

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



**INVESTIGACIÓN DE CASOS DE
OPTIMIZACIÓN PARA SERVICIOS DE
SEGURIDAD Y EMERGENCIAS**

**Trabajo de investigación para la obtención del grado de BACHILLER EN
CIENCIAS CON MENCIÓN EN INGENIERÍA INDUSTRIAL**

AUTOR:

León La Torre, Isaac Eduardo

ASESOR:

Ing. Wilmer Atoche Díaz

Lima, 29 de Octubre del 2020

RESUMEN EJECUTIVO

El presente trabajo tiene como objetivo presentar un conjunto de casos relacionados, principalmente, a las diversas formas de optimización de los servicios de seguridad, así como de los servicios de emergencias. Para un mejor entendimiento del estudio se optó por presentar los aspectos teóricos en el primer capítulo, el cual abarca los conceptos de las herramientas que se emplearon en la investigación. Hay presente un mayor énfasis en el uso de herramientas de investigación de operaciones ya que son una de las más usadas para temas de optimización en lo que respecta a este tipo de servicios; sin embargo, también se presenta una base conceptual de otros métodos empleados.

En el segundo capítulo se muestran los diversos casos de optimización que resolvieron varios problemas encontrados para los servicios mencionados anteriormente en diferentes partes del mundo. Dentro de cada caso se encuentra una descripción del problema que se resolvió, el procedimiento empleado por cada autor para solucionarlo, explicado mediante la base teórica vista en el primer capítulo, y las conclusiones a las que se llegaron una vez se obtuvieron los resultados. En la mayoría de casos se podrá observar que se hace uso de más de una herramienta para la optimización del servicio.

Finalmente, en el último capítulo se presentan las conclusiones encontradas sobre los resultados encontrados en cada caso mostrado a lo largo del estudio y también sobre la variedad de herramientas empleadas en cada problema, principalmente para mostrar mayor énfasis en los diferentes métodos de optimización que se han utilizado en la actualidad.

ÍNDICE GENERAL

CAPÍTULO 1. MARCO TEÓRICO	1
1.1 Programación lineal.....	1
1.2 Teoría de redes	5
1.2.1 Conceptos básicos.....	5
1.2.2 Problema de la ruta más corta.....	7
1.3 Teoría de colas	11
1.3.1 Descripción general de una línea de espera.....	12
1.3.2 Notación.....	14
1.3.3 Modelo hipercúbico de colas.....	15
1.4 Simulación.....	19
1.4.1 Método de Montecarlo.....	20
CAPÍTULO 2. CONTENIDO DE LA INVESTIGACIÓN.....	22
2.1 Determining optimal police patrol areas with maximal covering and backup covering location models	22
2.2 Optimizing police dispatch for incident response in real time.....	24
2.3 The effect of police response time on clearance rates.....	26
2.4 The optimal dispatch of traffic and patrol police service platforms	29
2.5 Service time, dispatch time, and demand for police services: helping more by serving less.....	30
2.6 Analysing the police patrol routing problem: a review.....	31
2.7 Police patrol districting method and simulation evaluation using agent-based model and GIS	33
2.8 Simulation optimization of police patrol districting plans using response surfaces	35
2.9 Covering models and optimization techniques for emergency response facility location and planning: a review	38
2.10 Optimal allocation of police patrol resources using a continuous-time crime model..	39
2.11 Queueing Models of Call Centers: An Introduction	40
2.12 A Queueing-Linear Programming Approach to Scheduling Police Patrol Cars.....	42
2.13 Queueing Model for Information Security Monitoring Systems	48
2.14 Testing the Validity of a Queueing Model of Police Patrol.....	49
2.15 Police officer scheduling using goal programming.....	54
CAPÍTULO 3. CONCLUSIONES	58
Bibliografía	61

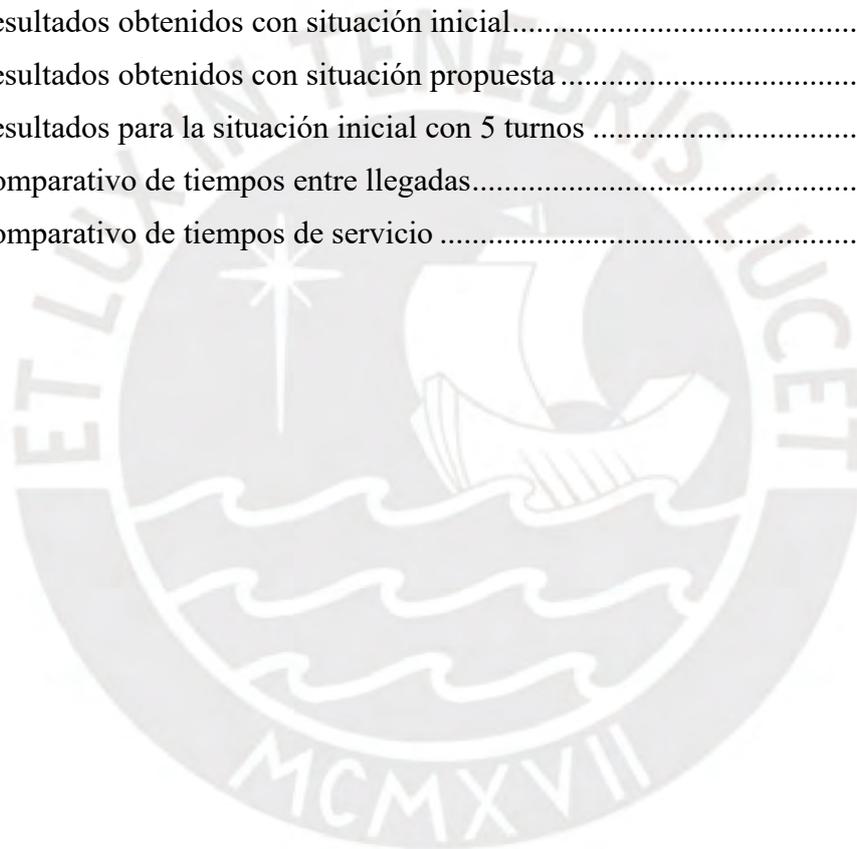
ÍNDICE DE TABLAS

Tabla 1: Relación de restricciones	3
Tabla 2: Puntos de la región factible.....	4
Tabla 3: Relación de paraderos.....	9
Tabla 4: Solución del ejercicio	10
Tabla 5: Parámetros del modelo hipercúbico.....	16
Tabla 6: Resumen estadístico.....	27
Tabla 7: Tiempo de servicio promedio	31
Tabla 8: Economías de escala para evaluación de modelo de <i>call centers</i>	42
Tabla 9: Resultados del experimento de 10 días.....	50
Tabla 10: Parámetros de entrada.....	51
Tabla 11: Comparativo de resultados observados vs calculados	52
Tabla 12: Cuadro de resultados actuales vs propuestos.....	57

ÍNDICE DE FIGURAS

Figura 1: Método gráfico	3
Figura 2: Región Factible.....	4
Figura 3: Red con cinco nodos y arcos	5
Figura 4: Ejemplo de árbol	6
Figura 5: Arco dirigido	6
Figura 6: Suma de distancias de arcos	7
Figura 7: Problema de la ruta más corta	8
Figura 8: Paraderos posibles y sus distancias (en km).....	9
Figura 9: Ruta más corta del ejercicio	11
Figura 10: Modelo hipercúbico de colas para 3 servidores	17
Figura 11: Incidentes en la división norte central de Dallas.....	23
Figura 12: Centros de patrulla óptimos.....	24
Figura 13: Mapa de Leicester	25
Figura 14: Unidades de envío disponibles	25
Figura 15: División sectorial de Manchester	27
Figura 16: Distribución del tiempo de respuesta por distancia.....	28
Figura 17: Volumen de llamadas y tiempo de servicio.....	30

Figura 18: Selección de unidades geográficas atómicas.....	34
Figura 19: Localización histórica de CFS en Charlottesville	34
Figura 20: Modelo de simulación de patrullas policiales	35
Figura 21: Centro óptimo para la atención de incidentes	36
Figura 22: Gráfica de densidad del volumen de llamadas	37
Figura 23: Gráfico Pareto de carga de trabajo vs tiempo de respuesta promedio.....	37
Figura 24: Modelo simple de un <i>call center</i>	41
Figura 25: Cuadro de demanda de ciudadanos vs vehículos disponibles	43
Figura 26: Modelo de optimización propuesto	45
Figura 27: Modelo de colas aplicado al proceso de patrullaje vehicular.....	46
Figura 28: Resultados obtenidos con situación inicial.....	46
Figura 29: Resultados obtenidos con situación propuesta	47
Figura 30: Resultados para la situación inicial con 5 turnos	47
Figura 31: Comparativo de tiempos entre llegadas.....	53
Figura 32: Comparativo de tiempos de servicio	53



CAPÍTULO 1. MARCO TEÓRICO

1.1 Programación lineal

De acuerdo con Luenberger (1984), la programación lineal es, sin duda alguna, el mecanismo más conocido para la formulación de varios tipos de problemas que requieren ser solucionados. Resulta ser un modelo de optimización, cuyo fin es resolver un problema basado en el uso de incógnitas que permitan maximizar o minimizar (dependiendo del caso) el objetivo. El problema se encuentra comprendido por un número de restricciones, las cuales pueden ser igualdades o desigualdades, siempre manteniendo la linealidad en cada restricción. Un modelo de programación lineal está comprendido por los siguientes elementos (Taha, 2017):

1. Variables de decisión que se busca determinar para resolver el problema.
2. Función objetivo, la cual se necesita optimizar (maximizar o minimizar).
3. Restricciones, las cuales debe satisfacer la solución encontrada.
4. No negatividad de las variables de decisión.

Para entender de mejor manera los elementos de un modelo de programación lineal, se procede a mostrar un ejemplo a continuación:

Variables de decisión: se desea saber cuántos productos deben fabricarse en cada línea para cumplir con la demanda planeada.

X_1 : número de productos terminados en la línea 1 de la fábrica.

X_2 : número de productos terminados en la línea 2 de la fábrica.

Función objetivo: se sabe que los productos de la línea 1 brindan un margen de 3 soles por producto terminado y los de la línea 2 un margen de 2 soles por producto terminado.

Maximizar $3X_1 + 2X_2$

Restricciones: se sabe que la línea 1 invierte 3 minutos en fabricar un producto mientras que la línea 2 invierte 4 minutos en ello; al día las dos líneas laboran por 8 horas. Además, teniendo en cuenta que, como mínimo, deben fabricarse 50 productos en la línea 2, se obtiene lo siguiente:

$$3X_1 + 4X_2 \leq 480$$

$$X_2 \geq 50$$

No negatividad: finalmente, colocamos las restricciones de no negatividad para que las variables de decisión se encuentren en el primer cuadrante.

$$X_1 \geq 0$$

$$X_2 \geq 0$$

Hay muchas formas de resolver un problema de programación lineal. El más conocido y sencillo es el método gráfico. Puesto que en los *papers* se utilizarán softwares para la resolución de estos problemas, se brindará el uso del método gráfico solo para este ejemplo. Este método se basa en la materialización de las restricciones por medio de una gráfica bi-dimensional. Sin embargo, esto implica que, como máximo, haya presentes dos variables de decisión, ya que de haber tres o más variables, un gráfico de dos dimensiones no será suficiente para encontrar el punto óptimo. Entonces, a continuación, se procede a mostrar las restricciones de manera gráfica:

Tabla 1: Relación de restricciones

Restricción	Descripción
1	$3X_1 + 4X_2 \leq 480$
2	$X_2 \geq 50$

Fuente: Elaboración propia.

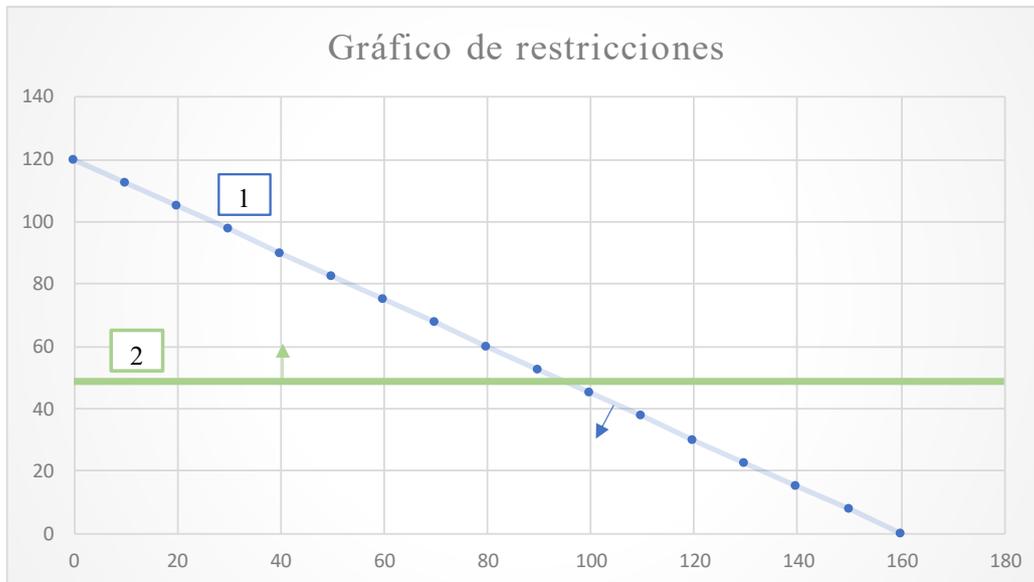


Figura 1: Método gráfico

Fuente: Elaboración propia.

Como podemos observar, el eje Y posee los datos de X_2 , mientras que el eje X muestra los datos de X_1 . Además, se puede notar que hay presente unas flechas en cada recta de las dos restricciones, esta flecha sirve de ayuda para entender qué región del cuadrante satisface las restricciones. Entonces, al intersectar las regiones que satisfacen las restricciones se obtiene lo que se conoce como **región factible**, la cual se observa a continuación:

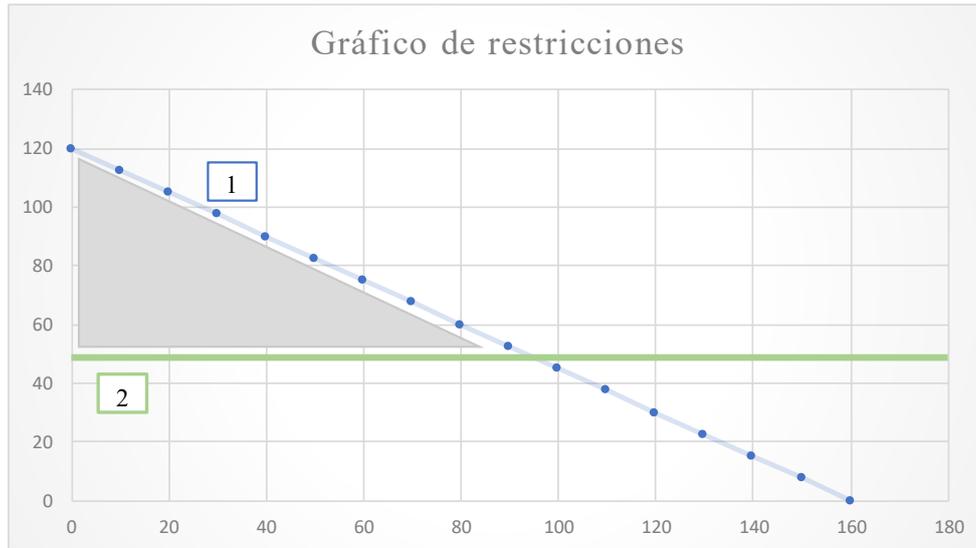


Figura 2: Región Factible
Fuente: Elaboración propia.

Se puede notar que esta región tiene presente tres vértices, los cuales serán considerados los puntos óptimos de la región; sin embargo, habrá un punto que brindará un mayor beneficio con respecto a los otros. Para ello, se procede a hacer el cálculo, el cual puede ser visto en la Tabla 2:

Tabla 2: Puntos de la región factible

X_1	X_2	$3X_1 + 2X_2$
0	120	240
0	50	100
93.33	50	380

Fuente: Elaboración propia.

