Herramientas para la toma decisiones: La Teoría de Colas

Categoría: Investigación de Operaciones

Indice

Datos de los autores

Fernando Marrero Delgado. Doctor en Ciencias Técnicas. Máster en Informática Aplicada. Ingeniero Industrial. Profesor Auxiliar. Departamento de Ingeniería Industrial. Universidad Central de Las Villas. Santa Clara. Cuba. Email: fmarrero@fce.uclv.edu.cu

Javier Asencio García. Doctor en Ciencias Técnicas. Ingeniero Industrial. Profesor Titular. Departamento de Ingeniería Industrial. Universidad Central de Las Villas. Santa Clara. Cuba. Email: asencio@fce.uclv.edu.cu

René Abreu Ledón. Máster en Ingeniería Industrial. Ingeniero Industrial. Profesor Asistente Departamento de Ingeniería Industrial. Universidad Central de Las Villas. Santa Clara. Cuba. Email: rabreu@fce.uclv.edu.cu

René Orozco Sánchez. Ingeniero Industrial. Aspirante a Máster del Departamento de Ingeniería Industrial. Universidad Central de Las Villas. Santa Clara. Cuba. Email: fmarrero@fce.uclv.edu.cu

Hugo R. Granela Martín. Doctor en Ciencias Técnicas. Ingeniero Industrial. Profesor Auxiliar. Departamento de Ingeniería Industrial. Universidad Central de Las Villas. Santa Clara. Cuba. Email: hugran@fce.uclv.edu.cu

Palabras claves

Teoría de Colas, Investigación de Operaciones, Optimización

Introducción

Muchas industrias de productos y de servicios tienen un sistema de colas en el que los ¨productos¨ (o clientes) llegan a una ¨estación ¨ esperan en una ¨fila¨ (o cola), obtienen algún ¨servicio¨ y luego salen del sistema. Considere los siguientes ejemplos:

- Los clientes llegan a un banco, esperan en una fila para obtener un servicio de uno

de los cajeros, y después salen del banco.
- Las partes de un proceso de producción llegan a una estación de trabajo particular desde diferentes estaciones, esperan en un compartimiento para ser procesadas por una máquina, y luego son enviadas a otra estación de trabajo.
- Después de hacer sus compras, los clientes eligen una fila en las cajas, esperan a que el cajero les cobre y luego salen de la tienda.
- Las llamadas telefónicas llegan al centro de reservaciones de una aerolínea, esperan al agente de ventas disponible, son atendidas por ese agente y dejan el sistema cuando el cliente cuelga.

Desarrollo

Los problemas administrativos relacionados con tales sistemas de colas se clasifican en dos grupos básicos:

1. **Problemas de análisis**. Usted podría estar interesado en saber si un sistema dado está funcionado satisfactoriamente. Necesita responder una o más de las siguientes preguntas:

a. ¿ Cuál es el tiempo promedio que un cliente tiene que esperar en la fila antes de ser atendido?
b. ¿Qué fracción de tiempo ocupan los servidores en atender a un cliente o en procesar un producto?
c. ¿Cuáles son el número promedio y el máximo de clientes que esperan en la fila

Basándose en estas preguntas, los gerentes tomarán decisiones como emplear o no más gente, agregar una estación de trabajo adicional para mejorar el nivel de servicio, o si es necesario o no aumentar el tamaño del área de espera.

2. **Problemas de diseño.** Usted desea diseñar las características de un sistema que logre un objetivo general. Esto puede implicar el planteamiento de preguntas como las siguientes:

a. ¿Cuántas personas o estaciones deben emplearse para proporcionar un servicio aceptable?
b. ¿Deberán los clientes esperar en una sola fila (como se hace en muchos bancos) o en diferentes filas (como en el caso de los supermercados)?
c. ¿Deberá haber una estación de trabajo separada que maneja las cuestiones ¨especiales¨(como el caso del acceso a primera clase en el mostrador de una aerolínea?
d. ¿Qué tanto espacio se necesita para que los clientes o los productos puedan esperar? Por ejemplo, en un sistema de reservaciones por teléfono, ¿qué tan grande debe ser la capacidad de retención? Esto es, ¿cuántas llamadas telefónicas se deben mantener en espera antes de que las siguientes obtenga la señal de ocupado?

