbano

Este escenario cuenta con un menor número de personas por cuadra, a fin de reflejar la mayor dispersión entre viviendas, es decir, una menor presencia de edificios, entre otros.

Por otro lado, la probabilidad que la llamada termine en un minuto es 0.4, la misma que se usa en Lima Metropolitana.

Finalmente, el ratio de arribo promedio al TUP es de 0.13. Dicho ratio es mayor al de Lima Metropolitana debido al mayor porcentaje de usuarios que se comunica solo a través de TUP, lo que a su vez determina un mayor número esperado de llamadas por usuario.

Gráfico No. XX:
Simulación y valores óptimos de densidad de TUP – Resto Urbano

No de cuadras	Densidad de telefonía pública			Tiempo de espera (minutos)	Tiempo hasta llegar al TUP	Media geométrica del tiempo
1 cuadra	9.5	7.1	8.3	0.93	0.72	**0.82**
2 cuadras	4.8	3.6	4.2	0.93	1.44	**1.16**
3 cuadras	3.2	2.4	2.8	0.92	2.16	**1.41**
4 cuadras	2.4	1.8	2.1	0.92	2.89	**1.63**
5 cuadras	**1.9**	**1.4**	**1.7**	**0.91**	**3.61**	**1.81**
6 cuadras	1.6	1.2	1.4	1.03	4.33	**2.11**
7 cuadras	1.4	1.0	1.2	17.66	5.05	**9.44**
8 cuadras	1.2	0.9	1.0	98.83	5.77	**23.88**
9 cuadras	1.1	0.8	0.9	236.72	6.49	**39.20**
10 cuadras	1.0	0.7	0.8	415.81	7.21	**54.77**

5. Conclusiones y recomendaciones de política

- El estudio permite determinar el número óptimo de personas deben incluirse en el ámbito de cobertura de cada Teléfono de Uso Público (TUP), basándose en criterios de bienestar de los usuarios expresado el tiempo de espera (y la distancia promedio) que enfrentan los usuarios del TUP.

- El modelo de colas desarrollado con este fin permite estimar el tiempo de espera máximo para utilizar un el servicio TUP (y la distancia recorrida hasta el TUP) para distintos números de cuadras entre terminales TUP en zonas urbanas.

- A partir de los resultados de simulación de este modelo, se infiere la densidad necesaria de TUP (por cada 1000 personas) para obtener una calidad adecuada del servicio. Este resultado puede servir como un sustento para definir políticas sobre la oferta del servicio. Para ello el nivel de medición debe ser el menor posible, idealmente a nivel de centros poblados, pues a niveles más agregados sería necesario considerar también una distribución adecuada de los terminales en relación a la población; es decir que no se supere niveles de distancia promedio adecuados para cada zona.

- En este sentido, la aplicación efectiva de este criterio requiere información detallada sobre la distribución de población y usuarios potenciales del servicio por centro poblado, y debe ser complementada con criterios como promedio y máximas entre servidores TUP.

6. Referencias Bibliográficas

Boffey, B., R. Galvao y L. Espejo (2007) *A review of congestion models in the location of facilities with inmobile servers*. European Journal of Operational Research 178, pag. 643-662.

Gallardo, J., K. López y C. Gonzales (2007) *Perú: Evolución del Acceso, la Cobertura y la Penetración en los Servicios de Telefonía.* Reporte No. 1. Gerencia de Políticas Regulatorias. Organismo Supervisor de la Inversión Privada en Telecomunicaciones.

Winston, W. (2004) *Operational Research: Aplications and Algorithms.* Duxbury Press; 4ta edición.