Se concluye que, cuando en la línea 1 se producen 93.33 productos terminados y en la línea 2 50 productos terminados, se obtendrán 380 soles. Como se mencionó anteriormente, es posible hacer uso de la programación lineal para una cantidad considerable de variables, pero para ello será necesario poseer un software que permita realizar el cálculo necesario. Finalmente, es necesario saber que a medida que se incrementa el número de variables, el número de restricciones deberá ser, por lo menos, igual al número de éstas.

1.2 Teoría de redes

Antes de ahondar en la explicación de los diferentes algoritmos que permiten mejorar la eficiencia de las distintas operaciones que involucran el uso de redes, se procederá a definir los conceptos básicos por los cuales se describe a una red.

1.2.1 Conceptos básicos

Definimos a un **grafo** como a una colección finita de elementos llamados **nodos**, los cuales se encuentran unidos en subconjuntos de pares de nodos por medio de **arcos**. Para representar a un nodo, podemos asignarle un número $(1, 2, \dots, n)$, mientras que a los arcos son representados por el par de nodos que unen (ej. Arco 12). Los nodos son designados por círculos que contienen el número asignado dentro de ellos. A continuación, se muestra un ejemplo de una red que comprende los elementos explicados:

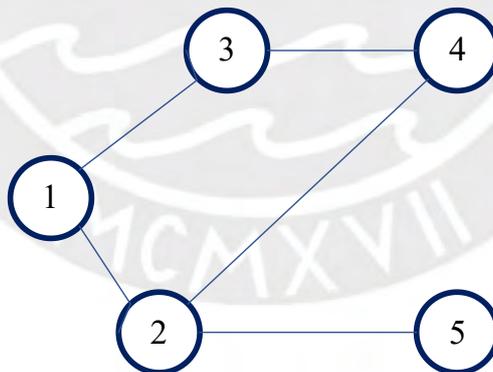


Figura 3: Red con cinco nodos y arcos

Fuente: Luenberger (1984).

A una red se le puede llamar **árbol** cuando están conectados todos los nodos y no hay **ciclos** presentes. Por ejemplo, en la figura 3 se nota que la red está conectada ya que hay arcos en cada par de nodos, pero no es un árbol puesto que, como se observa en el nodo 1, es posible

realizar un recorrido del nodo 1 al nodo 4 de dos maneras diferentes, lo cual implica la generación de un ciclo. Un ejemplo de árbol es el siguiente:

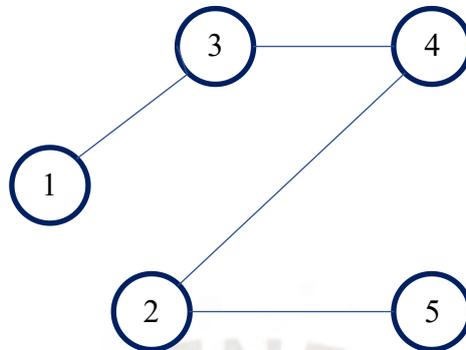


Figura 4: Ejemplo de árbol

Fuente: Elaboración propia.

Asimismo, se considera a un grafo **dirigido** cuando todos los arcos presentes que se encuentran entre los nodos presentan una orientación específica (como el arco 12 de la figura 5). Esta orientación se interpreta como “el nodo 1 es alcanzable a al nodo 2 pero el nodo 2 no es alcanzable al nodo 1”. Mientras haya un arco que no posea orientación, se le considerará al grafo como **no dirigido** y a los arcos que no posean dirección se les llamará **ligaduras**.

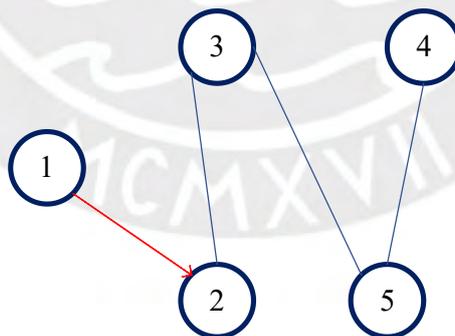


Figura 5: Arco dirigido

Fuente: Elaboración propia

Finalmente, a pesar de que el arco tenga una orientación, resulta necesario asignarle un peso a cada arco, para indicar la relación que puede haber entre los dos nodos unidos. Esta relación

puede indicar una distancia de un nodo a otro, un flujo de caudal que pasa por los dos nodos, una capacidad de corriente eléctrica, etc.

1.2.2 Problema de la ruta más corta

Como hemos podido observar, las redes pueden ser aplicadas a varios casos de la vida real: comunicación de datos, flujo de alguna sustancia líquida, etc. Sin embargo, resulta de interés mantener un enfoque en el estudio de rutas representadas por grafos dirigidos, esto es debido a que, en la mayoría de situaciones, se suele optar por escoger las rutas más rápidas, las más cortas, las más baratas o las más seguras. El problema de la ruta más corta, como su nombre lo dice, indicará, haciendo uso de un algoritmo conocido, qué ruta es la más corta para llegar a un destino objetivo. Para ello, es importante que se interprete a la red como la red de calles que se observa día a día en la ciudad. Según Jungnickel (2013), se debe comenzar definiendo que, a medida que se va recorriendo los arcos, la distancia acumulada es la suma de las distancias que se encuentran en cada arco; se muestra un ejemplo a continuación:



Figura 6: Suma de distancias de arcos

Fuente: Elaboración propia.

En el ejemplo en mención se puede suponer que un ciudadano desea salir de su casa para ir a trabajar. Para llegar a su trabajo a pie, debe caminar 200 metros para dirigirse hacia el parque, el cual se encuentra a 500 metros de su empresa. A partir de esta premisa podemos decir que

la ruta a pie de su casa al trabajo es la suma de los arcos que encontramos entre la casa y el parque, y entre el parque y la empresa (700 metros).

Como se pudo observar en el ejemplo de la Figura 6, hay presentes tres tipos de nodos para el problema de la ruta más corta: el nodo de origen del cual se empieza la movilización (la casa), el nodo de llegada al cual se desea dirigir (la empresa), y el nodo de transbordo, que sirve para llegar al nodo de llegada establecido (el parque). Un ejemplo más complejo se podría visibilizar en las calles que se encuentran en la ciudad de Lima, como vemos a continuación:

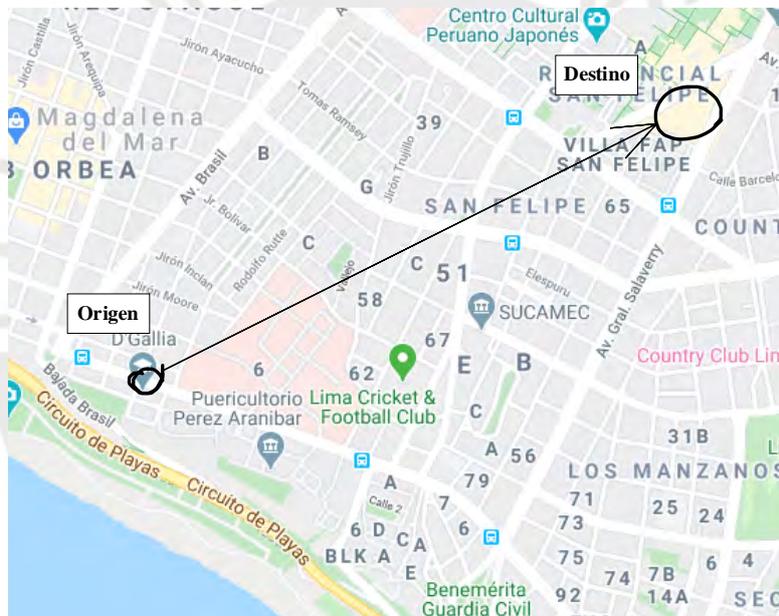


Figura 7: Problema de la ruta más corta

Fuente: Google Maps (2015).

Un estudiante de cocina en D'Gallia desea dirigirse a Real Plaza Salaverry para poder reunirse con un grupo de amigos que viven en el distrito de Jesús María, para ello planea movilizarse por medio del transporte público y así ahorrar algo de dinero. Sin embargo, no sabe qué ruta es la más corta para llegar a su destino; por eso, decide llamar a su hermano, a quien le provee los siguientes paraderos, los cuales podrían servirle de ayuda para hallar la ruta óptima:

Tabla 3: Relación de paraderos

Nodos	Descripción
O	Instituto D'Gallia
A	Paradero Av. El Ejército con Av. Salaverry
B	Paradero Av. El Ejército con Av. Brasil
C	Paradero Av. Sucre con Av. La Marina
D	Paradero Av. Brasil con Av. Javier Prado
E	Real Plaza Salaverry

Fuente: Elaboración propia.

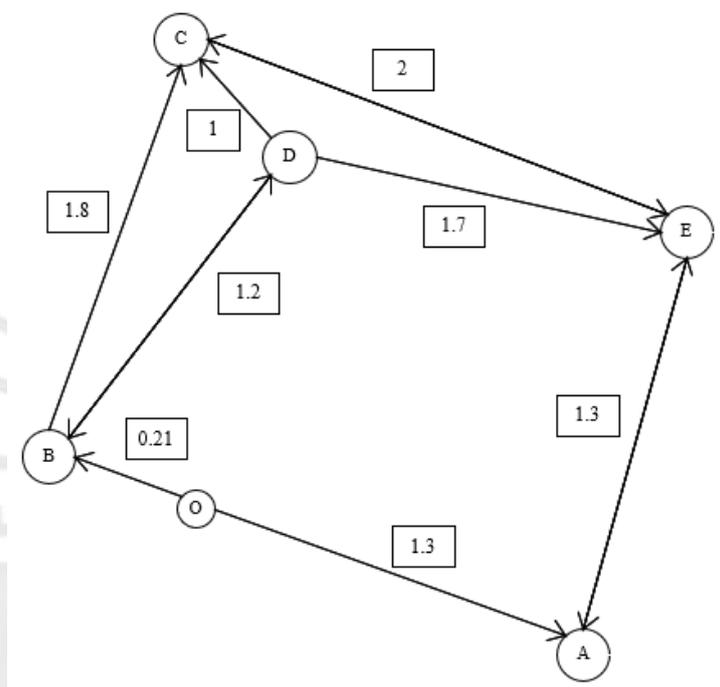


Figura 8: Paraderos posibles y sus distancias (en km)

Fuente: Elaboración propia.

Para solucionar este tipo de casos es requerido buscar un algoritmo que permita encontrar la ruta más corta de manera eficiente, por eso se procederá a usar uno de los algoritmos más conocidos: el **algoritmo de Dijkstra**. El algoritmo de Dijkstra contempla las siguientes definiciones para su uso (Van Steen, 2010):

- Se define a un vértice $a \in V(G)$ donde G es una red no dirigida, y a la serie $S(a)$ de los vértices cuyas rutas más cortas desde a fueron encontradas.

- En cada paso se considera al conjunto de vértices que son adyacentes a algún vértice perteneciente a $S(a)$ que no son parte de $S(a)$.
- Del conjunto de vértices adyacentes, se escoge aquel que es el más cercano a a y finalmente se le añade a la serie $S(a)$.
- Se inicia con un nodo de origen, el cual se define de acuerdo al ejercicio.
- Se etiqueta a los vértices que pertenecen a $S(a)$ con (m,n) donde n es el vértice perteneciente a $S(a)$ que era precedente al vértice que se desea etiquetar, y m la distancia entre el vértice actual y su precedente. La etiqueta del nodo de origen es $(0, -)$ por defecto.
- El algoritmo finaliza cuando todos los nodos pertenecen a $S(a)$.

Tomando en cuenta los pasos anteriormente mencionados, se procede a resolver el ejercicio planteado.

Tabla 4: Solución del ejercicio

Nodo	Etiqueta	Iteración
O	(0,-)	-
A	(1.3, O)	2da. Iteración
B	(0.21, O)	1ra. Iteración
C	(2.01, B)	4ta. Iteración
D	(1.41, B)	3ra. Iteración
E	(2.6, A)	5ta. Iteración

Fuente: Elaboración propia.

A partir de los resultados obtenidos y las iteraciones del algoritmo podemos observar que la ruta más cercana a E (Real Plaza Salaverry) es por medio de A. Teniendo entonces la ruta O-A-E (2.6 kilómetros) como la ruta óptima para llegar al destino planteado.

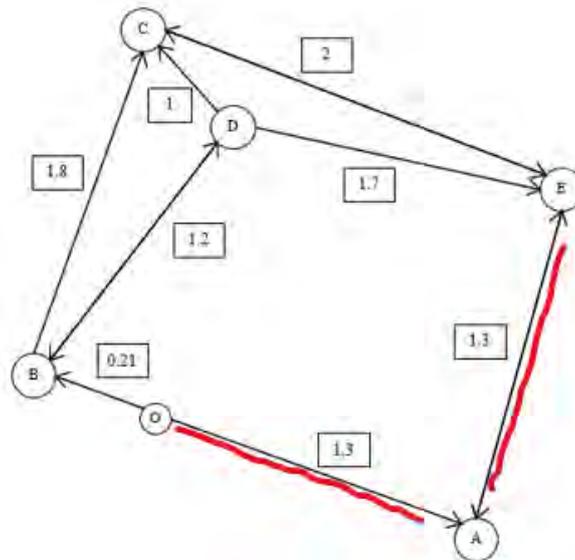


Figura 9: Ruta más corta del ejercicio

Fuente: Elaboración propia.

1.3 Teoría de colas

La teoría de colas, en términos generales, hace alusión al estudio de las líneas de espera que se encuentran presentes en varias ocasiones que se encuentran en la vida real, por ejemplo: las filas de espera presentes en un supermercado, la cola que se presenta cuando se procede a realizar alguna transferencia en algún banco o cuando esperamos nuestro turno para que el doctor de una clínica nos atienda. Para este estudio se hará un enfoque general en las distintas partes que describen a un modelo de colas y finalmente se describirá un caso especial que tendrá cabida en el siguiente capítulo.

Como bien se sabe, en una cola hay presentes ciertos elementos considerados básicos, pero que serán de gran utilidad para conocer el proceso por el cual se analiza a las colas, así como para poder introducir los casos especiales. Antes de describir a estos elementos, se debe definir a un *sistema*, esto es, el espacio físico (banco, hospital, supermercado, fábrica, etc.) o virtual (tiendas virtuales, servicios telefónicos, etc.) en el cual los *servidores* (personal de atención al

cliente, cajeros automáticos, etc.), que laboran en ese espacio, se encargan de atender a los *clientes* (personas en un banco, autos en un *car wash*, etc.).

1.3.1 Descripción general de una línea de espera

Dentro de una línea de espera se encuentran varios elementos, cada uno será definido a continuación:

- La entrada y el patrón de llegada de los clientes: para hacer uso de la teoría de colas ante algún problema referente a una línea de espera es necesario saber cuántos clientes ingresan al sistema de estudio, así como la distribución, ya sea continua o discreta, que presentan los tiempos entre llegadas de los clientes. Usualmente los tiempos entre llegadas suelen aproximarse a una distribución Exponencial con parámetro λ (unidades de llegada por unidad de tiempo); en los diversos modelos de colas se emplea la distribución Exponencial para las llegadas.
- El mecanismo de servicio: este elemento contempla datos referentes al servicio brindado, los cuales son
 - El número de servidores: es necesario saber la cantidad de servidores que hay presentes en la instalación para poder saber si este es uno de los factores por los cuales se forma una cola. Además, es un dato requerido para los modelos de colas.
 - El tiempo de servicio: el tiempo en el cual se brinda el servicio es un dato esencial para aplicar teoría de colas. Este dato suele calcularse en términos del promedio de clientes atendidos por unidad de tiempo para poder hallar el tiempo promedio por cliente atendido.

- Canales de servicio: una instalación puede tener distintos canales de servicio. Es necesario saber cuántos canales hay para poder identificar el problema en cuestión. Por ejemplo: en un banco hay una cola para aquellas personas que quieren utilizar el cajero automático y otra para aquellas que quieren realizar una transferencia en ventanilla.
 - Fases de servicio: además de la cantidad de canales que pueda presentar un servicio en un sistema, también es importante saber por cuántas fases debe pasar el cliente para que finalice su servicio. Esto se puede apreciar en un *car wash*: primero se procede a utilizar *shampoo* para lavado de automóviles y después se realiza el secado en otra plataforma.
- La capacidad del sistema: un dato necesario para los modelos de colas es la capacidad o el aforo que hay presente en una instalación, en algunos casos es posible que haya aforo ilimitado, pero en la mayoría de sistemas de carácter físico habrá una capacidad limitada de personas (u objetos) que podrán estar dentro de éste. Cabe recalcar que para la capacidad de un sistema se considera tanto a la cantidad de servidores como a la cantidad de clientes que hay presentes.
- La disciplina de la cola: se basa en la conducta que tiene la cola al momento de asignar clientes a un determinado servidor. En la gran mayoría de casos, la disciplina es FIFO, esto es, primer cliente que llega es el primer cliente a atender. También hay otras disciplinas como LIFO (*last in, first out*), selección aleatoria, tiempo de proceso más corto, entre otras.