Estas decisiones de diseño se toman mediante la evaluación de los méritos de las diferentes alternativas, respondiendo a las preguntas de análisis del grupo 1 y luego seleccionando la alternativa que cumpla con los objetivos administrativos. En el presente capítulo se proporcionan las técnicas para analizar un sistema de colas dado. Sin embargo, las técnicas matemáticas específicas dependen de la clase de sistema a la cual pertenece su modelo de colas.

Características de un sistema de colas

Para analizar un sistema de colas, es mejor primero identificar las características importantes que aparecen en la siguiente sección características claves, y que se ilustran en la figura 1.1

Características claves

Las siguientes características se aplican a los sistemas de colas:

Una población de clientes, que es el conjunto de los clientes posibles.

Un proceso de llegada, que es la forma en que llegan los clientes de esa población

Un proceso de colas, que está conformado por (a) la manera que los clientes esperan para ser atendidos y (b) la disciplina de colas, que es la forma en que son elegidos para proporcionarles el servicio.

Un proceso de servicios, que es la forma y la rapidez con la que es atendido el cliente

Proceso de salida, que son de los siguientes dos tipos:

a. Los elementos abandonan completamente el sistema después de ser atendidos, lo que tiene como resultado un sistema de colas de un paso. Por ejemplo como se muestra en la figura 1.2 (a) los clientes de un banco esperan en una sola fila, son atendidos por uno de los tres cajeros y, después que son atendidos abandonan el sistema.

b. Los productos, ya que son procesados en una estación de trabajo, son trasladados a alguna otra parte para someterlos a otro tipo de proceso, lo que tiene como resultado una red de colas. Por ejemplo, los productos que se muestran en la figura 1.2 (b) primero son procesadas en la estación de trabajo A y después son enviadas a la estación de trabajo B o C. Los productos terminados en ambas estaciones, B y C, luego son procesados en la estación D, antes de abandonar el sistema.

En el presente capítulo solamente se considerarán sistemas de un paso Se necesitan diferentes análisis matemáticos para cada uno de estos dos tipos de procesos de salida.

La población de clientes

Al tomar en cuenta la base de clientes, la principal preocupación es el tamaño de la población. Para problemas como los de un banco o un supermercado, en donde el número de clientes potenciales es bastante grande (cientos de miles), el tamaño de la población se considera, para fines prácticos, como si fuera infinita.

Al contrario, considere una fábrica que tiene cuatro máquinas, que a menudo se descomponen y requieren servicio de reparación en un taller especializado. En este caso, es de solamente cuatro. El análisis de poblaciones finitas (es decir de tamaño limitado) es más complicado que el análisis en donde la base de población se considera infinita.

El proceso de llegada

El proceso de llegada es la forma en que los clientes llegan a solicitar un servicio. La característica más importante del proceso es el tiempo entre llegadas, que es la cantidad de tiempo entre dos llegadas sucesivas. Este lapso es importante porque mientras menor sea el intervalo de tiempo, con más frecuencia llegan los clientes, lo que aumenta la

demanda de servidores disponibles.

Características claves

Existen dos clases básicas de tiempo entre llegadas:

Determinístico, en el cual clientes sucesivos llegan en un mismo intervalo de tiempo, fijo y conocido. Un ejemplo clásico es el de una línea de ensamble, en donde los artículos llegan a una estación en intervalos invariables de tiempo (conocido como ciclos de tiempo)

Probabilístico, en el cual el tiempo entre llegadas sucesivas es incierto y variable. Los tiempos entre llegadas probabilísticos se describen mediante una distribución de probabilidad.

En el caso probabilístico, la determinación de la distribución real, a menudo, resulta difícil. Sin embargo, una distribución , la distribución exponencial, ha probado ser confiable en muchos de los problemas prácticos. La función de densidad, para una distribución exponencial depende de un parámetro, digamos λ (letra griega lambda), y está dada por:

$$f(t)=(1/\lambda)e^{-\lambda t}$$

en donde λ (lambda) es el número promedio de llegadas en una unidad de tiempo.

