

**ANÁLISIS DE REDES DE COLAS MODELADAS CON TIEMPOS ENTRE LLEGADAS
EXPONENCIALES E HÍPER ERLANG PARA LA ASIGNACIÓN EFICIENTE DE LOS
RECURSOS**

CAMILO ERNESTO MARTÍNEZ ERASO

**PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA
INGENIERÍA INDUSTRIAL
BOGOTÁ D.C.
2009**

**ANÁLISIS DE REDES DE COLAS MODELADAS CON TIEMPOS ENTRE LLEGADAS
EXPONENCIALES E HÍPER ERLANG PARA LA ASIGNACIÓN EFICIENTE DE LOS
RECURSOS**

CAMILO ERNESTO MARTÍNEZ ERASO

Trabajo de Grado

Director:
LINA MARCELA RANGEL
Ingeniera Industrial

**PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA
INGENIERÍA INDUSTRIAL
BOGOTÁ D.C.
2009**

CONTENIDO

INTRODUCCIÓN	12
1 PLANTEAMIENTO DEL PROBLEMA	13
1.1 ANTECEDENTES	13
1.2 FORMULACIÓN DEL PROBLEMA	18
2 JUSTIFICACIÓN DEL PROYECTO	19
3 OBJETIVOS	23
3.1 OBJETIVO GENERAL	23
3.2 OBJETIVOS ESPECÍFICOS	23
4 MARCO TEÓRICO	24
4.1 PROCESOS ESTOCÁSTICOS	24
4.2 TEORÍA DE COLAS	24
4.2.1 Sistema de clasificación	26
4.2.2 Medidas de desempeño	27
4.3 REDES DE COLAS	27
4.3.1 Clasificación de las redes de colas	28
4.3.2 Redes de Jackson	29
4.3.2.1 Medidas de desempeño	29
4.4 APLICACIONES DE TEORÍA DE COLAS	30
4.5 SIMULACIÓN	30
4.6 OPTIMIZACIÓN	31
4.6.1 Tipos de Optimización	31
5 REDES DE COLAS EN CENTROS DE SERVICIO	33
5.1 DEFINICIÓN DE LA RED DE COLAS	33
5.2 FACTORES DE INFLUENCIA	35
5.2.1 Distribuciones de probabilidad de los tiempos entre llegadas	35
5.2.2 Distribuciones de probabilidad de tiempos entre llegadas a contrastar	38
5.2.3 Cantidad de Nodos	40
5.2.4 Cantidad de servidores por nodo	40

5.2.5	Porcentaje de utilización por nodo	41
5.2.6	Distribución de tiempos de servicio y Tasa de servicio por nodo	42
5.2.7	Capacidad del sistema	43
5.3	DISEÑO DEL EXPERIMENTO	44
5.3.1	Objetivo del estudio	45
5.3.2	Variables de respuesta	45
5.3.3	Factores controlables	46
5.3.4	Tipo de mediciones/observaciones	46
5.3.5	Niveles de los factores	47
5.3.6	Tamaño del experimento	47
6	SIMULACIÓN DEL EXPERIMENTO	49
6.1	CONFIGURACIÓN DE LA SIMULACIÓN	49
6.1.1	Definición de réplicas y duración	49
6.1.2	Técnicas de reducción de varianza	50
6.2	TIEMPO DE EJECUCIÓN DE LA SIMULACIÓN	53
6.3	REQUERIMIENTOS DE SISTEMA	53
6.4	CONSTRUCCIÓN DEL LOS MODELOS DE SIMULACIÓN	53
6.4.1	Configuración de los módulos	54
6.4.1.1	Create	54
6.4.1.2	Decide	55
6.4.1.3	Process	55
6.4.1.4	ReadWrite	56
6.4.1.5	Dispose	56
6.4.2	Modelos de simulación	57
7	RESULTADOS DEL EXPERIMENTO	58
7.1	GENERALIDADES: INTERACCIONES ENTRE LOS FACTORES	58
7.2	ANÁLISIS DEL NÚMERO PROMEDIO DE PERSONAS EN COLA	59
7.2.1	Relación entre número esperado de personas en cola y nodo de atención	59
7.2.2	Relación entre número esperado de personas en cola y factor de utilización	62
7.2.3	Relación entre número esperado de personas en cola y número de servidores	63
7.2.4	Relación entre número esperado de personas en cola y capacidad del nodo	64
7.2.5	Relación entre número esperado de personas en cola, cantidad de servidores y nivel de utilización	66
7.2.6	Relación entre número esperado de personas en cola, capacidad del nodo y nivel de utilización	67
7.2.7	Relación entre número esperado de personas en cola, capacidad del nodo y cantidad de servidores	69