1.3.2 Notación

La notación a emplear para los distintos sistemas de colas que se encuentran en el contenido de la investigación será la de Kendall (1953), la cual se muestra a continuación:

$$(a/b/c : d/e/f)$$

Donde

a: hace alusión a la distribución que siguen los tiempos entre llegadas. Las distribuciones de probabilidad que se aproximan a este dato son

- Exponencial (M)
- Erlang con parámetro k (E_k)
- General (G)
- Determinística (D)

b: esta parte del modelo indica la distribución que sigue el tiempo de servicio. Las distribuciones de probabilidad que se aproximan a esta variable aleatoria son

- Exponencial (M)
- Erlang con parámetro k (E_k)
- General (G)
- Determinística (D)

c: corresponde al número de servidores presentes en el sistema de estudio (banco, supermercado, hospital, etc.).

d: referido a la disciplina de la cola. Como se explicó en el acápite anterior, hay diferentes tipos de disciplina para la recepción del servicio:

- FIFO (primeros que entran, primeros que salen)
- LIFO (últimos que entran, primeros que salen)
- Orden aleatorio (RS)
- Disciplina general (DG)

e: representa la capacidad máxima del sistema. Esta toma en cuenta el aforo máximo de objetos (o personas) que debe haber en la instalación.

f: hace alusión a la población (demanda) que posee el servicio y que en algún momento requerirá ser atendida por este. Esta población también suele ser llamada como *centro emisor* y puede ser de carácter finito o infinito.

1.3.3 Modelo hipercúbico de colas

Según Larson (1974), a pesar de que la mayoría de análisis de problemas de localización o envío de unidades aplicados a servicios de emergencia urbanos han presentado cierta efectividad, hay presentes al menos tres deficiencias. Primero, la mayoría de análisis se han enfocado solamente en las unidades de respuesta *intradistritales* (distrito de estudio), ignorando las respuestas *interdistritales* (distritos aledaños), lo cual puede afectar en los resultados que se obtengan. Segundo, los estudios previos se han enfocado en solo una medida de *performance*, esto es, indicadores relacionados al tiempo de espera de la unidad o tiempo de viaje de ésta; ignorando así otros indicadores de *performance* que caracterizan la efectividad operacional de estos sistemas, y que pueden servir de utilidad para el estudio. Finalmente, los estudios han fallado en incorporar la naturaleza probabilística de un sistema de servicio de emergencia urbana, lo cual se puede ajustar a una distribución de Poisson en el sentido del proceso de llamadas de arribos y en la variabilidad de tiempos de servicio. Para ello, se brinda un modelo capaz de generar una cantidad mayor de indicadores que pueden servir para un mayor análisis en distintos casos de estudio referentes a la atención de algún servicio.

El modelo hipercúbico de colas es un modelo de carácter descriptivo, basado en la distribución espacial de diferentes sistemas de colas con modelo M/M/N, es decir, sistemas en los cuales los tiempos entre llegadas y los tiempos de servicio se ajustan a una distribución exponencial

y, además, hay presente una cantidad N de servidores. Se dice que posee una distribución espacial porque representa a los N servidores de manera individual en un sistema donde los servidores tienen que dirigirse al cliente. Esto permite el acceso a casos de estudio donde haya políticas de envío más complejas. Por otro lado, el modelo requiere del uso de varios parámetros, los cuales son descritos en la tabla siguiente:

Tabla 5: Parámetros del modelo hipercúbico

Notation	Meaning	Measure unit
N_A	Number of atoms	Number
N	Number of servers	Number
T_{ij}	Travel time from atom i to atom j	Minutes
$T_{ik,jh}$	Travel time from subatom ik to subatom jh	Minutes
λ_j	Arrival rate for atom j ($\lambda_j = \sum_k \lambda_{jk}$)	Calls/hour
λ_{jk}	Arrival rate for subatom jk	Calls/hour
λ_k	Arrival rate for priority k	Calls/hour
λ	System total arrival rate	Calls/hour
μ_j	Service rate for server i	Calls/hour
μ	System total service rate	Calls/hour
μ^{-1}	Service time	Minutes

Fuente: Beojone, Caio Vitor (2017)

Como se puede observar, la gran mayoría de parámetros corresponden a un modelo de colas genérico, a excepción del término átomo correspondiente a los parámetros N_A , T_{ij} , $T_{ik,jh}$, entre otros. Se describe a los átomos como la cantidad de particiones que se le emplea a una zona geográfica. Entonces, por ejemplo, si la zona de estudio fuese Lima se podría considerar como un átomo a Magdalena del Mar y otro a San Juan de Lurigancho. Cada átomo puede tener la forma de cualquier figura geométrica y es recomendable que se haga una repartición sustentada para que no afecte al análisis.

Asimismo, en un modelo hipercúbico, los espacios de estado indican la disponibilidad de cada servidor individualmente. Por ejemplo, de tener un sistema con 2 servidores, habrá 2^2 espacios de estado posibles en el sistema: (0,0), (1,0), (0,1), (1,1) donde 0 significa que se encuentre libre y 1 que el servidor se encuentra ocupado. Un cubo podría representar un sistema para 3

servidores, y para un número mayor de servidores se tendrá un hipercubo. En la siguiente figura se puede observar el caso para 3 servidores:

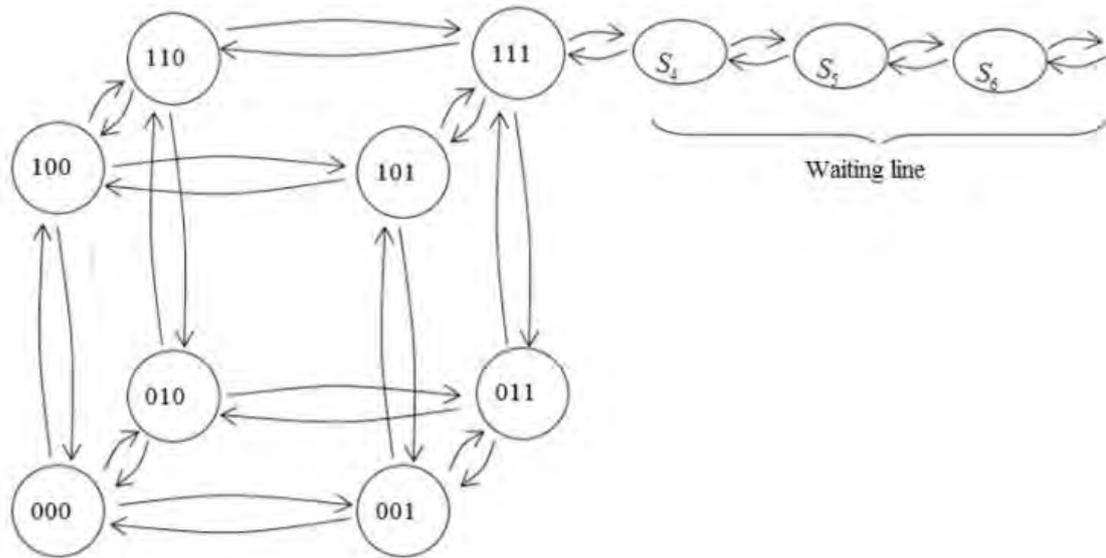


Figura 10: Modelo hipercúbico de colas para 3 servidores

Fuente: Beojone, Caio Vitor (2017)

Por otro lado, según Larson y Odoni (1981) se tienen que tomar las siguientes asunciones al momento de la preparación del modelo:

1. Átomos geográficos: el área en la cual el sistema provee el servicio puede ser dividida en un número N_A de áreas de reporte estadísticas o, como se mencionaba anteriormente, átomos. Estos átomos pueden corresponder a, por ejemplo, bloques de censo, bloques de la ciudad o áreas de protección policial (áreas que protegen las comisarías).
2. Arribos Poisson independientes: las peticiones (llamadas, alertas de emergencia) hechas por los clientes son generadas como un proceso de Poisson, independientemente de cada uno de los átomos. La tasa de arribos Poisson se puede observar en la tabla 5, esta puede ser conocida o estimada.

3. Tiempos de viaje: la data que se posea puede permitir el cálculo de la estimación del tiempo de viaje promedio de cada átomo i a otro átomo j . El símbolo de este indicador se encuentra en la tabla 5 y su símbolo es el de τ_{ij} .
4. Servidores: como se explicó anteriormente, hay N servidores espacialmente distribuidos y cada uno de estos puede transportarse a cualquiera de los átomos geográficos de la región definida de servicio.
5. Localización de los servidores: la localización de cada unidad de respuesta, mientras no se encuentre ocupada realizando el servicio, será conocida (al menos estadísticamente). Por ejemplo, un vehículo de patrulla policial puede encontrarse el 50% de su horario en el átomo 3 y un 25% de su horario en los átomos 1 y 2. Otro caso es el que ocurre con las ambulancias o unidades de delivery, ya que se encuentran específicamente en un átomo cuando no están de servicio. En el caso de las patrullas policiales, al conjunto de átomos por los cuales la unidad de servicio pasa durante su horario se les conoce como área de patrulla, también se les puede llamar sectores o rutas.
6. Asignación de servidores: asumiendo que al menos una unidad de servicio se encuentre disponible entre la región, solamente una unidad será enviada a la zona de requerimiento. Si ninguna unidad se encuentra disponible, el requerimiento entrará en cola junto a otros requerimientos, o será servido por alguna unidad de apoyo (por ejemplo, un vehículo patrulla que asiste a una ambulancia en caso se requiera más ayuda), o simplemente se retirará el requerimiento hecho por el cliente.
7. Preferencia de envío: si un cliente necesita que se acerque una patrulla policial, se preferirá que lo atienda una unidad proveniente del átomo más cercano a que lo atienda una unidad de cualquier otro átomo.

8. Tiempos de servicio: en general, cada unidad de respuesta puede poseer su propio tiempo de servicio promedio. Este tiempo contiene tanto el tiempo de viaje como el tiempo que se toma el servidor en brindar soporte al cliente. Al hacer uso de un tiempo de servicio promedio, no hay presente un análisis de la variabilidad de la media, para estos casos suele asumirse que la desviación estándar es aproximadamente igual a la media. El análisis matemático indica que los tiempos de servicio, en su mayoría, siguen una distribución Exponencial.
9. Dependencia del tiempo de servicio en el tiempo de viaje: las variaciones en el tiempo de servicio que se encuentran explicadas por las variaciones en el tiempo de viaje suelen pasar a un segundo plano comparadas a las variaciones del tiempo de soporte al cliente o derivados. Sin embargo, esto no significa que haya un menor interés en el tiempo de viaje promedio para el cálculo del tiempo de servicio promedio, ya que el cálculo del tiempo de servicio involucra el tiempo de viaje empleado por el servidor para llegar al usuario.

1.4 Simulación

La simulación, según Hillier (1981), es una técnica que sirve para imitar la operación de un proceso o sistema completo por medio de una computadora. Esto es debido a que las computadoras poseen una velocidad de procesamiento tal que se pueden simular años de operación inclusive. Sin embargo, antes de hacer uso de la simulación es requerido haber desarrollado lo siguiente:

- Implementación de un análisis teórico preliminar, aplicando modelos matemáticos.
- Crear un diseño básico del sistema, incluyendo los procedimientos de operación.

Luego de ello, es posible aplicar la simulación para experimentar con los diseños respectivos y estimar la eficiencia real del proceso. A partir de estos resultados, se emplea el sistema real para realizar los últimos detalles del diseño final. Los pasos para la generación de un modelo de simulación son los siguientes:

1. Definir el estado del sistema.
2. Identificar los posibles estados.
3. Describir los eventos probables.
4. Hacer uso de un reloj de simulación, que registrará el paso del tiempo.
5. Buscar un método para generar los registros de manera aleatoria.
6. Encontrar una fórmula que permita hallar las transiciones de los estados que generan los eventos.

1.4.1 Método de Montecarlo

La simulación de Montecarlo, o también llamado método de pruebas estadísticas, analiza el resultado de un modelo un número de veces considerable para poder proporcionar una distribución estadística de los resultados calculados. Es imprescindible que, para usar una simulación de Montecarlo, se deben tener tres parámetros (probable, optimista y pesimista) y los valores de probabilidad que representan cada uno de ellos. De acuerdo a Sobol (1983), el método de Montecarlo presenta dos puntos importantes:

- Su algoritmo presenta una estructura básica: primero se elabora un programa para la generación de las pruebas aleatorias. Después, repetir el experimento un número determinado de veces, de tal forma que cada prueba sea independiente de las anteriores. Finalmente, se toma en cuenta la media de los resultados obtenidos de todas las pruebas.

- El error resulta ser, por defecto, proporcional a $\sqrt{D/N}$, donde D es una constante y N el número de experimentos realizados. Con este dato, se puede observar que el error puede disminuir en gran porcentaje a medida que se aumenta el volumen de pruebas a realizar.

Sin embargo, cabe recalcar que no es sencillo llegar a un nivel de exactitud muy alto aplicando este método. Por ello, el método de Montecarlo es eficaz para la resolución de problemas que impliquen conocer el resultado con una exactitud relativamente baja.



CAPÍTULO 2. CONTENIDO DE LA INVESTIGACIÓN

2.1 Determining optimal police patrol areas with maximal covering and backup covering location models

La mejora de áreas de patrulla policiales con modelos que permitan establecer una cobertura máxima (Curtin, 2007), que se basa en la aplicación del algoritmo del problema MCLP (*Maximal Covering Location Problem*), cuya variación es el modelo PPAC (*Police Patrol Area Covering*). Además, se emplean sistemas de información geográfica (GIS) y la programación lineal para generar diversas soluciones óptimas. Para ello, se describe que la división de las fuerzas policiales en un área establecida representa, en efecto, un problema geográfico. Esto es debido a que utilizando herramientas de investigación de operaciones es posible administrar de manera más eficiente el envío de oficiales e inclusive diseñar el área de patrullaje. Sin embargo, se considera a los GIS una herramienta más importante, ya que se pueden determinar las zonas con mayor actividad criminal. El objetivo de los modelos MCLP y PPAC es buscar la solución al problema de localizar instalaciones que maximicen la cobertura de la demanda de servicios por medio de una distancia aceptable (o tiempo de respuesta óptimo). Para ello se toma en cuenta la información de incidentes presentes en el estado de Dallas (Estados Unidos), específicamente, en la zona norte central, que se visualizará a continuación:

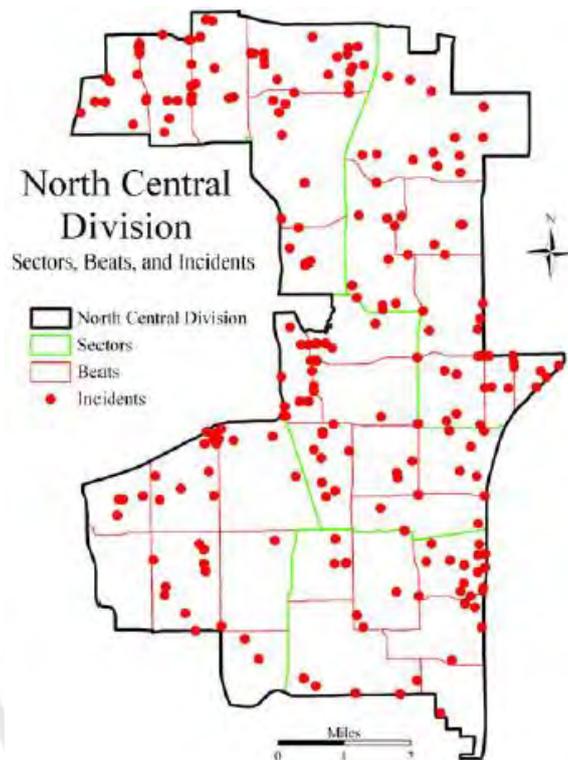


Figura 11: Incidentes en la división norte central de Dallas

Fuente: Curtin (2007).