Con una cantidad, T, de tiempo usted puede hacer uso de la función de densidad para calcular la probabilidad de que el siguiente cliente llegue dentro de las siguientes T unidades a partir de la llegada anterior, de la manera siguiente:

$$P(\text{tiempo entre llegadas} <= T)=1-e^{-\lambda t}$$

El proceso de colas

Parte del proceso de colas tiene que ver con la forma en que los clientes esperan para ser atendidos. Los clientes pueden esperar en una sola fila, como en un banco, observe la figura 1.3(a) éste es un sistema de colas de una sola línea. Al contrario, los clientes pueden elegir una de varias filas en las que deben esperar para ser atendidos, como en las cajas cobradoras de un supermercado observe la figura 1.3(b), éste es un sistema de colas de líneas múltiples

Otra característica del proceso de colas es el número de espacios de espera en cada fila, es decir, el número de clientes que pueden esperar (o que esperarán) para ser atendidos en cada línea. En algunos casos, como en un banco, ese número es bastante grande y no significa ningún problema práctico, pues para cuestiones de análisis la cantidad de espacio de espera se considera infinita. En contraste, un sistema telefónico puede mantener un número finito de llamadas (es decir limitado) de llamadas, después del cual las llamadas subsecuentes no tienen acceso al sistema. Las condiciones de espacio de espera infinito y finito requieren análisis matemáticos diferentes

Características claves

Otra característica del proceso de colas es la disciplina de cola, es decir la forma en que los clientes que esperan son seleccionados para ser atendidos. A continuación presentamos algunas de las formas más comunes.

Primero en entrar, primero en salir (PEPS). Los clientes son atendidos en el orden en que van llegando a la fila. Los clientes de un banco y de un supermercado son atendidos de esa manera.

Último en entrar, primero en salir (VEPS). El cliente que ha llegado más recientemente es el primero en ser atendido. Un ejemplo de esta disciplina se da en un proceso de producción en el que los productos llegan a una estación de trabajo y son apilados uno encima del otro. El trabajador elige, para su procesamiento, el producto que esta encima de la pila, que fue el último que llegó para ser procesado o para brindarle un servicio.

Selección de prioridad. A cada cliente que llega se le da una prioridad y se le elige según ésta para brindarle el servicio. Un ejemplo de esta disciplina son los pacientes que llegan a la sala de urgencias de un hospital. Mientras más severo sea el caso, mayor será la prioridad del ¨cliente ¨

El proceso de servicio

El proceso de servicio define cómo son atendidos los clientes. En algunos casos, puede existir más de una estación en el sistema en el cual se proporcione el servicio requerido. Los bancos y los supermercados, de nuevo, son buenos ejemplos de lo anterior. Cada ventanilla y cada registradora son estaciones que proporcionan el mismo servicio. A tales estructuras se les conoce como sistemas de colas de canal múltiple. En dichos sistemas, los servidores pueden ser idénticos, en el sentido en que proporcionan la misma clase de servicio con igual rapidez, o pueden no ser idénticos. Por ejemplo, si todos los cajeros de un banco tienen la misma experiencia, pueden considerarse como idénticos. En este capítulo, se tomarán en cuenta servidores idénticos.

Al contrario de un sistema de canal múltiple, considere un proceso de producción con una estación de trabajo que proporciona el servicio requerido. Todos los productos deben pasar por esa estación de trabajo; en este caso se trata de un sistema de colas de canal sencillo. Es importante hacer notar que incluso en un sistema de canal sencillo pueden existir muchos servidores que, juntos, llevan a cabo la tarea necesaria. Por ejemplo, un negocio de lavado a mano de automóviles, que es una sola estación, puede tener dos empleados que trabajan en un auto de manera simultánea

Otra característica del proceso de servicio es el número de clientes atendidos al mismo tiempo en una estación. En los bancos y en los supermercados (sistema de canal sencillo), solamente un cliente es atendido a la vez. Por el contrario, los pasajeros que esperan en una parada de autobús son atendidos en grupo, según la capacidad del autobús que llegue. En este capítulo se verá el servicio de uno a la vez.