7.2.8	Resumen de interacción de factores sobre el número esperado de personas en cola	71
7.3	TIEMPO PROMEDIO DE ESPERA EN LA COLA	74
7.3.1	Relación entre tiempo esperado en cola y nodo de atención	74
7.3.2	Relación entre tiempo esperado en cola y el factor de utilización	75
7.3.3	Relación entre tiempo esperado en cola y número de servidores	76
7.3.4	Relación entre tiempo esperado en cola y capacidad del nodo	78
7.3.5	Relación entre tiempo esperado en cola, cantidad de servidores y nivel de utilización	79
7.3.6	Relación entre tiempo esperado en cola, capacidad del nodo y nivel de utilización	80
7.3.7	Relación entre tiempo esperado en cola, capacidad del nodo y cantidad de servidores	82
7.3.8	Resumen de interacción de factores sobre el tiempo esperado en cola	84
7.4	ANÁLISIS GENERAL DEL SISTEMA	86
7.4.1	Subconjuntos homogéneos	86
7.5	RESUMEN MEDIDAS DE DESEMPEÑO	87
7.5.1	Generalidades	87
7.5.2	Distribución Exponencial	88
7.5.3	Distribución Híper-Erlang	88
8	DISEÑO DEL MODELO DE OPTIMIZACIÓN	89
8.1	COSTOS ASOCIADOS CON LAS REDES DE COLAS	90
8.1.1	Costos de mano de obra	91
8.1.2	Costos de utilización del sistema	92
8.1.3	Costos de espera	92
8.1.4	Costos de puestos de trabajo	93
8.1.5	Costos de oportunidad	94
8.1.6	Otros Costos	94
8.2	MODELO DE OPTIMIZACIÓN	94
8.2.1	Definición de variables	94
8.2.2	Funciones objetivo	95
8.2.3	Restricciones	96
8.2.3.1	Restricciones asociadas con los servidores	96
8.2.3.2	Restricciones asociadas con la capacidad	96
8.2.3.3	Restricciones asociadas con los costos	96
8.2.4	Modelo Final	97
8.3	SOLUCIÓN DEL MODELO	98
8.3.1	Evaluación y selección de la heurística vía teoría de juegos	98
8.3.2	Diseño del Algoritmo	102

8.3.2.1	Configuración inicial	102
8.3.2.2	Generación del vecindario	102
8.3.2.3	Selección del mejor vecino	102
8.3.2.4	Evaluación de los vecinos	103
8.3.2.5	Criterio de parada	104
8.3.3	Funcionamiento	106
8.3.4	Evaluación	107
8.4	EJEMPLO	107
9	HERRAMIENTA DE APLICACIÓN	111
9.1	GENERALIDADES	111
9.1.1	Construcción	111
9.1.2	Información Utilizada	111
9.1.3	Requerimientos de Funcionamiento	111
9.1.4	Funcionamiento de la Herramienta	113
9.2	DISEÑO DE LA HERRAMIENTA	113
9.2.1	Información de entrada	113
9.2.1.1	Parámetros	114
9.2.1.2	Costos	115
9.2.1.3	Modelo	115
9.2.1.4	Información de salida	116
9.3	USO APROPIADO DE LA HERRAMIENTA	118
9.3.1	Verificar Compatibilidad	118
9.3.2	Recolección de la información	118
9.3.3	Procedimiento de uso	118
10	CONCLUSIONES	120
12	RECOMENDACIONES	122
13	BIBLIOGRAFÍA	123