El estudio concluye en que, utilizando los modelos en mención, se puede reducir la distancia entre las instalaciones propuestas y el lugar de un incidente aleatorio, minimizando, por ende, el tiempo de respuesta. Además, también se contempla que, a pesar de que cada estación policial emplea distintos métodos para repartir geográficamente a sus propias patrullas, hay una variedad de opciones para optimizar de manera generalizada la repartición de oficiales y patrullas en una zona geográfica por medio de distintos modelos que involucren el uso de programación lineal o heurísticas. Finalmente, se muestra la repartición óptima de las instalaciones propuestas a continuación:

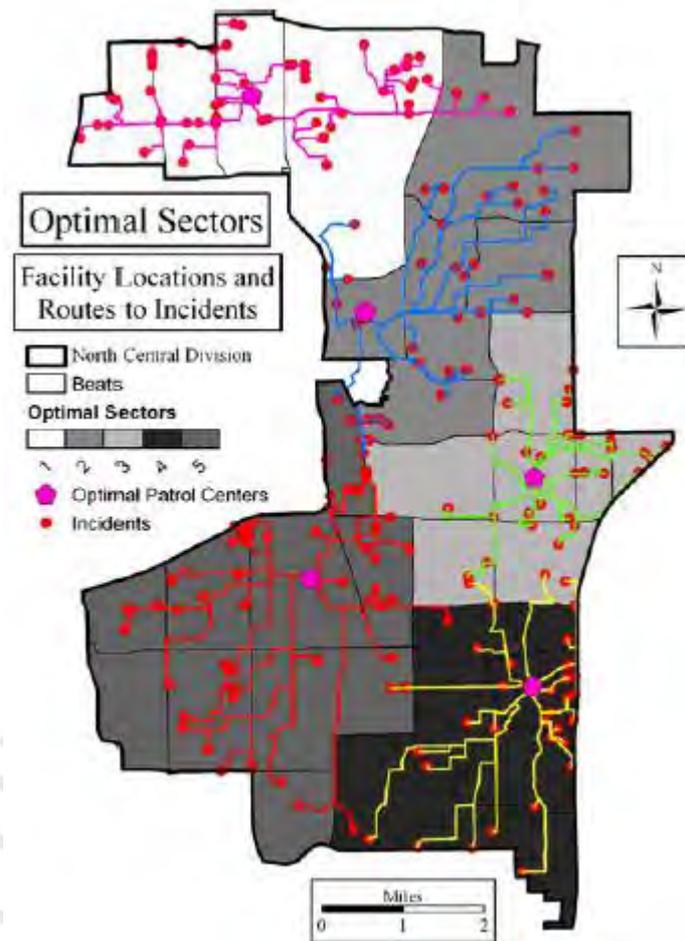


Figura 12: Centros de patrulla óptimos

Fuente: Curtin (2007).

2.2 Optimizing police dispatch for incident response in real time

La propuesta de optimización del envío de policías para la respuesta a una emergencia en tiempo real (Dunnett, 2018) emplea la combinación de grafos dirigidos y el algoritmo de Dijkstra para obtener la ruta más rápida en vez de la ruta más corta. Para llegar a obtener resultados más precisos se empleó la simulación como herramienta a la cual se le brinda la duración del tiempo empleado en un incidente, que es determinado por la distribución empleada, y el tipo de accidente, que se encuentra compuesto por un proceso de decisión propuesto. Los resultados recogidos por la simulación son los tiempos de respuesta, el tiempo total que las unidades están disponibles y la distancia total recorrida para llegar a los incidentes.

El estudio se lleva a cabo en la ciudad de Leicestershire, en el Reino Unido, la cual se puede observar en la Figura 12. Además, también se muestran las cuatro instalaciones disponibles en la ciudad (A, B, C, D).

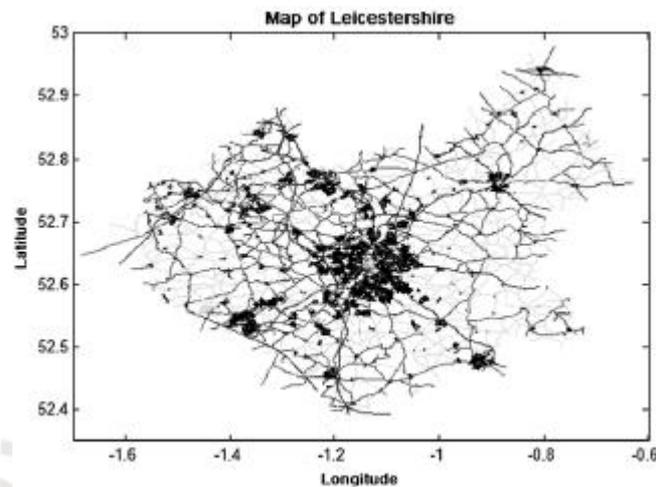


Figura 13: Mapa de Leicester

Fuente: Dunnett (2018).

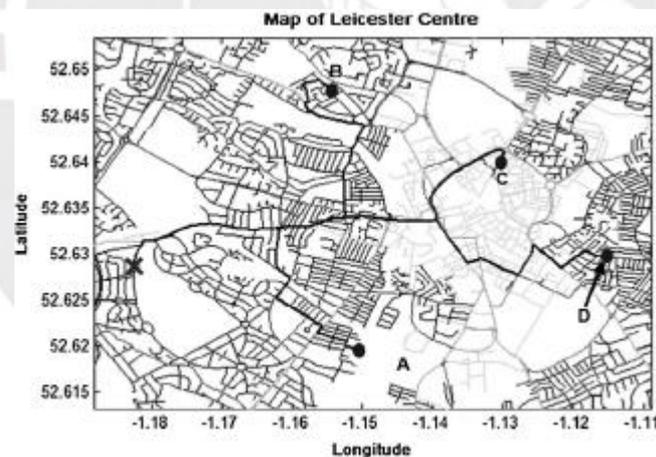


Figura 14: Unidades de envío disponibles

Fuente: Dunnett (2018).

A partir de los resultados, se concluyó en que el proceso de decisión que fue planteado en el estudio conlleva a una eficiente selección de oficiales, además de una reducción del tiempo de respuesta y un incremento de la disponibilidad de unidades policiales. Sin embargo, estos resultados varían de acuerdo a factores como el número de unidades disponibles y el tiempo

empleado en un incidente, que dependen de acuerdo a las fuerzas policiales que se van a escoger como elemento de estudio.

2.3 The effect of police response time on clearance rates

En este estudio, elaborado por Blanes (2018), se hace uso de una base de datos del 2008 que muestra las discontinuidades en los tiempos de respuesta de los efectivos policiales para incidentes que sucedieron en los límites territoriales de cada división policial; además, la base de datos también brinda características del crimen cometido y si este fue resuelto o no. El lugar donde se lleva a cabo el estudio es Manchester, Reino Unido. Para el estudio en mención se asume que cuando se recibe una llamada para que se efectúe el servicio, el oficial que responde se encuentra en la estación correspondiente, mas no en un punto aleatorio asociado a la ruta de patrulla que tiene que seguir. La Gran Policía de Manchester (GMP por sus siglas en inglés) posee alrededor de 6200 efectivos policiales que sirven a una población de 2.6 millones de habitantes aproximadamente; para ello, la GMP divide el territorio en 11 sectores, los cuales coinciden tanto en número como en alcance territorial a las municipalidades presentes en Manchester.



Figura 15: División sectorial de Manchester

Fuente: Blanes (2018).

A partir de la información brindada se pudo agrupar los datos para obtener un resumen estadístico de los campos de importancia del estudio, entre los que destacan los siguientes: tiempo de respuesta, distancia del punto del crimen a la estación más cercana, porcentaje de crímenes violentos o robos. Este resumen se puede visualizar a continuación:

Tabla 6: Resumen estadístico

	Mean	Median	SD	Min	Max
Cleared	0.38	0	0.48	0	1
Time to clearance (days)	34.47	4	113.44	0	2,218
Immediate arrest	0.12	0	0.32	0	1
Suspect named	0.24	0	0.43	0	1
Response time (min)	73.3	16.82	315.01	0.07	21,613.58
Distance to station (km)	3.25	2.26	3.68	0	19
Grade 1	0.31	0	0.46	0	1
Grade 2	0.69	1	0.46	0	1
Violent crimes	0.24	0	0.43	0	1
Theft crimes	0.53	1	0.5	0	1
Other crimes	0.23	0	0.42	0	1

Fuente: Blanes (2018).

Además, haciendo un enfoque más preciso, se pudo distribuir el tiempo de respuesta de acuerdo a la distancia a la cual se encontraba el crimen desde la estación más cercana al punto del

crimen, confirmando así la hipótesis de que a mayor sea la distancia, más grande será la probabilidad de que el tiempo de respuesta sea mayor. Para ello, también se muestra la densidad del tiempo de respuesta, tal y como se puede admirar en la siguiente figura:

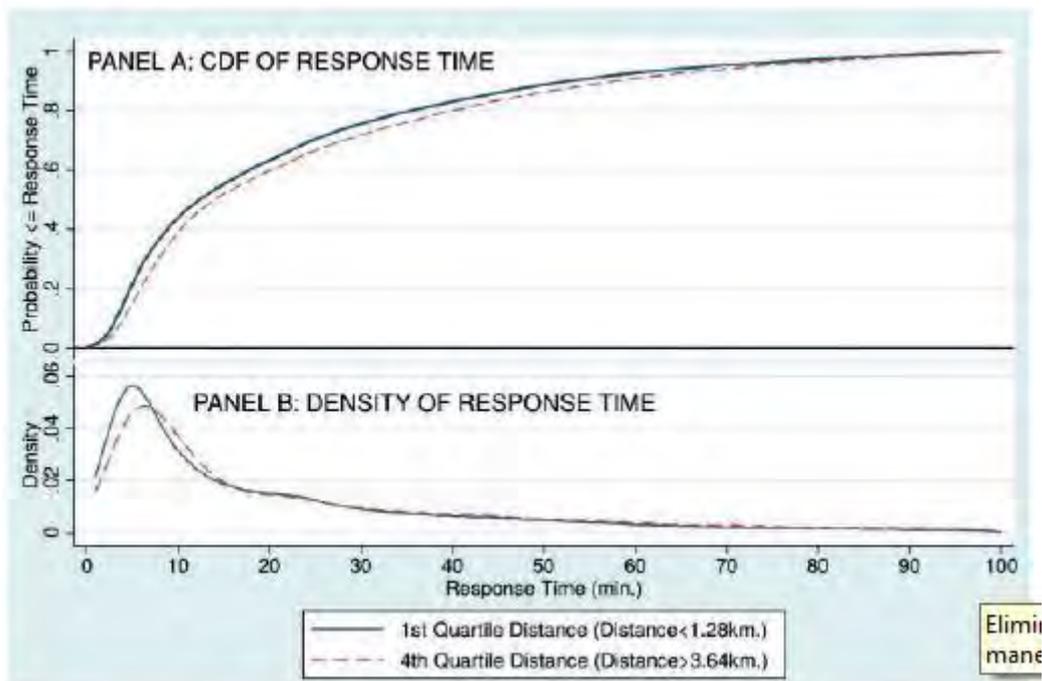


Figura 16: Distribución del tiempo de respuesta por distancia

Fuente: Blanes (2018).

Finalmente, a partir de un análisis de datos más robusto del resumen estadístico, hay evidencia de que hay un efecto del tiempo de respuesta brindado por los efectivos policiales con respecto a los índices de despeje criminalísticos que suceden en Manchester. Hay presentes, además, mejores resultados en los crímenes relacionados a hurto o robo que en los relacionados a la violencia.

2.4 The optimal dispatch of traffic and patrol police service platforms

Este artículo de investigación sostiene que las plataformas de servicio presentes en China que involucran a las patrullas policiales y al tránsito no pueden ejercer un correcto envío de unidades de seguridad debido a tres factores principales (Quan, 2012):

- La congestión vehicular presente en la ciudad
- La localización de los recursos que posee la policía
- La falta de precisión con respecto al flujo vehicular diario

Para ello, se aplica el algoritmo de Floyd, el cual sirve para resolver el problema de la ruta más corta, problema ya abordado en el marco teórico. Asimismo, se hace uso de sistemas de información geográfica (GIS) y de la obtención de imágenes en tiempo real que muestran las cámaras de video para poder comprobar que, por medio del algoritmo de Floyd, se puede hallar fácilmente la ruta más corta. Por otro lado, se hace uso de programación lineal y del problema del *Travelling Salesman Problem* para poder generar modelos con la presencia de congestión vehicular, así como sin la presencia de éste. Ambos modelos son de programación lineal entera y también presentan en su función objetiva la función *minmax*.

Se concluye en que, luego de haber usado la información de varios proyectos y utilizado el software Matlab para la obtención de resultados para distintos tiempos de respuesta, la congestión vehicular encontrada entre los nodos de una red de tráfico urbano influiría en el envío óptimo de patrullas policiales en cierto grado cuando hay presente un incidente.

2.5 Service time, dispatch time, and demand for police services: helping more by serving less

En este *paper* se busca administrar de manera más eficiente las operaciones relacionadas a los patrullajes realizados por los efectivos policiales de la ciudad de San Francisco, Estados Unidos. Para ello se realiza un análisis enfocado tanto en el volumen de llamadas recibidas por las patrullas como en el tiempo de respuesta y el tiempo de servicio que son brindados. Se detalla, además, que, al haber una carga de trabajo muy alta para estos efectivos, suelen rehusarse a recibir algunas llamadas debido al cansancio acumulado por estos, y proceden a trasladar la llamada a otra unidad que es muy probable se encuentre más lejos de la localización del incidente. A continuación, se observa la cantidad de llamadas presentes promedio por hora diaria en la ciudad, así como la duración del tiempo de servicio promedio:

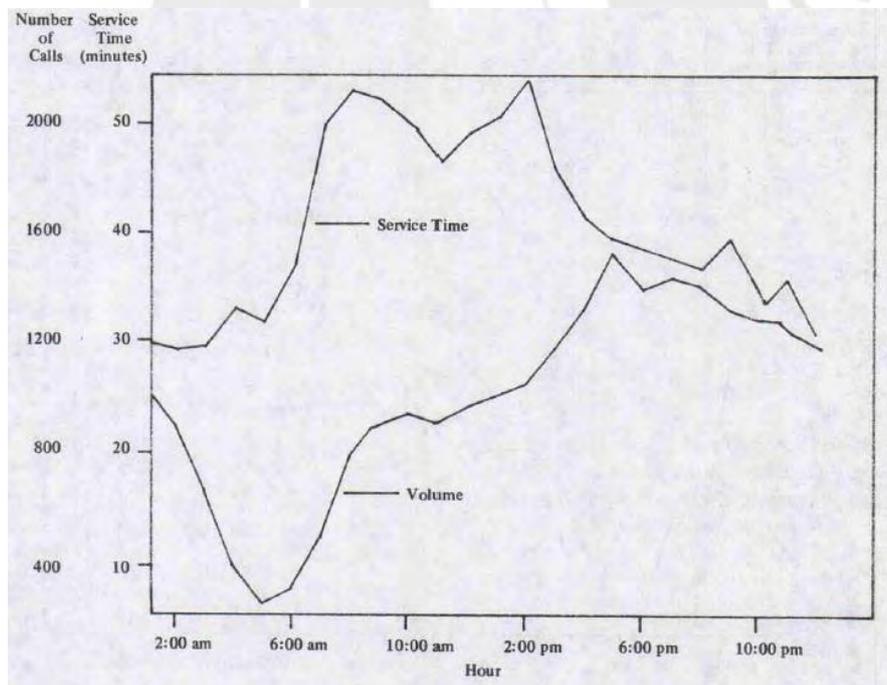


Figura 17: Volumen de llamadas y tiempo de servicio

Fuente: Maxfield (1982).