Otra característica más de un proceso de servicio es si se permite o no la prioridad, esto es ¿puede un servidor detener el proceso con el cliente que está atendiendo para dar lugar a un cliente que acaba de llegar? Por ejemplo, en una sala de urgencia, la prioridad se presenta cuando un médico, que está atendiendo un caso que no es crítico es llamado a atender un caso más crítico. Cualquiera que sea el proceso de servicio, es necesario tener una idea de cuánto tiempo se requiere para llevar a cabo el servicio. Esta cantidad es importante debido a que cuanto más dure el servicio, más tendrán que esperar los clientes que llegan. Como en el caso del proceso de llegada, este tiempo pude ser determinístico o probabilístico . Con un tiempo de servicio determinístico, cada cliente requiere precisamente de la misma cantidad conocida de tiempo para ser atendido. Con un tiempo de servicio probabilístico, cada cliente requiere una cantidad distinta e incierta de tiempo de servicio. Los tiempos de servicio probabilísticos se describen matemáticamente mediante una distribución de probabilidad. En la práctica resulta difícil determinar cuál es la

distribución real, sin embargo, una distribución que ha resultado confiable en muchas aplicaciones , es la distribución exponencial .En este caso, su función de densidad depende de un parámetro, digamos (la letra griega my) y esta dada por

$s(t)=(1/\mu)e^{-\mu t}$

en la que

μ **=número promedio de clientes atendidos por unidad de tiempo, de modo que**

$1/\mu$ =tiempo promedio invertido en atender a un cliente

En general, el tiempo de servicio puede seguir cualquier distribución, pero, antes de que pueda analizar el sistema, usted necesita identificar dicha distribución.

Clasificaciones de los modelos de cola

Como se menciono al inicio del presente capitulo, para aplicar las técnicas matemáticas apropiadas, usted debe identificar las características de un sistema de colas, basado en la población de clientes y en los procesos de llegada, de colas y de servicio. El método de clasificación presentado aquí pertenece a un sistema de colas en el que el tamaño de la población de clientes es infinita, los clientes que llegan esperan en una sola fila y el espacio de espera en cada línea es efectivamente infinito.

Características claves

En este modelo, los símbolos describen las características del sistema

- EL proceso de llegada. Este símbolo describe la distribución de tiempo entre llegadas, que es uno de los siguientes:

a. D para denotar que el tiempo entre llegadas es determinístico.
b. M para denotar que los tiempos entre llegadas son probabilístico y siguen una distribución exponencial.
c. G para denotar que los tiempos entre llegadas son probabilísticos y siguen una distribución general diferente a la exponencial.

- El proceso de servicio. Este símbolo describe la distribución de tiempos de servicio, que es uno de los siguientes:

a. D para describir un tiempo de servicio determinístico.
b. M para denotar que los tiempos de servicio son probabilísticos y siguen una distribución exponencial
c. G para denotar que los tiempos de servicio son probabilísticos y siguen una distribución diferente a la exponencial.

- El proceso de colas. Este número, c, representa cuántas estaciones o canales paralelos existen en el sistema. (Recuerde que se supone los servidores idénticos en su rapidez de servicio.)

Cuando de espera y/o el tamaño de la población de clientes es finito, los dos siguientes símbolos adicionales se incluyen para indicar estas limitaciones:

- Un número k que representa el número máximo de clientes que pueda estar en el sistema en cualquier momento(es decir, en servicio o en espera en la fila). Este número es igual al de estaciones paralelas más el número total de clientes para ser atendidos.
- Un número L que representa el número total de clientes de la población

Medidas de rendimiento para evaluar un sistema de colas

El objetivo último de la teoría de colas consiste en responder cuestiones administrativas pertenecientes al diseño y a la operación de un sistema de colas. El gerente de un banco puede querer decidir si programa tres o cuatro cajeros durante la hora de almuerzo. En una estructura de producción, el administrador puede desear evaluar el impacto de la compra de una nueva máquina que pueda procesar los productos con más rapidez.

Cualquier sistema de colas pasa por dos fases básicas. Por ejemplo, considere un día como se muestra en la figura 1.4. Cuando el banco abre en la mañana, no hay nadie en el sistema, de modo que el primer cliente es atendido de forma inmediata. Conforme van llegando más clientes, lentamente se va formando la cola y la cantidad de tiempo que tienen que esperar se empieza a aumentar. A medida que avanza el día, el sistema

Llega a una condición en la que el efecto de la falta inicial de clientes ha sido eliminado y el tiempo de espera de cada cliente ha alcanzado niveles bastante estables.