ÍNDICE DE FIGURAS

Figura 1. Estructura de sistemas de colas	25
Figura 2. Clasificación Kendall y Lee	26
Figura 3. Tipos de distribuciones	26
Figura 4. Redes de colas	27
Figura 5. Clasificación de las redes de colas.	28
Figura 6. Tipos de redes	33
Figura 7. Estructura de la red de colas definida	34
Figura 8. Cantidad de nodos en serie	40
Figura 9. Indicadores dentro de los reportes de simulación	45
Figura 10. Factores Anidados	47
Figura 11. Configuración de réplicas en Arena	50
Figura 12. Configuración del modulo Create	54
Figura 13. Configuración del modulo Decide	55
Figura 14. Configuración del modulo Process	56
Figura 15. Modelos de simulación.	57
Figura 16. Redes de colas con capacidad finita	60
Figura 17. Resultados de Lq según el nodo de atención con tiempos entre llegadas exponenciales	60
Figura 18. Resultados de Lq según el nodo bajo la distribución Híper-Erlang.	61
Figura 19. Resultados de Lq según el factor de utilización bajo la distribución exponencial en el primer nodo.	62
Figura 20. Comparativo distribuciones según el número de servidores	64
Figura 21 Comparativo distribuciones según la capacidad (Lq)	65

Figura 22. Resultados de L_q en el primer nodo según el número de servidores bajo la distribución Exponencial	66
Figura 23. Resultados de L_q en el primer nodo según el número de servidores bajo la distribución Híper-Erlang.	67
Figura 24. Resultados de L_q según la capacidad en primer nodo bajo la distribución Exponencial	68
Figura 25. Resultados de L_q según la capacidad en el primer nodo bajo la distribución Híper-Erlang.	69
Figura 26. Resultados de L_q según el número de servidores y la capacidad el primer nodo bajo la distribución Exponencial	70
Figura 27. Resultados de L_q según el número de servidores y la capacidad en el primer nodo bajo la distribución Híper-Erlang	71
Figura 28. Comparativo cambio del W_q a lo largo de los nodos entre las distribuciones	74
Figura 29. Resultados de W_q en el primer nodo según el factor de utilización.	75
Figura 30. Resultados de W_q según el factor de utilización bajo la distribución Híper-Erlang.	76
Figura 31. Comparativo distribuciones según la capacidad	77
Figura 32. Comparativo distribuciones según capacidad (W_q)	78
Figura 33. Resultados de W_q en todos los nodos según el número de servidores bajo la distribución Exponencial	79
Figura 34. Resultados de W_q en el primer nodo según el número de servidores bajo la distribución Híper-Erlang.	80
Figura 35. Resultados de W_q según la capacidad en el primer nodo bajo la distribución Exponencial	81
Figura 36. Resultados de W_q según la capacidad en el primer nodo bajo la distribución Híper-Erlang	81
Figura 37. Resultados de W_q según el número de servidores y la capacidad en todos los nodos bajo la distribución Exponencial	82
Figura 38. Resultados de W_q según el número de servidores y la capacidad en el primer nodo bajo la distribución Híper-Erlang	83
Figura 39. Costos del servicio.	90

Figura 40. Espacio en R2 de soluciones multiobjetivo.	101
Figura 41. Fraccionamiento de la red en nodos.	107
Figura 42. Frontera Pareto por nodos en una red de colas	110
Figura 43. Listado de referencias VBA Project	112
Figura 44. Esquema de la herramienta	113
Figura 45. Pestañas del menú de entrada.	114
Figura 46. Formulario de Parámetros	114
Figura 47. Formulario de Costos	115
Figura 48. Formulario del Modelo	115
Figura 49. Solución del modelo	116
Figura 50. Información impresa generada en el archivo aplicativo.xlsx	117