Además, también se aprecia de manera más detallada el tiempo de servicio promedio para cada tipo de incidente ocurrido en la ciudad de San Francisco:

Tabla 7: Tiempo de servicio promedio

	Mean Service Time	Standard Deviation	Coefficient of Variation
Alarm	17.1	21.9	1.28
Robbery	43.1	44.1	1.02
Assault	47.9	49.7	1.04
Disturbance	26.1	29.0	1.11
Family Fight	33.7	35.5	1.05
Burglary	51.5	38.7	0.75
Person in Need	39.4	39.8	1.01
Vehicle Theft	46.4	35.3	0.76
Meet Complainant	44.4	38.1	0.86
Suspicious Circumstances	27.9	33.3	1.19

Fuente: Maxfield (1982).

Se concluye en que la generación de las colas durante el envío de patrullas se da debido a la alta variabilidad que presenta la demanda, esto es, la cantidad de llamadas recibidas durante el día, además de la falta de disponibilidad de unidades para atender las llamadas. Y aunque un exceso de demanda para el servicio es esperado, las largas demoras ocurridas pueden significar la pérdida de vidas dependiendo del tipo de llamada.

2.6 Analysing the police patrol routing problem: a review

El artículo en mención tiene el objetivo de encontrar un método de solución que pueda ser usado para desarrollar una estrategia de ruteo de patrullaje policial en la ciudad de Kansas que pueda cumplir con ciertos objetivos que resultan ser importantes. Estos objetivos son los siguientes:

- Cumplir con el indicador costo-eficiencia utilizando solo los recursos disponibles que poseen las unidades policiales, ya que estos recursos son financiados con fondos públicos.
- Asegurar el monitoreo de la eficiencia producida por las patrullas en su labor de reducir el crimen y el desorden presente en la ciudad.

- Se debe hacer énfasis en la óptima salida espacio-temporal, con esto se refiere a que se debe elaborar una estrategia que permita administrar bien el espacio y tiempo en el cual que se encuentran las patrullas policiales.

Tomando en cuenta esos objetivos, se planea usar los modelos generados por el problema del ruteo vehicular (VRP) y el problema del viajante (TSP). El VRP determina cómo se puede servir a una serie de consumidores que se encuentran dispersos geográficamente, empleando las rutas que deben seguir los vehículos que salen de un depósito para regresar al mismo una vez que se ha realizado la entrega correspondiente. El VRP optimiza las rutas que toman estos vehículos para poder reducir la distancia total recorrida y, por ende, varios costos asociados a ella. También se presenta al DVRP como una variación del VRP con la diferencia que el DVRP resulta ser un modelo dinámico en el cual se toma en cuenta el problema del ruteo de manera tal que se puedan observar resultados en tiempo real. Esto facilita el ingreso de nuevos clientes que requieren ser atendidos a la brevedad.

A lo largo del estudio se aplican distintas herramientas para resolver el DVRP y observar la variación en los resultados encontrados. Sin embargo, se concluye en que el ruteo de las patrullas policiales formulado como un problema de ruteo vehicular dinámico virtualmente no brinda resultados que se encuentren cercanos a lo que se obtiene realmente. Esto es debido a que, a diferencia de estudios que obtienen resultados paralelos a los reales para temas asociados a servicios médicos de emergencia o brigadas contra incendios, el ruteo de patrullas posee sus propias especificaciones. Si se adentra en el detalle, se observa que hay una alta concordancia en cómo se reacciona a la emergencia y se logran reducir los tiempos de espera del ciudadano, pero no se logra ejercer un patrullaje preventivo que pueda anticipar de manera más eficiente

futuros crímenes; siendo esta, también, la razón por la cual las operaciones policiales difieren de las operaciones que aplican los bomberos o las ambulancias.

2.7 Police patrol districting method and simulation evaluation using agent-based model and GIS

Para este artículo se emplea la simulación para medir el *performance* de los distritos de la ciudad Charlottesville usando dos indicadores: el tiempo de respuesta promedio y la carga de trabajo asociada a los efectivos policiales. La simulación que se usa revela complejidades importantes para la decisión de los planes de repartición. La generación de estos planes de repartición se basa en la utilización de unidades geográficas atómicas, como los bloques de censo que hay repartidos en toda la ciudad o los efectivos que se encuentran transitando alrededor de estos bloques, reagrupando estas unidades de tal manera que se vuelvan a generar nuevos distritos. La idea principal para la implementación de un algoritmo de repartición es similar al algoritmo CPSC (*Constraint-Based Polygonal Spatial Clustering*). Para el algoritmo CPSC se agrupan las unidades geográficas atómicas de forma poligonal y equidistante una de otra. Este polígono se adjunta alrededor de una circunferencia y se va agrupando así a un conjunto de polígonos hasta que se puedan cubrir todos los distritos de la ciudad de estudio. Esto se puede entender de mejor manera con la siguiente figura, la cual muestra las unidades geográficas atómicas repartidas en los distritos de Charlottesville.

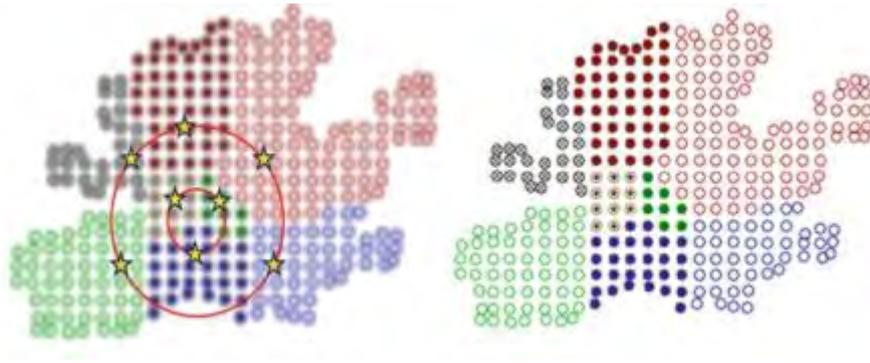


Figura 18: Selección de unidades geográficas atómicas

Fuente: Zhang (2013).

Cabe mencionar que Charlottesville es una ciudad con un diámetro de 11.2 kilómetros aproximadamente y una población de 40000 personas; hay presentes 8 distritos y desde hace 20 años una patrulla se encarga de vigilar un distrito, por lo que hay 8 patrullas en total. Entonces, con la información brindada de los registros de CFS (llamadas de servicio) se pudo conseguir la cantidad histórica de incidentes (figura 20, puntos rojo oscuro) y también, por medio de la simulación, la localización óptima en la cual se deben encontrar las patrullas para reducir el tiempo de respuesta y la carga de trabajo (figura 21, puntos rosados).

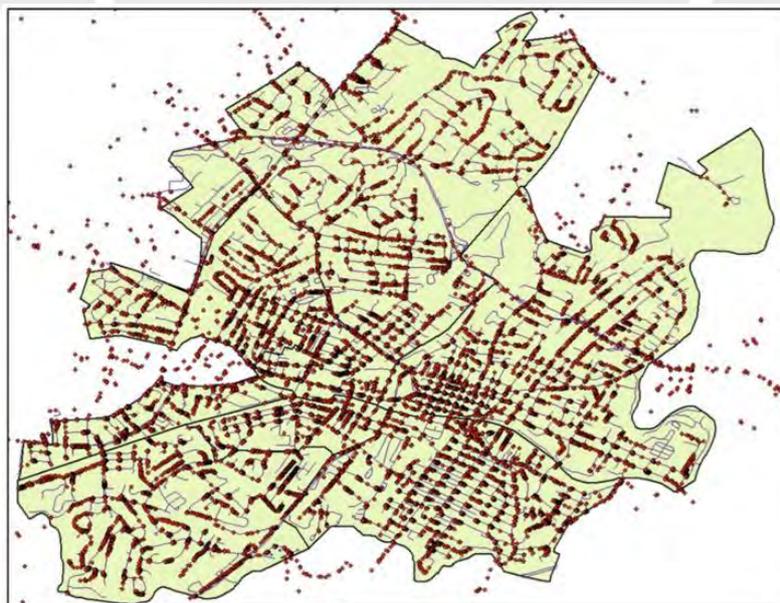


Figura 19: Localización histórica de CFS en Charlottesville

Fuente: Zhang (2013).

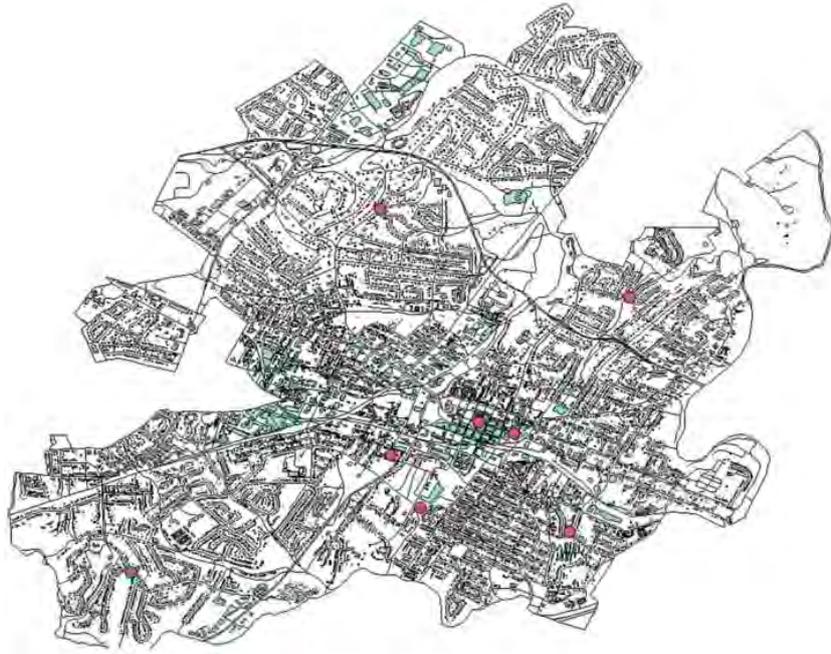


Figura 20: Modelo de simulación de patrullas policiales

Fuente: Zhang (2013).

Como se puede apreciar, se pudo hallar la mejor manera de distribuir los recursos que se tenía. Sin embargo, se debe considerar que los resultados del estudio en cuestión se pueden ver afectados tanto por cambios climáticos o incremento de índices de congestión vehicular. Es por ello que se recomienda aplicar el mismo estudio para un año entero, lo cual implicaría manejar una gran cantidad de datos.

2.8 Simulation optimization of police patrol districting plans using response surfaces

Así como en el acápite anterior, se medirá el performance del distrito de Charlottesville, con la diferencia de que, en este caso, se utilizará una simulación de evento discreto de tal manera que se pueda emplear un modelo de colas *hypercube*, el cual sirve para analizar sistemas de respuesta ante emergencias como sistemas de colas espacialmente distribuidos. Asimismo, también se aplica el algoritmo CPSC para la partición de distritos correspondiente y se obtiene

la repartición que se encontraba en la figura 19. Cabe mencionar que los objetivos planteados en el artículo en cuestión son los mismos con respecto a los del capítulo anterior.

Entonces, aplicando el modelo de Kriging para encontrar la ubicación del centro circular concéntrico, y por medio del uso de 200 puntos experimentales en el diseño *hypercube* del modelo de colas, se puede obtener el siguiente centro:

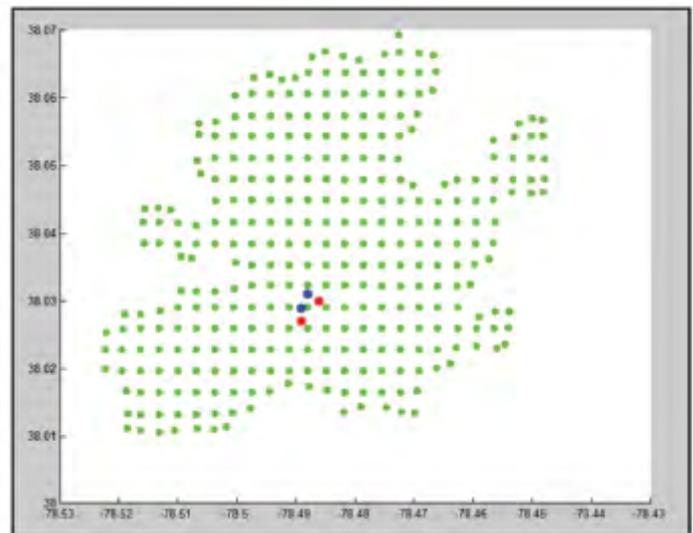


Figura 21: Centro óptimo para la atención de incidentes

Fuente: Zhang (2014).

Se puede notar que este centro, al cambiar la gráfica por una gráfica de densidad como la vista en la figura 23, en la cual el color azul representa un mayor volumen de llamadas empleadas por un ciudadano para que sea atendido por un efectivo policial, resulta concordar con la zona que recibe más llamadas en la ciudad históricamente.

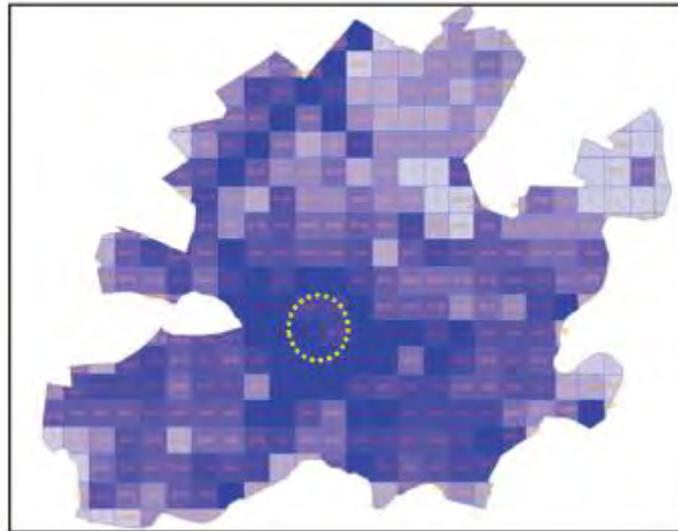


Figura 22: Gráfica de densidad del volumen de llamadas
 Fuente: Zhang (2014).

Finalmente, por medio de la simulación de evento discreto y tomando en cuenta el centro circular concéntrico se pueden obtener los siguientes resultados teniendo como ejes el tiempo de respuesta promedio y la variación de la carga de trabajo de las patrullas policiales.

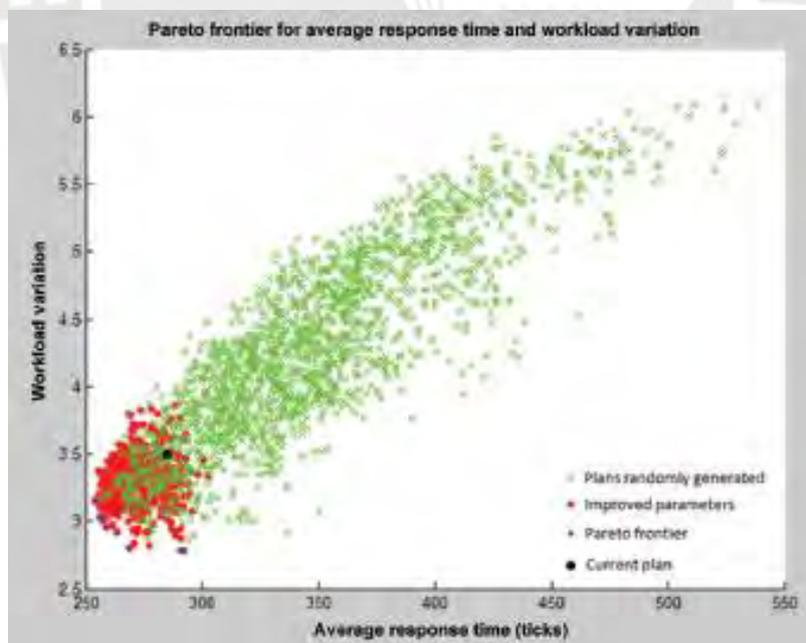


Figura 23: Gráfico Pareto de carga de trabajo vs tiempo de respuesta promedio
 Fuente: Zhang (2014).

Se concluye en que los resultados, a pesar de ser más eficientes que los del acápite anterior, muestran que los tiempos de respuesta son más sensibles al tiempo entre llegada de las llamadas correspondientes y la velocidad con la cual se responde éstas, ya que explican una larga proporción de variación del tiempo de respuesta.