Como se indica en la figura 1.4, la fase inicial, que conserva los efectos de las condiciones iniciales, se conoce como fase transitoria .Después de que los efectos de las condiciones son eliminados, el sistema entra en un estado estable. A pesar de que las preguntas pertenecientes a ambas faces son importantes, esta sección trata solamente el comportamiento del estado estable.

Algunas medidas de rendimiento comunes

Existen muchas medidas de rendimiento diferentes que se utilizan para evaluar un sistema de colas en estado estable, algunas de las cuales se describen en la presente sección. Para diseñar y poner en operación un sistema de colas, por lo general, se los administradores se preocupan por el nivel de servicio que recibe un cliente, así como el uso apropiado de las instalaciones de servicio de la empresa. Algunas de las medidas que se utilizan para evaluar el rendimiento surgen de hacerse las siguientes preguntas:

Preguntas relacionadas con el tiempo, centradas en el cliente, como:

a. ¿Cuál es el tiempo promedio que un cliente recién llegado tiene que esperar en la fila antes de ser atendido?. La medida de rendimiento asociada es el tiempo promedio de espera, representado con Wq
b. ¿Cuál es el tiempo que un cliente invierte en el sistema entero, incluyendo el tiempo de espera y el de servicio?. La medida de rendimiento asociada es el tiempo promedio en el sistema, denotado con W

Preguntas cuantitativas relacionadas al número de cliente, como:

a. En promedio ¿cuántos clientes están esperando en la cola para ser atendidos?. La medida de rendimiento asociada es la longitud media de la cola, representada con Lq
b. ¿Cuál es el número promedio de clientes en el sistema?. La medida de rendimiento asociada es el número medio en el sistema, representado con L

Preguntas probabilísticas que implican tanto a los clientes como a los servidores, por ejemplo:

a. ¿Cuál es la probabilidad de que un cliente tenga que esperar a ser atendido?. La medida de rendimiento asociada es la probabilidad de bloqueo, que se representa por, pw
b. En cualquier tiempo particular, ¿cuál es la probabilidad de que un servidor esté ocupado?. La medida de rendimiento asociada es la utilización, denotada con U. Esta medida indica también la fracción de tiempo que un servidor esta ocupado.
c. ¿Cuál es la probabilidad de que existan n clientes en el sistema?. La medida de rendimiento asociada se obtiene calculando la probabilidad Po de que no haya clientes en el sistema , la probabilidad Pi de que haya un cliente en el sistema, y así sucesivamente. Esto tiene como resultado la distribución de probabilidad de estado, representada por Pn, n=0,1.....
d. Si el espacio de espera es finito, ¿Cuál es la probabilidad de que la cola esté llena y que un cliente que llega no sea atendido?. La medida de rendimiento asociada es la probabilidad de negación del servicio, representada por Pd

Preguntas relacionadas con los costos, como:

a. ¿Cuál es el costo por unidad de tiempo por operar el sistema?
b. ¿Cuántas estaciones de trabajo se necesitan para lograr mayor efectividad en los costos?

El cálculo específico de estas medidas de rendimiento depende de la clase de sistema de colas. Algunas de estas medidas están relacionadas entre sí. Conocer el valor de una medida le permita encontrar el valor de una medida relacionada. Tales relaciones generales se describen primeramente en la sección 1.1.5, antes de que se presenten los métodos utilizados para calcular estas medidas de rendimiento para un sistema de colas dado

Relaciones entre medidas de rendimiento

El cálculo de muchas de las medidas de rendimiento depende de los procesos de llegadas y de servicio del sistema de colas en específico. Recuerde de la sección 1.1, que el caso probabilístico, estos procesos son descritos matemáticamente mediante distribuciones de llegada y de servicio. Incluso sin conocer la distribución especifica, las relaciones entre algunas de las medidas de rendimiento pueden obtenerse para ciertos sistemas de colas, únicamente mediante el uso de los siguientes parámetros de los procesos de llegada y de servicio.