ÍNDICE DE TABLAS

Tabla 1. Resumen parámetros de distribuciones de llegada	39
Tabla 2. Utilización del sistema.	41
Tabla 3. Resumen de niveles por factores.	47
Tabla 4. Número de combinaciones del experimento	48
Tabla 5. Cambio entre distribuciones según la longitud de la red (l_q)	62
Tabla 6. Cambio entre distribuciones según el factor de utilización (L_q)	63
Tabla 7. Cambio entre distribuciones según el número de servidores (L_q)	64
Tabla 8. Comparativo distribuciones según la capacidad (L_q)	65
Tabla 9. Resumen relación entre número esperado de personas en cola y nodo de atención	71
Tabla 10. Resumen relación número esperado de personas en cola y factor de utilización	72
Tabla 11. Resumen relación número esperado de personas en cola y número de servidores	72
Tabla 12. Resumen relación número esperado de personas en cola y capacidad del nodo	73
Tabla 13. Resumen Relación número esperado de personas en cola, capacidad del nodo y cantidad de servidores	73
Tabla 14. Comparativo distribuciones según la longitud (W_q)	75
Tabla 15. Comparativo Distribuciones según el factor de utilización (W_q)	76
Tabla 16. Comparativo distribuciones según el numero de servidores (W_q)	77
Tabla17. Comparativo distribuciones según capacidad	78
Tabla 18. Resumen relación tiempo esperado en cola y nodo de atención	84
Tabla 19. Resumen relación tiempo esperado en cola y el factor de utilización	84
Tabla 20. Resumen relación tiempo esperado en cola y número de servidores	85

Tabla 21. Resumen relación tiempo esperado en cola y capacidad del nodo	85
Tabla 22. Resumen relación número esperado de personas en cola, capacidad del nodo y cantidad de servidores	86
Tabla 23. Juego entre los objetivos.	100

ÍNDICE DE ANEXOS

Anexo A. Macro de configuración y ejecución de la simulación	129
Anexo B. Macro para la administración y unificación de la información desde Excel.	137
Anexo C. Resultado En SPSS	144
Anexo D. Resultado Lq en Redes de colas	145
Anexo E. Resultados Wq redes de colas.	154
Anexo F. Medias de dispersión de la frontera Pareto	166
Anexo G. Herramienta de optimización	167

INTRODUCCIÓN

Los centros de servicio con sistemas de atención en red viven día a día problemas de saturación, ineficiencia y falta de planeación en el diseño adecuado de su modelo de atención. Tanto su concepción como rediseño implica un estudio basado en medidas de desempeño que se ven afectadas constantemente por muchas variables dentro y fuera del sistema.

Cada una de estas variables impacta de alguna forma el modo como el sistema se comporta y cómo éste es capaz de satisfacer las necesidades del cliente eficientemente. Especialmente el recurso humano es el factor clave en un buen diseño, ya que tiene mayor impacto en las medidas de evaluación y por lo tanto representan el mejor punto de partida para aplicar procesos de mejora, su estudio y análisis solo es posible mediante pruebas estadísticas que evidencien su efecto en los indicadores alrededor de las interacciones con los demás factores que inciden en un modelo de colas.

Este estudio se encarga de realizar dicho análisis combinando diferentes tipos de factores como tasas de atención y llegada, distribuciones de probabilidad, longitud de una de una red en serie, número de servidores y capacidad del sistema para modelar problemas de colas en diferentes aplicaciones y con resultados significativamente diferentes.

Además a partir del análisis obtenido, es posible construir un algoritmo de optimización que involucre el conocimiento de los factores y su influencia en los indicadores, de tal manera que sea inteligente en la escogencia del mejor conjunto de soluciones para el problema, es decir, la cantidad de servidores y la capacidad necesaria para satisfacer las necesidades del cliente. Las técnicas de optimización multi-objetivo resultan muy útiles al momento de minimizar dos objetivos opuestos en donde no existen parámetros que optimicen ambos objetivos simultáneamente.