2.9 Covering models and optimization techniques for emergency response facility location and planning: a review

Para este *paper* se busca cubrir de manera óptima el área que representan los sistemas de servicio médico de emergencia, puesto que su tiempo de respuesta es crucial para atender a pacientes que requieren ayuda inmediata. Por eso, se emplean modelos de cobertura que permitan obtener una solución ideal a este problema. Para ello, se aplican los modelos del *Location Set Covering Problem* (LSCP), el cual es un modelo de cobertura obligatoria en el que se debe cubrir todos los puntos de demanda al menos una vez; además, también se utiliza el modelo de *Maximal Coverage Location Problem* (MCLP), el cual tiene como objetivo maximizar la cobertura brindada como objetivo, con los recursos limitados que se dan en el problema.

A partir del uso de estos dos modelos en simultáneo se encontró una desventaja que impedía que se pudiera encontrar la solución óptima, y es que una vez que un sistema de servicio médico de emergencia es enviado para atender una llamada, otras demandas posibles que puedan surgir en ese lapso de tiempo no podrán ser atendidas aún cuando pertenecen al área de cobertura de estos servicios. Ante estas circunstancias, se procedió a añadir el modelo *Double Standard Model* (DSM), el cual cubría, mediante la instalación de sistemas EMS en sitios de emergencia potenciales para proveer una cobertura completa.

2.10 Optimal allocation of police patrol resources using a continuous-time crime model

Este artículo se enfoca en minimizar el tiempo de respuesta desde diversos enfoques. Se comienza por la asignación de patrullas policiales limitadas a ciertos puntos del espacio, para lo cual se genera un modelo simplificado de dos escenarios en el que el primer escenario la policía determina la localización espacial óptima de una serie de unidades vehiculares, mientras que en el segundo escenario las unidades responden a los incidentes de carácter criminal que ocurren en el momento. Se observa que no se puede tener certeza de las decisiones asociadas al primer escenario, lo cual no sucede en el segundo, ya que se puede asumir que las decisiones ya fueron tomadas. Esto es debido a que se considera que la colocación de una unidad vehicular o un oficial a una zona aleatoria reduce los índices de criminalidad. Para hacer uso de esta consideración se emplea un modelo estocástico regenerativo de criminalidad. Sin embargo, se requieren de otros parámetros para poder encontrar la solución.

Esto conlleva a la minimización del tiempo de respuesta para un número determinado de incidentes criminales. Considerando un número limitado de servidores en el espacio que deben atender una serie de incidentes sin algún orden establecido, se plantea el problema como un caso múltiple del problema del viajante (*Travelling Salesman Problem o TSP*). Entonces, empleando la información del año 2009 de Davidson County, Tennessee, Estados Unidos, se obtienen alrededor de 4600 incidentes, los cuales incluyen coordenadas y tiempos de ocurrencia. Además, se utilizaron los datos de los vehículos de patrullaje policial, encontrando alrededor de 31 millones de registros en el condado de Davidson en el año 2009. También se consideró al clima como un posible factor y para ello también se poseía registros correspondientes a la ciudad.

Finalmente, y por medio de los softwares R y Python, se pudo hacer uso de un algoritmo de programación estocástico que permite contar con la dependencia entre la distribución espacio-temporal de los incidentes criminales y la localización de los efectivos policiales. Para evaluar la optimización, se presentó un modelo de espacio discreto y tiempo continuo para pronosticar el crimen como función de una serie de variables que incluyen el despliegue de unidades vehiculares. Por último, el experimento supera el estado de arte de modelos de tiempo discreto o continuo de predicción de incidentes criminales tanto en términos de efectividad de predicción y tiempos de corrida.

2.11 Queueing Models of Call Centers: An Introduction

En este *paper* se presenta un conjunto de modelos relacionados a la atención de la industria de los *call centers*, los cuales también están relacionados a una mejora en el ámbito de atención de emergencias que involucran el envío de oficiales a algún incidente. Para ello, Koole (2001) muestra el siguiente sistema de colas aplicado a un call center:

A Simple Call Center

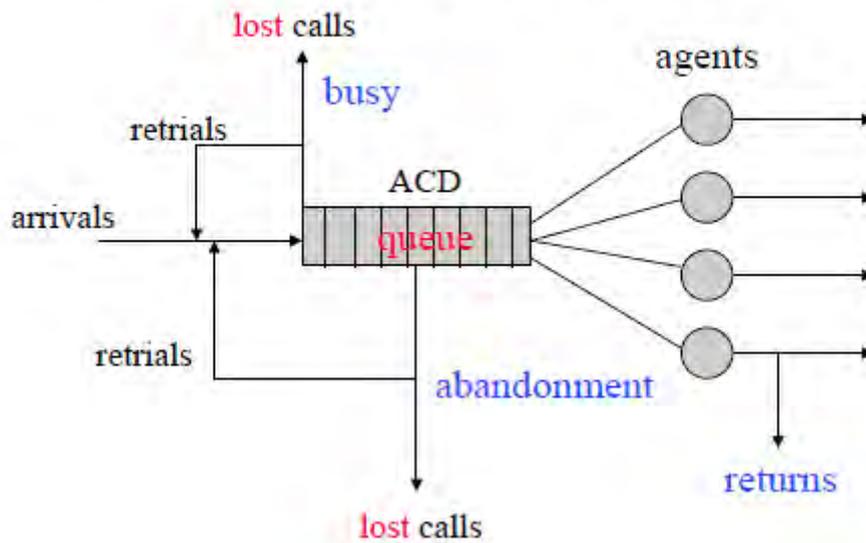


Figura 24: Modelo simple de un *call center*

Fuente: Koole y Mandelbaum (2001).

Como se observa en la figura 26, hay presentes tres salidas en todo el proceso: una cuando no se contesta la llamada en cuestión, la cual se considera como llamada perdida; otra salida se muestra cuando, durante la llamada en cuestión, el cliente procede a colgar por lo que se considera también como llamada perdida; finalmente, está el caso en el que el cliente pasa por el ACD (Automatic Calls Distribution) y luego procede a conversar con un agente, el cual luego retornaría la llamada. Como se explicó anteriormente, este modelo se considera simple porque abarca temas generales de los *call centers* pero puede presentar ciertas variaciones de acuerdo al proceso que se lleva a cabo.

Por otro lado, los indicadores de performance propuestos se encargan de calcular métricas relacionadas al abandono, espera y rescate de llamadas en cuestión. Para el factor de abandono, se divide la cantidad de clientes que abandonaron la cola entre la cantidad total de clientes. El factor de espera se calcula como el promedio del tiempo del tiempo en cola por el cual pasan los clientes durante el proceso de llamada. Finalmente, los retornos se calculan como la

fracción de aquellos clientes cuyos requerimientos fueron satisfechos en el primer intento entre el total de clientes. A continuación, se muestran las economías de escala (indicadores de evaluación) para poder minimizar los tiempos de espera en *call centers*:

Tabla 8: Economías de escala para evaluación de modelo de *call centers*

	Base Case	Efficiency-driven	Quality-driven	Rationalized
Offered load	$R = \frac{\lambda}{\mu}$	mR	mR	mR
Safety staffing	Δ	Δ	$m\Delta$	$\sqrt{m}\Delta$
Number of agents	$N = R + \Delta$	$mR + \Delta$	$mR + m\Delta$	$mR + \sqrt{m}\Delta$
Service grade	$\beta = \frac{\Delta}{\sqrt{R}}$	$\frac{\beta}{\sqrt{m}}$	$\beta\sqrt{m}$	$\boxed{\beta}$
Erlang-C = $P\{\text{Wait} > 0\}$	$P(\beta)$	$P\left(\frac{\beta}{\sqrt{m}}\right) \uparrow 1$	$P(\beta\sqrt{m}) \downarrow 0$	$\boxed{P(\beta)}$
Occupancy	$\rho = \frac{R}{R + \Delta}$	$\frac{R}{R + \frac{\Delta}{m}} \uparrow 1$	$\boxed{\rho = \frac{R}{R + \Delta}}$	$\frac{R}{R + \frac{\Delta}{\sqrt{m}}} \uparrow 1$
ASA = $E\left[\frac{\text{Wait}}{E(S)} \mid \text{Wait} > 0\right]$	$\frac{1}{\Delta}$	$\boxed{\frac{1}{\Delta} = \text{ASA}}$	$\frac{1}{m\Delta} = \frac{\text{ASA}}{m}$	$\frac{1}{\sqrt{m}\Delta} = \frac{\text{ASA}}{\sqrt{m}}$
TSF = $P\left\{\frac{\text{Wait}}{E(S)} > T \mid \text{Wait} > 0\right\}$	$e^{-T\Delta}$	$\boxed{e^{-T\Delta} = \text{TSF}}$	$e^{-mT\Delta} = (\text{TSF})^m$	$e^{-\sqrt{m}T\Delta} = (\text{TSF})^{\sqrt{m}}$

Fuente: Koole y Mandelbaum (2001).

Se debe tomar en cuenta que el TSF representa el factor total del servicio ejecutado y que ASA representa la velocidad promedio de respuesta en minutos. Se concluye en que la industria de los *call centers* es un área de estudio que se encuentra abierta para la aplicación de varios modelos de colas. Sin embargo, aún falta ahondar en la formulación de casos más complejos que se asemejen a la realidad y así poder asegurar una mayor precisión en los resultados.

2.12 A Queueing-Linear Programming Approach to Scheduling Police

Patrol Cars

El siguiente *paper* (Kolesar, 1975) presenta una metodología para la creación de nuevos horarios que buscan mejorar la correspondencia entre la disponibilidad de patrullaje y las

demandas del servicio. Este estudio se llevó a cabo en la ciudad de Nueva York. Se explica que la demanda de unidades policiales varía considerablemente durante todo el día, para ello se muestra la siguiente gráfica que explica lo sucedido:

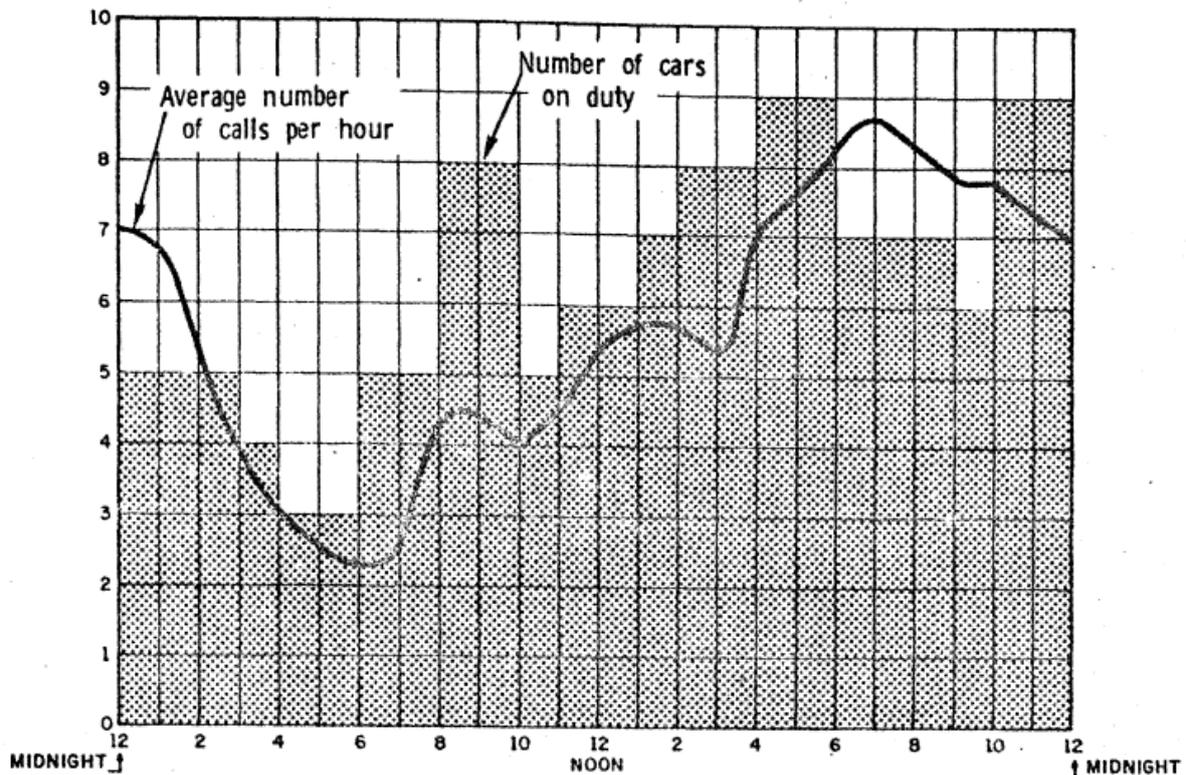


Figura 25: Cuadro de demanda de ciudadanos vs vehículos disponibles

Fuente: Kolesar (1975).

La figura 25 indica que al empezar el día hay presente una alta demanda de ciudadanos, pero no hay una alta disponibilidad de vehículos patrulla; sucede lo contrario entre las 8 y las 10 de la mañana, horario en el que hay más disponibilidad de vehículos patrulla que demanda de ciudadanos. Para ello, se hará uso de la programación lineal para poder optimizar la disponibilidad de vehículos correspondiente. El proceso consta de los siguientes pasos:

1. Especificar las políticas de restricciones que se poseen para la generación del horario.

2. Generar estimaciones para el requerimiento de vehículos por hora, empleando un modelo matemático.
3. Obtener un horario óptimo, empleando los dos primeros pasos.
4. Evaluar el horario obtenido.
5. Revisar las restricciones, puesto que se utiliza modelos aproximados, no necesariamente se cumplirá con todos los indicadores deseados.

Entonces, una vez que se obtienen las políticas, será posible elaborar el problema de programación lineal, el cual posee las siguientes variables de decisión:

n_i = número de carros asignados para trabajar, empezando en el turno i , donde i pertenece a S , el cual posee un conjunto de tiempos de comienzo de cada turno, los cuales son brindados por el departamento policial. Cada turno de trabajo dura alrededor de 8 horas.

m_{ij} = número de carros trabajando en el turno i que son asignados para la hora de refrigerio a la hora j , donde $i + e(i) \leq j \leq i + l(i)$ donde $e(x)$ y $l(x)$ representan las horas mínima y máxima a las cuales se puede llevar a cabo el refrigerio.

Puesto que se desea buscar la asignación óptima de turnos de inicio y de refrigerio que satisfaga el requerimiento de vehículos usando el número mínimo de carros. A partir de estas premisas y empleando la programación lineal entera (puesto que las unidades vehiculares se reflejan como variable entera), se muestra el modelo a continuación:

$$\begin{aligned}
& \text{minimize } \sum_{i \in S} n_i \\
& \text{subject to} \\
& \sum_{j=i+e(i)}^{i+l(i)} m_{ij} = n_i, \quad i \in S \\
& \sum_{i \in S} n_i - m_{it} \geq r_t, \quad t = 0, 23 \\
& \quad \quad \quad t-7 \leq i \leq t \\
& n_i \geq 0, m_{ij} \geq 0, \text{ and integer.}
\end{aligned}$$

Figura 26: Modelo de optimización propuesto

Fuente: Kolesar (1975).

El valor de la función objetivo es el número total de carros usados durante todo el día. La primera serie de restricciones asegura que cada carro tenga asignado un turno de refrigerio, y la segunda serie de restricciones que la cantidad actual de carros que se encuentran en turno de trabajo satisfaga cada uno de los requerimientos horarios.

Una vez desarrollado el modelo de programación lineal, se debe evaluar los resultados obtenidos. Para ello, se pudieron aplicar dos métodos: uno que involucraba la simulación del proceso completo de patrullaje vehicular, y otro referente a un modelo de colas del tipo M/M/n. Debido a la facilidad del cálculo, se optó por escoger el segundo método. Para lo cual se determinó que los arribos de las llamadas para el servicio en cualquier hora dada podían representarse por un proceso Poisson. El modelo, sin embargo, asume que cada llamada de servicio es atendida por una sola unidad vehicular (caso que sucede la mayoría de veces en la realidad). El modelo que se obtiene por el segundo método es el siguiente:

The probability that there is at least one call in queue:

$$P[X(t) \geq n(t) + 1] = \sum_{j=n(t)+1}^{\infty} p_{ij}(t_0, t) \quad (1)$$

The probability that all cars are busy:

$$P[X(t) > n(t)] = \sum_{j=n(t)}^{\infty} p_{ij}(t_0, t) \quad (2)$$

The expected number of calls in queue:

$$EQ = \sum_{j=n(t)+1}^{\infty} [j - n(t)] p_{ij}(t_0, t) \quad (3)$$

Figura 27: Modelo de colas aplicado al proceso de patrullaje vehicular

Fuente: Kolesar (1975).