λ **=número promedio de llegadas por unidad de tiempo**

m **= número promedio de clientes atendidos por unidad de tiempo en una sección**

Suponga una población de clientes infinita y una cantidad limitada de espacio de espera en la fila. El tiempo total que un cliente invierte en el sistema es la cantidad de tiempo invertido en la fila más el tiempo durante el cual es atendido:

- **Tiempo promedio tiempo promedio tiempo promedio En el sistema = de espera + de servicio**

El tiempo promedio en el sistema y el tiempo promedio de espera están representados por

las cantidades W y Wq, respectivamente. El tiempo promedio de servicio puede expresarse en términos de parámetros de &. Por ejemplo, si & es cuatro clientes por hora, entonces , en promedio, cada cliente requiere un cuarto de hora para ser atendido. En general, el tiempo de servicio es 1/&, lo cual nos conduce a la siguiente relación :

- **W = Wq + 1/**m

Considere ahora la relación entre el número promedio de clientes en el sistema y el tiempo promedio que cada cliente pasa en el sistema. Imagine que un cliente acaba de llegar y se espera que permanezca en el sistema un promedio de media de hora. Durante esta media hora, otros clientes siguen llegando a una tasa ¿¿¿, digamos doce por hora. Cuando el cliente en cuestión abandona el sistema, después de media hora, deja tras de sí un promedio de (1/2)*12 = 6 clientes nuevos. Es decir, en promedio, existen seis clientes en el sistema en cualquier tiempo dado. En términos de & y de las medidas de rendimiento, entonces:

Numero promedio número promedio tiempo promedio de clientes = de llegadas por * en el sistema

en el sistema unidad de tiempo de modo que:

L = l ***W**

Utilizando una lógica parecida se obtiene la relación entre el número promedio de clientes que esperan en la cola y el tiempo promedio de espera en la fila:

- **Número promedio número promedio tiempo promedio de clientes = de llegadas por * en la cola en la cola unidad de tiempo de manera que:**

  Lq = l *** Wq**

ANÁLISIS ECONÓMICO DE LOS SISTEMAS DE COLAS

En la sección anterior se vio la ventaja de tener más de un servidor, a saber la reducción del tiempo de espera y el número de clientes que esperan para ser atendidos. Claramente, mientras más servidores se tengan, mejor será el servicio a los clientes. Sin embargo, cada servidor implica costos de operación. ¿De que manera evalúa usted este equilibrio entre el nivel de servicio y el costo?

EJEMPLO 1.1: Problema de colas de American Weavers, Inc.

American Weavers, Inc, tiene una fabrica de manufactura en Georgia. La planta tiene un gran numero de maquinas tejedoras que con frecuencia se atascan. Estas maquinas son reparadas basándose en el procedimiento de la primera en entrar, la primera en ser revisada, por uno de los 7 miembros del personal de reparación. Durante varios recorridos, la gerente de producción ha observado que, en promedio, de 10 a 12 maquinas están fuera de operación en cualquier momento debido a que están atascadas. Ella sabe que contratar personal de reparaciones adicional bajaría el número de máquinas sin funcionar, lo cual traería como consecuencia un aumento en la producción, pero no sabe a cuantas personas más debería contratar. Se desea determinar dicho número.

MODELO Y ANALISIS DEL SISTEMA DE COLA ACTUAL.

El primer paso que se debe dar consiste en analizar las condiciones de operación actuales. Se debe reconocer que las maquinas tejedoras conforman un modelo de colas. Los clientes están constituidos por las maquinas que se atascan de vez en cuando. Existe un gran numero de tales maquinas, de modo que se podría suponer razonablemente, que la población de clientes es infinita. Se tienen 7 servidores independientes e idénticos que reparan las maquinas basándose en una estrategia de primera en entrar, primera en darle servicio. Se puede pensar en estas maquinas formando una sola fila en espera de pasar con el siguiente servidor que este disponible.

Para modelar esta operación, el siguiente paso consiste en reunir y analizar los datos correspondientes a los procesos de llegada y de servicio. Se supone lo siguiente:

1. La aparición de maquinas atascadas puede ser aproximada por un proceso de llegada de Poisson con una tasa promedio de 25 por hora.
2. Cada máquina atascada requiere una cantidad aleatoria de tiempo para su reparación, que puede ser aproximada por una distribución exponencial con un tiempo promedio de servicio de 15 minutos, lo cual, para cada servidor, significa una tasa promedio de cuatro maquinas por hora.

Con estas observaciones, el sistema actual puede modelarse como un sistema de colas M / M / 7, con λ = 25 maquinas por hora, μ = 4 maquinas por hora y una población y un área de espera infinita.

TABLA 1:**Medidas de rendimiento obtenidas con ‘ Queuing Analysis ‘ en el WinQSB .**

02-08-2000	Performance Measure	Result
23:12:22		
1	System: M/M/7	From Formula
2	Customer arrival rate (lambda) per hour =	25.0000
3	Service rate per server (mu) per hour =	4.0000
4	Overall system effective arrival rate per hour =	25.0000
5	Overall system effective service rate per hour =	25.0000
6	Overall system utilization =	89.2857 %
7	Average number of customers in the system (L) =	12.0973
8	Average number of customers in the queue (Lq) =	5.8473
9	Average number of customers in the queue for a busy system (Lb) =	8.3333
10	Average time customer spends in the system (W) =	0.4839 hours
11	Average time customer spends in the queue (Wq) =	0.2339 hours
12	Average time customer spends in the queue for a busy system (Wb) =	0.3333 hours
13	The probability that all servers are idle (Po) =	0.1017 %

14	The probability an arriving customer waits (Pw or Pb) =	70.1674 %
15	Average number of customers being balked per hour =	0

Como se puede ver, el gerente de producción había estimado con bastante precisión el hecho de que entre 10 y 12 maquinas están atascadas, en promedio; en cualquier momento. De hecho, ese numero en el informe es de 12.09. La línea 10 del reporte indica que las maquinas atascadas están fuera de operación durante un tiempo promedio de 0.4839 horas, aproximadamente 29 minutos.

Es necesario determinar el numero de reparadores adicionales que se necesitarían contratar. Se conocen las medidas de rendimiento de un total de 7 trabajadores.

¿ De que manera cambian las medidas de rendimiento si se aumenta el personal de reparación ? Las medidas de rendimiento asociadas para un numero entre 7 y 11 reparadores se muestran en la TABLA 2

A medida que aumenta el tamaño del personal de 7 a 11, el numero promedio de maquinas fuera de operación disminuye de aproximadamente 12 a 6.333. Similarmente , la cantidad promedio de tiempo que una maquina esta fuera de operación disminuye de 0.4839 horas (aproximadamente 29 minutos) a 0.2533 horas (aproximadamente 15 minutos).

Ahora se necesita información sobre los costos para determinar cuantos reparadores adicionales, si se requieren, deben contratarse.

TABLA 2: Medidas de rendimiento con diferentes tamaños de personal de reparación.

		Numero	de	reparadores	
	7	8	9	10	11
Utilización (%)	89.2857	78.1250	69.4444	62.5000	56.8182
Numero esperado en la cola	5.8473	1.4936	0.5363	0.2094	0.0830
Numero esperado en el sistema	12.0973	7.7436	6.7863	6.4594	6.3330
Probabilidad de que un cliente tenga que esperar	0.7017	0.4182	0.2360	0.1257	0.0630
Tiempo esperado en cola	0.2339	0.0597	0.0215	0.0084	0.0033
Tiempo esperado en el sistema	0.4839	0.3097	0.2715	0.2584	0.2533

ANALISIS DE COSTOS DEL SISTEMA DE COLAS.

Al analizar los méritos de contratar personal de reparación adicional en American Weavers, Inc., se deben identificar dos componentes importantes:

1. Un costo por hora basado en el tamaño del personal.
2. Costo total de = Costo por hora para * Numero de personal por hora cada reparador reparadores
3. Un costo por hora basado en el numero de maquinas fuera de operación.