Cabe destacar que t_0 es el estado del sistema en el tiempo $t = 0$ y que P_{ij} es la función de probabilidad de transición vista en las leyes de Markov. Finalmente, al emplear este modelo para la evaluación del horario obtenido, se muestran los resultados finales (empleando 24 patrullas como situación inicial y 29 como situación propuesta) a continuación:

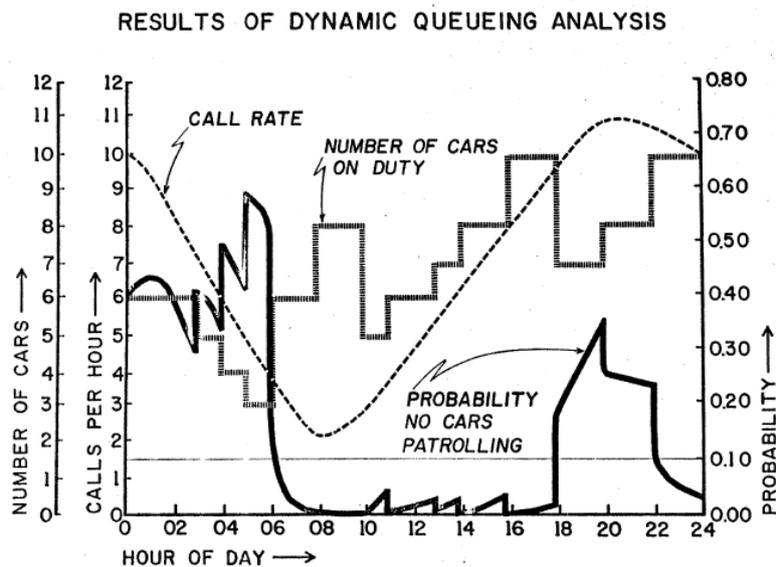


Figura 28: Resultados obtenidos con situación inicial

Fuente: Kolesar (1975).

RESULTS OF DYNAMIC QUEUEING ANALYSIS

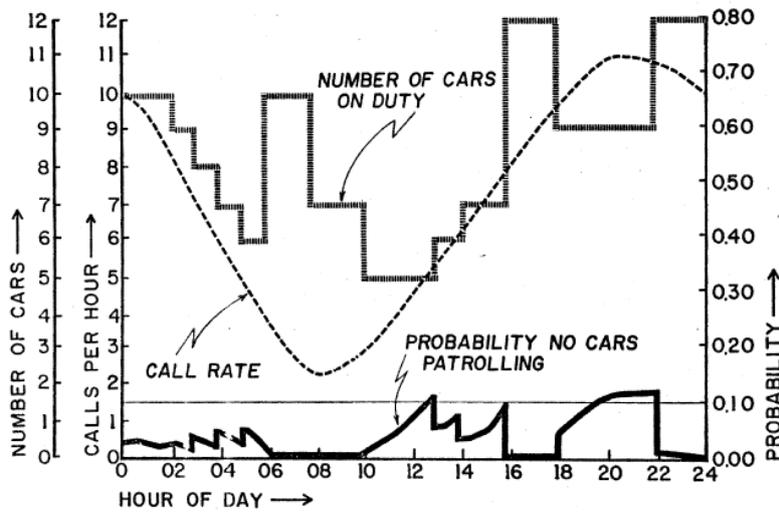


Figura 29: Resultados obtenidos con situación propuesta
Fuente: Kolesar (1975).

Se puede observar que empleando 29 vehículos patrulla es posible cumplir con las demandas en toda la ciudad de Nueva York. Sin embargo, de no poder adquirir dichos vehículos, se pudo obtener una solución más viable si se repartía el horario en 5 turnos. Dicha situación es la siguiente:

RESULTS OF DYNAMIC QUEUEING ANALYSIS

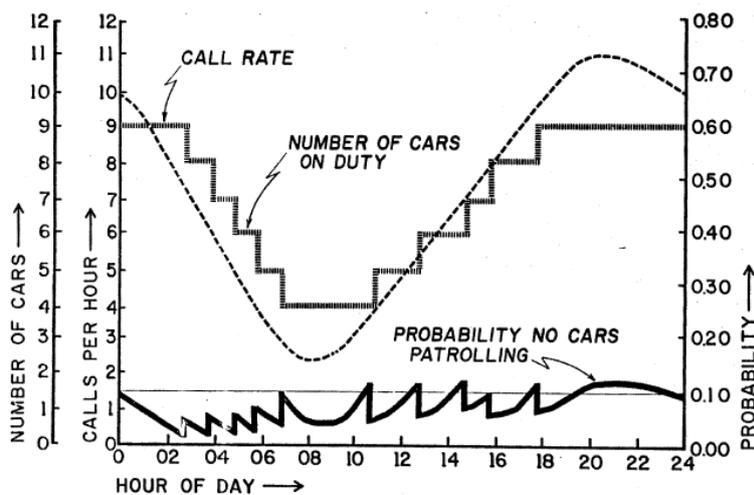


Figura 30: Resultados para la situación inicial con 5 turnos
Fuente: Kolesar (1975).

Se concluye en que, a pesar de que las aproximaciones llevan consigo un porcentaje de error y que las asunciones hechas podrían derivar en posibles fallas de lo programado en el paso 1, se comprueba que con una mayor cantidad de turnos puede haber un mejor análisis. Además, se recomienda emplear la simulación debido a la precisión que lleva consigo puesto que puede simular el proceso completo.

2.13 Queueing Model for Information Security Monitoring Systems

Los autores del *paper* en cuestión (Imamverdiyev, 2016) proponen un modelo para el proceso de manejo de incidentes de un sistema de gestión de seguridad de la información, dicho modelo se denomina como modelo M/G/1. Esto es necesario, ya que hoy en día, se ha incrementado el tráfico de contenido en red por lo que se requieren equipos computacionales potentes que permitan monitorear procesos con gran cantidad de información de manera ininterrumpida, manteniendo la integridad, confidencialidad y la calidad de la data de las empresas.

Se habla de un flujo accidental de incidentes de seguridad de información que ingresa por medio de sensores y que es manejado de forma aleatoria. Para manejar el flujo de información se puede hacer uso de herramientas teóricas como *IP protocol*. Se hace uso de varios métodos de colas para minimizar el tiempo de monitoreo y optimizar la supervisión en línea de los nodos de la red corporativa dentro de los recursos de red asignados. Para la aplicación del método, el servicio de seguridad de la información, divide los incidentes registrados en función a su criticidad como “absoluto”, “relativo” y “no prioritario”, cada uno de los cuales sigue un proceso de atención, de este modo, al ingresar un incidente absoluto, se para la atención de cualquier otro caso y se forma una cola de los mismos. Los no prioritarios pueden ser atendidos bajo las modalidades FIFO, LIFO o RAND, pues el resultado del tiempo de espera medio calculado, no se ve afectado significativamente.

A continuación, se presentan las fórmulas para el cálculo de dichos tiempos dependiendo de la clasificación según prioridad de atención:

$$W_k^{AP} = \frac{R_{k-1} b_k}{1 - R_{k-1}} + \frac{\sum_{i=1}^k \lambda_i b_i^{(2)}}{2(1 - R_k)(1 - R_{k-1})}$$

Prioridad absoluta

$$W_k^{RP} = \frac{R_{M_1} b_k}{1 - R_{M_1}} + \frac{\sum_{i=1}^M \lambda_i b_i^{(2)}}{2(1 - R_{k-1})(1 - R_k)}$$

Prioridad relativa

$$W_k^{WP} = \frac{R_{M_1} b_k}{1 - R_{M_1}} + \frac{\sum_{i=1}^M \lambda_i b_i^{(2)}}{2(1 - R_{M_1+M_2})(1 - R)}$$

No prioritario

2.14 Testing the Validity of a Queueing Model of Police Patrol

Para este *paper*, Green y Kolesar (1989) buscan otro método para hallar la forma de ajustar más a la realidad un trabajo que llevaron a cabo en la ciudad de Nueva York en 1984. Para este caso se utiliza un modelo de colas para el envío de múltiples vehículos de patrullaje, cuyos aspectos clave son los siguientes:

1. Hay presente un número ajustado, S , de carros idénticos (servidores) ejerciendo el patrullaje a lo largo de todo el turno.
2. Las llamadas de emergencia llegan una a la vez para representar un proceso Poisson con tasa λ .

3. Hay K clases de prioridad de llamadas, $k = 1, 2, \dots, K$ y P_k es la probabilidad que una llamada particular tenga prioridad k . Prioridad 1 es considerada la más alta.
4. Si una llamada que requiere de j servidores llega a encontrar suficientes vehículos disponibles, el servicio empieza de inmediato. Todos los vehículos necesarios empiezan a ejecutar el servicio de manera simultánea.
5. Si no se encuentra la cantidad de vehículos necesarios, la llamada entra en cola con una prioridad asignada.
6. La llamada que se encuentre primera en la cola tendrá vehículos asignados, los cuales se encontrarán en estado disponible hasta que la cantidad necesaria de vehículos esté libre y pueda llevarse a cabo el servicio.
7. Los tiempos de servicio de los diferentes vehículos en cada llamada son independientes y distribuidos como variables aleatorias exponenciales con media $1/\mu$, sin importar la prioridad o el número de vehículos que se requieran.

El estudio en cuestión se realizó en tres recintos de la ciudad de Nueva York (recintos 50, 26 y 77). Para el cálculo de los parámetros se hizo una toma de datos durante 10 días, en los cuales se obtuvo lo siguiente:

Tabla 9: Resultados del experimento de 10 días

No. of Responding Cars		26th Precinct		50th Precinct		77th Precinct	
		High Priority	Low Priority	High Priority	Low Priority	High Priority	Low Priority
1	Count	62	131	54	114	149	265
	Percent	63.9%	85.6%	75.0%	86.3%	68.7%	85.2%
2	Count	23	20	16	14	51	38
	Percent	23.7%	13.1%	22.2%	10.6%	23.5%	12.2%
3	Count	7	1	1	3	10	6
	Percent	7.2%	0.6%	1.4%	2.3%	4.6%	1.9%
4	Count	3		1		5	2
	Percent	3.1%		1.4%		2.3%	0.6%
5	Count	2	1		1	2	
	Percent	2.1%	0.6%		0.7%	0.9%	
Total		97	153	72	132	217	311

Fuente: Green y Kolesar (1989).

Además, con los datos se pudieron hallar los parámetros para la formulación del modelo de colas, lo cual se muestra a continuación:

Tabla 10: Parámetros de entrada

	Precinct 26	Precinct 50	Precinct 77
Average No. Calls/ Hr. (λ)	3.125	2.550	6.600
Average Service Time (Mins.) ($1/\mu$)	28.04	37.33	30.06
Proportion of High Priority Calls p_H	0.388	0.353	0.411
Dispatch Probabilities			
$p(1 H)$	0.639	0.750	0.687
$p(2 H)$	0.237	0.222	0.235
$p(3 H)$	0.124	0.028	0.078
$p(1 L)$	0.856	0.864	0.852
$p(2 L)$	0.144	0.136	0.148
Average No. of Effective Cars on Duty (\bar{n})	3.30	4.21	5.95

Fuente: Green y Kolesar (1989).

Cabe destacar que los datos fueron tomados desde las 8 am hasta las 4 pm. Entonces, con la toma de datos hecha, se desea comparar los resultados obtenidos experimentalmente con los resultados que se obtendrán a partir de los parámetros para observar cuán grandes son las variaciones porcentuales de la comparación a realizar. Para ello, se tiene que hallar una cantidad entera de vehículos y no la cantidad promedio que es posible hallar con la toma de datos, entonces se propuso dos métodos heurísticos: un método basado en el gráfico de tiempo de espera vs cantidad de vehículos patrulla y otro enfocado en la interpolación lineal entre tiempos de espera generados por el múltiple envío de carros y el número entero de vehículos cercanos al promedio. Para cada método se muestra el comparativo final calculado:

Tabla 11: Comparativo de resultados observados vs calculados

Precinct	Call Category	MCD Delay Estimates		Stationary** Estimate	Observed Delay (Standard Error)
26	All Calls	3 Cars	33.61	NA	8.28 (1.19)
		3.30 Cars*	24.89		
		4 Cars	4.54		
	High Priority	3 Cars	13.63	9.20	7.53 (2.00)
		3.30 Cars*	10.41		
		4 Cars	2.89		
50	All Calls	4 Cars	4.82	3.50	5.60 (0.51)
		4.21 Cars*	4.05		
		5 Cars	1.14		
	High Priority	4 Cars	2.94	2.20	4.78 (0.76)
		4.21 Cars*	2.49		
		5 Cars	0.81		
77	All Calls	5 Cars	50.33	13.20	13.31 (1.39)
		5.95 Cars*	48.30		
		6 Cars	9.73		
	High Priority	5 Cars	10.22	4.20	5.91 (0.86)
		5.95 Cars*	9.91		
		6 Cars	3.99		

* Obtained from linear interpolation.

** Obtained from graph.

Fuente: Green y Kolesar (1989).

Además, se muestra la similaridad que hay entre los tiempos entre llegadas observados con los calculados por el modelo en cuestión para el Recinto 26 y los tiempos de servicio observados con los calculados por el modelo generado para el Recinto 50.

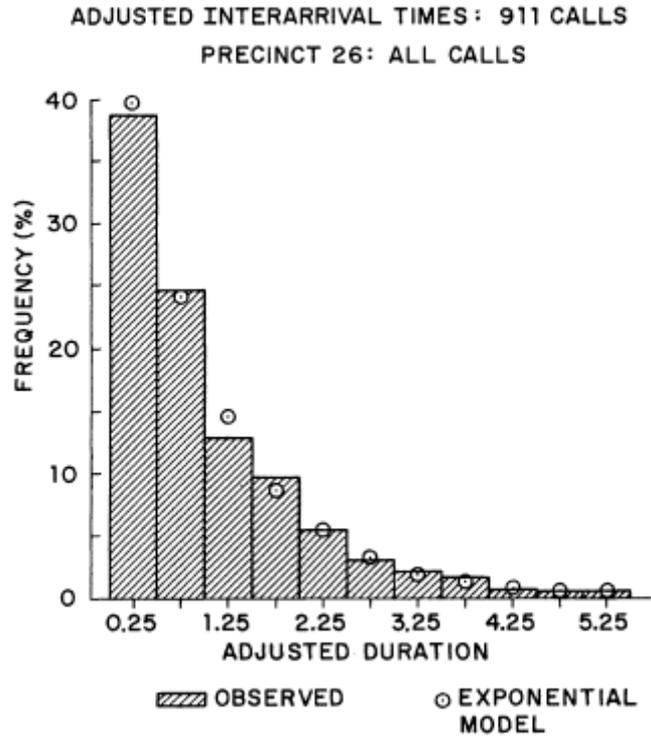


Figura 31: Comparativo de tiempos entre llegadas
Fuente: Green y Kolesar (1989).

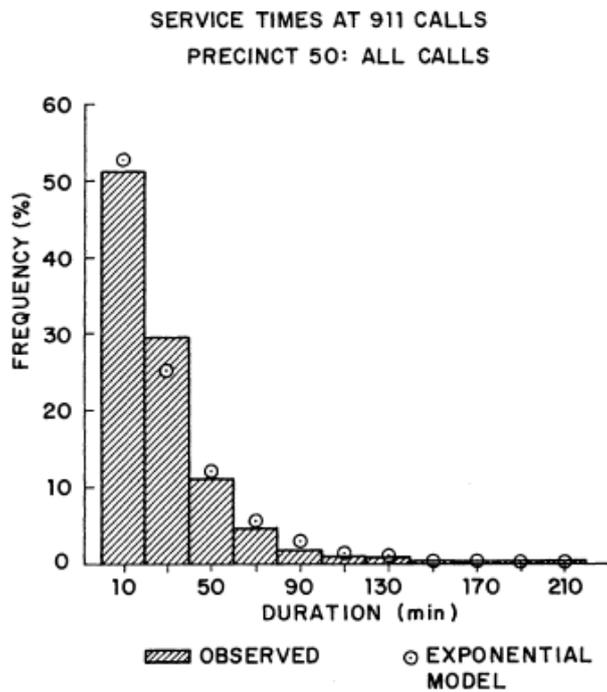


Figura 32: Comparativo de tiempos de servicio
Fuente: Green y Kolesar (1989).