Costo total por = Costo por hora para cada * Numero promedio la espera maquina fuera de operación máquina fuera de operación

Para seguir adelante, se necesita ahora conocer el costo por hora de cada miembro del personal de reparación (denotado con Cs) y el costo por hora de una maquina fuera de operación (denotado Ce), que es el costo de una hora de producción perdida. Suponga que el departamento de contabilidad le informa que cada reparador le cuesta a la compañía $ 50 por hora, incluyendo impuestos, prestaciones, etc. El costo de una hora de producción perdida deberá incluir costos explícitos, como la cantidad de ganancias no obtenidas, y costos implícitos, como la perdida de voluntad del cliente no se cumple con la fecha limite de entrega.

Sin embargo, suponga que el departamento de contabilidad estima que la compañía pierde $ 100 por cada hora que una maquina este fuera de operación.

Ahora se puede calcular un costo total para cada uno de los tamaños de personal. Para un personal de 7 reparadores, el numero esperado de maquinas en el sistema es 12. 0973.

Costo total = Costo del personal + Costo de la espera

Costo por hora por * Numero de + Costo por hora por * Número persona reparadores cada maquina fuera esperado de de operación máquinas fuera de operación = (50 * 7) + (100 * 12.0973) = $ 1559.73 por hora.

Realizando cálculos parecidos para cada uno de los tamaños de personal restantes se tiene como resultado los costos por hora de cada alternativa presentada en la TABLA 3.

De los resultados, se puede ver que la alternativa que tiene el menor costo por hora, $ 1128.63, es tener un total de 9 reparadores. En consecuencia, la recomendación a la gerencia de producción, es contratar a dos reparadores adicionales. Estos dos nuevos empleados tendrán un costo de $ 100 por hora, pero este costo adicional esta mas que justificado por los ahorros que se tendrán con menos maquinas fuera de operación. La recomendación reducirá el costo por hora de $1559.73 a $ 1128.63, un ahorro de aproximadamente $ 430 por hora, mayor que la cantidad que cubre sus honorarios.

TABLA 3: **Costo por hora para diferentes tamaños de personal de reparación**.

Tamaño de personal	Numero esperado en el sistema	Costo por hora ($)
7	12.0973	(50 * 7) + (100 * 12.0973) = 1559.73
8	7.7436	(50 * 8) + (100 * 7.7436) = 1174.36
9	6.7863	(50 * 9) + (100 * 6.7863) = 1128.63
10	6.4594	(50 * 10) + (100 * 6.4594) = 1145.94
11	6.3330	(50 * 11) + (100 * 6.3330) = 1183.30

Características claves

En resumen, para evaluar un sistema de colas en el que usted controla el número servidores o su tasa de servicio, se necesitan las siguientes estimaciones de costos y

medidas de rendimiento:

- El costo por servidor por unidad de tiempo (Cs)
- El costo por unidad de tiempo por cliente esperando en el sistema (Ce)
- El número promedio de clientes en el sistema (L)

Modelos que caracterizan los diferentes procesos de servicio.

A continuación se analizarán los diferentes modelos que caracterizan los procesos de servicio.

- **Líneas de espera con salidas y llegadas combinadas.**

Se estudiarán a continuación modelos para líneas de espera que combinan procesos de llegadas y salidas. Solo analizaremos líneas de espera donde los clientes son atendidos por c servidores que ofrecen servicios iguales desde el punto de vista del tiempo que requieren para prestar el mismo.

El objetivo final de analizar situaciones de espera consiste en generar medidas de desempeño para evaluar sistemas reales. Se hace necesario para analizar los sistemas de espera decidir con anterioridad si nos interesa analizar el sistema en condiciones transitorias o de estado estable. Las líneas de espera que combinan salidas y llegadas se inician en condiciones transitorias y llegan gradualmente al estado estable después de haber transcurrido un tiempo lo suficientemente grande, siempre que los parámetros permitan se alcance el estado estable.

Utilizando el modelo general de Poisson, obtenemos la ecuación general de equilibrio:

λ_{n}-1p_{n-1} + μ_{n} + 1p_{n+1} = (λ_{n}+μ_{n})p_{n} , n =1,2,...

Las ecuaciones de equilibrio se resuelven en forma recursiva comenzando con p, y procediendo por inducción para determinar $p_{n.}$ En general, podemos demostrar por inducción que la probabilidad de que el sistema alcance el estado estable es:

p_{n} = [($\lambda_{n-1} * \lambda_{n-2} * \ldots \lambda_{0}$) / ($\mu_{n} *$