Se concluyó en que el modelo de envío de múltiples unidades vehiculares es una herramienta válida para su objetivo: el envío de patrullas a través de recintos y por medio de turnos. Dentro de las presunciones destacan que la validez de la herramienta puede deberse debido a la zona de estudio por lo que no necesariamente en otra área de estudio el modelo de envío sea una herramienta válida.

2.15 Police officer scheduling using goal programming

En este *paper* (Todovic, 2015) se realiza un estudio en las estaciones policiales de Bosnia-Herzegovina para determinar la programación mensual de oficiales. De acuerdo a los datos obtenidos, se tuvo que los efectivos policiales laboraban en dos turnos de 12 horas. Además, el comandante de cada estación tiene como prioridad mantener un buen indicador de performance para todos los oficiales de su estación. Debido a la imposibilidad de cumplir con todas las demandas que presentan los comandantes, resulta necesario asociar este problema como un problema de programación por metas. Las asunciones del modelo a crear son las siguientes:

1. La estación policial de estudio posee 11 efectivos policiales.
2. Los oficiales serán programados para el periodo de un mes.
3. El trabajo está organizado en dos horarios (día y noche) que duran 12 horas.
4. Es obligatorio completar las 180 horas de trabajo estándar mensuales.
5. Durante la programación de horarios de los oficiales, la regla de que luego de las 12 horas de trabajo siguen 12 horas de descanso es tomada en cuenta.
6. Se respetan requerimientos de los empleados para la toma de días libres o vacaciones.
7. Durante el mes, un número de oficiales puede ausentarse. Por casusas del modelado de la programación de horarios, las ausencias se dividirán en dos grupos: ausencia de 1 día y ausencia de 2 días.

Las variables de decisión binarias se muestran a continuación:

$$a_{kj} = \begin{cases} 1 & \text{if schedule } k \text{ includes shift } j \\ 0 & \text{otherwise} \end{cases}, \quad j \in D, k \in R$$

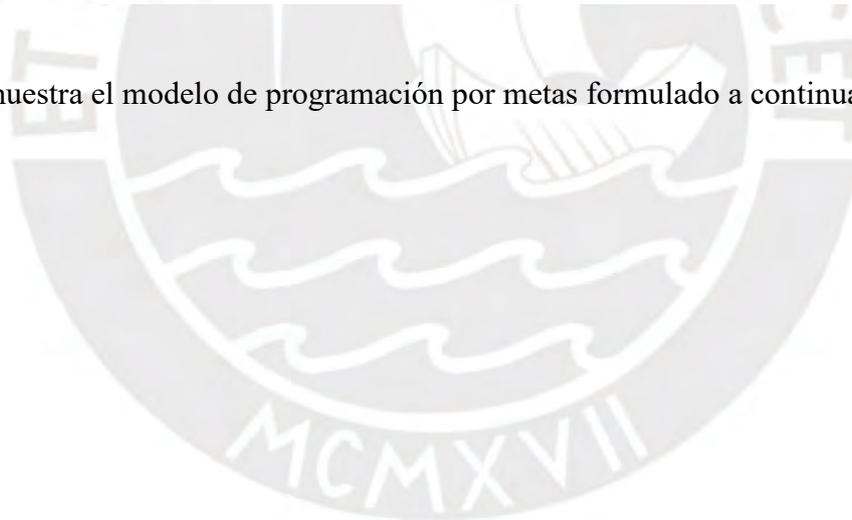
R es la serie de horarios que pueden ser utilizados durante el mes. D_d es la serie de cambios durante el día, D_n es la serie de cambios durante la noche y D es la unión entre D_d y D_n y Z es la serie de efectivos policiales en la estación policial observada.

$$c_{ij} = \begin{cases} 1 & \text{if the police officer } i \text{ is absent in the shift } j \\ 0 & \text{otherwise} \end{cases}, \quad i \in Z, j \in D - \text{for one - day absence,}$$

$$b_{ij} = \begin{cases} 0 & \text{if the police officer } i \text{ is absent in the shift } j \\ 1 & \text{otherwise} \end{cases}, \quad i \in Z, j \in D - \text{for two or more days absence;}$$

$$x_{ik} = \begin{cases} 1 & \text{if schedule } k \text{ is assigned to the police officer } i \\ 0 & \text{otherwise} \end{cases}, \quad i \in Z, k \in R;$$

Además, se muestra el modelo de programación por metas formulado a continuación:



$$\min \sum_{i \in Z} (d_i^- + d_i^+) + \sum_{j \in D_d} dd_j + \sum_{j \in D_n} dn_j$$

$$\sum_{k \in R} x_{ik} = 1, \quad i \in Z$$

$$\sum_{k \in R} a_{kj} \cdot \sum_{i \in Z} b_{ij} \cdot x_{ik} + dd_j \geq pd_j, \quad j \in D_d$$

$$\sum_{k \in R} a_{kj} \sum_{i \in Z} b_{ij} \cdot x_{ik} + dn_j \geq pn_j, \quad j \in D_n$$

$$\sum_{k \in R} a_{kj} \cdot \sum_{i \in Z} b_{ij} \cdot x_{ik} \geq pdm_j, \quad j \in D_d$$

$$\sum_{k \in R} a_{kj} \cdot \sum_{i \in Z} b_{ij} \cdot x_{ik} \geq pmm_j, \quad j \in D_n$$

$$\sum_{k \in R} \sum_{j \in D} a_{kj} \cdot c_{ij} \cdot x_{ik} = 0, \quad i \in Z$$

$$\sum_{k \in R} \sum_{j \in D} a_{kj} \cdot b_{ij} \cdot x_{ik} \cdot rt + d_i^- - d_i^+ = t \cdot pr_i, \quad i \in Z$$

$$x_{ik} \in [0, 1], \quad i \in Z, \quad k \in R$$

$$d_i^- \geq 0, \quad d_i^+ \geq 0, \quad i \in Z$$

$$dd_j \geq 0, \quad j \in D_d$$

$$dn_j \geq 0, \quad j \in D_n$$

Figura 33: Modelo de programación por metas

Fuente: Elaboración propia.

Las reglas de existencia son las siguientes:

$$x_{ik} \in [0, 1], \quad i \in Z, \quad k \in R$$

$$d_i^- \geq 0, \quad d_i^+ \geq 0, \quad i \in Z$$

$$dd_j \geq 0, \quad j \in D_d$$

$$dn_j \geq 0, \quad j \in D_n$$

Con el modelo en cuestión se pudo obtener el horario óptimo que pueden seguir los 11 efectivos policiales de cada estación para poder obtener la mayor eficiencia posible. Además, se presenta un cuadro comparativo entre el modelo actual con el modelo propuesto que presenta las horas de trabajo mensuales que laboraba cada oficial en conjunto.

Tabla 12: Cuadro de resultados actuales vs propuestos

Police officer	Old scheduling	Optimal solution two-shifts
P1	180	180
P2	192	180
P3	192	180
P4	192	180
P5	192	180
P6	204	180
P7	192	180
P8	180	180
P9	108	108
P10	156	180
P11	216	180

Fuente: Todovic y otros (2015).

CAPÍTULO 3. CONCLUSIONES

Luego de haber observado todos los artículos y *papers* mencionados en el capítulo anterior, se puede concluir en lo siguiente:

1. Para resolver el problema de envío de efectivos policiales hay presente una vasta cantidad de herramientas de investigación de operaciones, desde la programación lineal, el uso de teoría de colas hasta la simulación de procesos.
2. Las herramientas de investigación de operaciones presentan cierta flexibilidad para combinarse con otras, como los sistemas de información geográfica, algoritmos, entre otros, las cuales permiten obtener parámetros que brindan mayor precisión al problema que se desea modelar.
3. A lo largo de los años, se puede observar que hay presente un uso de más herramientas que pueden facilitar el cálculo y asegurar resultados que albergan una mayor cantidad de datos o cuya zona geográfica alberga una mayor área.
4. La reducción de los tiempos de espera de los efectivos policiales resulta ser un tema que ha sido de interés desde hace más de dos décadas y aún sigue habiendo estudios al respecto para proponer mejores opciones que permitan mejorar los sistemas de seguridad que hay presentes en la actualidad.
5. La aplicación de algún modelo y cuán preciso es puede depender en gran parte de la localización de estudio, por lo que antes de asumir que el modelo puede funcionar en

una nueva zona se debería iniciar el estudio desde la toma de datos experimentales para comprobar la validez del modelo.

6. No hay una herramienta que sea mejor que todas, mejor dicho, en cada herramienta a emplear siempre habrá ciertas presunciones, las cuales impiden que los resultados del proceso a analizar tiendan a parecerse a los resultados obtenidos en la realidad. Sin embargo, cada herramienta puede emplearse de mejor manera para ciertos casos y no para otros, por lo que es recomendable diferenciar entre todas las herramientas presentes cuál es la indicada para el proceso que se desea estudiar.
7. El uso de la programación lineal resulta ser más práctico a comparación de otras herramientas. Sin embargo, la estimación de parámetros a emplear debe poseer un menor grado de error para obtener resultados más precisos. Un caso similar sucede con el uso de la teoría de colas para la resolución de algún problema de envío de unidades de emergencia.
8. Aunque se vio en una menor cantidad de artículos, el modelo hipercúbico de colas representa una herramienta que se está empleando con mayor tendencia en los últimos años, esto es debido a la practicidad que presenta para poder establecer prioridades o estratificar la cola por servidor.
9. La simulación resulta ser una herramienta de uso amplio y que presenta gran cantidad de facilidades para poder modelar procesos de alta complejidad. Por ello, este instrumento es recomendable cuando se desea representar un proceso en la vida real y

que también cuente con una gran cantidad de información para realizar un análisis de datos más confiable.

10. Los sistemas de información geográfica proveen mediante coordenadas las localizaciones donde puede requerirse un mayor número de envíos y sirven de utilidad en la elaboración del modelo ya que esta información puede reemplazar a posibles presunciones que pudieron tomarse en cuenta para la simulación respectiva.
11. Por la cantidad de artículos vistos es notable destacar que la búsqueda de un mejor sistema de seguridad es un tema de relevancia para la mayoría de autores, y aunque no haya presente un modelo general a aplicar para todos los lugares del mundo, es posible decir que la elaboración de cada uno de estos modelos sirvió para optimizar al menos uno o más lugares de estudio por lo que podría tratar de aplicarse alguno de ellos o emplear una variación de éstos a este país.

Bibliografía

DUNNETT, Sarah y otros

2019 “Optimizing police dispatch for incident response in real time”. *Journal of the Operational Research Society*, Loughborough, volumen 70, número 2, pp. 269-279. Consulta: 17 de mayo de 2020.

ZHAO, Ke Quan y otros

2012 “The optimal dispatch of traffic and patrol police service platforms”. *Journal of Applied Mathematics*, Chongqing, volumen 2012, pp. 1-9. Consulta: 25 de julio de 2020.

MAXFIELD, Michael

1982 “Service time, dispatch time, and demand for police services: helping more by serving less”. *Public Administration Review*, San Francisco, volumen 42, número 3, pp. 252-263. Consulta: 25 de julio de 2020.

DEWINTER, Maite y otros

2020 “Analysing the police patrol routing problem: a review”. *International Journal of Geo-Information*, Kansas, volumen 2020, número 9, pp. 157-174. Consulta: 25 de julio de 2020.

ZHANG, Yue y BROWN, Donald

2013 “Police patrol districting method and simulation evaluation using agent-based model & GIS”. *Security Informatics*, volumen 2, número 7. Consulta: 25 de julio de 2020.

BLANES, Jordi y KIRCHMAIER, Tom

2018 “The effect of police response time on clearance rates”. *Review of Economic Studies*, volumen 85, número 2, pp. 855-891. Consulta: 17 de mayo de 2020.

CURTIN, Kevin y otros

2010 “Determining optimal police patrol areas with maximal covering and backup covering location models”. *Networks and Spatial Economics*, volumen 10, pp. 125-145. Consulta: 17 de mayo de 2020.

ZHANG, Yue y BROWN, Donald

2014 “Simulation optimization of police patrol districting plans using response surfaces”. *Simulation: Transactions of the Society for Modeling and Simulation International*, volumen 90, número 6, pp. 687-705. Consulta: 25 de julio de 2020.

LI, Xueping y otros

2011 “Covering models and optimization techniques for emergency response facility location and planning: a review”. *Mathematical Methods of Operations Research*, volumen 74, pp. 281-310. Consulta: 25 de julio de 2020.

MUKHOPADHYAY, Ayan

2016 “Optimal allocation of police patrol resources using a continuous-time crime model”. *Game Sec 2016: Decision and Game Theory for Security*, volume 9996, pp. 139-158. Consulta: 25 de julio de 2020.

NARAYAN, U.

2008 *An introduction to queueing theory*. Boston: Birkhäuser

MEDHI, Jyotiprasad

2003 *Stochastic models in queueing theory*. Segunda edición. Massachusetts: Academic Press.

BAUM, Dieter y BREUER, L.

2005 *An introduction to queueing theory and matrix analytic methods*. Dordrecht: Springer.

COOPER, Robert

1981 *Introduction to queueing theory*. New York: North Holland.

LIPSKY, Lester

2009 *Queueing theory: a linear algebraic approach*. Segunda edición. New York: Springer.

HILLIER, Frederick y LIEBERMAN, Gerald

1981 *Introduction to Operations Research*. Novena edición. New York: McGraw-Hill.

TAHA, Hamdy

2017 *Operations research: an introduction*. Décima edición. Essex: Pearson Education Limited.

KENDALL, David

1953 “Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain”. *The Annals of Mathematical Statistics*, pp. 338-354. Consulta: 24 de mayo de 2020.

LUENBERGER, D.G.

1984 *Linear and nonlinear programming*. Segunda edición. Massachusetts: Addison-Wesley.

BRICKMAN, Louis

1989 *Mathematical introduction to linear programming and game theory*. New York: Springer-Verlag.

JUNGNICKEL, Dieter

2013 *Graphs, networks and algorithms*. Berlin: Springer-Verlag.

VAN STEEN, Maarten

2010 *Graph theory and complex networks: An Introduction*. Twente: Maarten Van Steen.

SOBOL, I. M.

1983 *Método de Montecarlo*. Moscú: Mir.

BEOJONE, Caio Vitor y SOUZA, Regiane

2017 “Application of the hypercube models with queue priorities and more than one preferential server: a case study on a SAMU”. *Gestão & Produção*, volumen 24(4), pp 814-828. Consulta: 25 de octubre de 2020.

LARSON, Richard y ODONI, Amedeo

1981 *Urban Operations Research*. Cambridge: Prentice-Hall.

LARSON, Richard

1974 “A hypercube queueing model for facility location and redistricting in urban emergency services”. *Computers & Operations Research*, volume 1(1), pp 67-95. Consulta: 20 de octubre de 2020.

KOOLE, Ger y MANDELBAUM, Avishai

2001 “Queueing Models of Call Centers: An Introduction”. *Annals of Operations Research*, volume 113 (2002), pp 41-59. Consulta: 28 de octubre de 2020.

KOLESAR, Peter y otros

1975 *A Queueing-Linear Programming Approach to Scheduling Police Patrol Cars*. California: RAND Corporation.

<https://www.rand.org/pubs/papers/P5260-1.html>

GREEN, Linda y KOLESAR, Peter

1989 “Testing the Validity of a Queueing Model of Police Patrol”. *Management Science*, volume 35 (2), pp 127-148. Consulta: 27 de octubre de 2020.

IMAMVERDIYEV, Yadigar y NABIYEV, Babak

2016 “Queueing Model for Information Security Monitoring Systems”. *Problems of information technology*, 2016 (1), pp 28-32. Consulta: 27 de octubre de 2020.

TODOVIC, Dragana y otros

2015 “Police officer scheduling using goal programming”. *Policing an International Journal of Police Strategies and Management*, volume 38 (2), pp 295-313. Consulta: 27 de octubre de 2020.