La inclusión de técnicas conocidas en la investigación de operaciones a un algoritmo de optimización concebido desde la teoría de juegos, desarrollado con procesos de búsqueda local y evaluado mediante principios económicos de igualdad, brinda una herramienta de análisis poco tradicional e innovadora con diferentes conclusiones acerca del estado óptimo de un sistema.

De esta manera se concibe un estudio de investigación y análisis de redes de colas que brinde criterios de mejora y eficiencia en la asignación eficiente del recurso humano.

1 PLANTEAMIENTO DEL PROBLEMA

1.1 ANTECEDENTES

El aumento significativo de la demanda de servicios crea al interior de las organizaciones un crecimiento y mejoramiento como parte dinámica de su evolución corporativa; así mismo, el aumento en sus utilidades ofrece la posibilidad de invertir en ellas mismas con el ánimo de ampliar y expandir su influencia en el mercado al que pertenecen. A pesar de esto, los problemas por la inadecuada planeación de los recursos en pro de satisfacer las necesidades de sus clientes continúan y de hecho se intensifican conforme va creciendo la organización. Situaciones de espera y congestiones se presentan en la medida en que la oferta del servicio es inferior respecto a la demanda del mismo y se crean las denominadas colas, que además de molestas, restan puntos a la calidad percibida del servicio¹.

El estudio de las líneas de espera o colas inicia en 1909 cuando A. K. Erlang, ingeniero telefonista danés, analizó los sistemas telefónicos de tráfico en donde las llamadas esperaban para ser conmutadas a su destino², pero dado el crecimiento del servicio, la espera para ser comunicado aumentaba y en ocasiones era imposible completar la operación. A partir de esto, Erlang formalizó una investigación que después sería conocida como teoría de colas con el fin de analizar la espera y los factores que intervenían en ella.

A pesar de que la idea de la teoría de colas surgió de una red, solo fue concebida, trabajada y determinada como tal en los estudios realizados por Jackson³ cincuenta años después, en donde se analizaron las redes abiertas y cerradas* de colas y se evidenció que las tasas de llegada que siguen la propiedad markoviana** son independientes del estado de la red. Bajo una demostración matemática el estudio concluyó que dada esta

¹LOVELOCK, Christopher H. Mercadotecnia de servicios. Primera edición. México: Prentice-Hall Hispanoamericana, 1997. p. 217-222.

²NAHMIAS, Steven. Análisis de la producción y las operaciones. Primera edición. México: CECSA, 1999. p. 425.

³JACKSON, J. R. Networks of waiting lines. *En*: Operation Research. Vol. 5 (1957); p. 518–521.

* Las redes son abiertas cuando existen llegadas de entidades al sistema desde el exterior de la red. En las redes cerradas los nodos forman un ciclo que nunca permite el flujo de entidades al exterior de la misma.

** Es equivalente a establecer una probabilidad condicional de cualquier "evento" futuro dados cualquier "evento" pasado y el estado actual $X_i = i$, es independiente del evento pasado y sólo depende del estado actual del proceso.

propiedad era posible analizar la red como un conjunto de sistemas de espera individuales e independientes de la tasa de llegada⁴.

Este hallazgo impactó positivamente el desarrollo teórico de las redes de espera permitiendo analizar sistemas más complejos que reflejaban la verdadera estructura de los sistemas y se ajustaban en una mejor medida a la realidad. En 1962, Mazumdar, Mason y Douligeris⁵ utilizaron las redes de Jackson para representar los modelos de telecomunicaciones y determinar la capacidad máxima de estos sistemas antes de que colapsaran. A su vez, P. J. Burke⁶ en 1976 estudió el comportamiento de colas M/M/1 al interior de una red de Jackson y analizó cómo las salidas de un proceso a otro se comportaban equivalentemente con la distribución de llegada cuyas llegadas se ajustaban a distribuciones de probabilidad clásicas.

La importancia del estudio de Jackson ha tenido un alto impacto en la teoría de colas y su aporte ha trascendido a lo largo de los años. Autores como Downb⁷ en su investigación de 1998, trabajaron con la combinación de redes abiertas y cerradas de Jackson para determinar la estabilidad de recurrencia y tránsito de un sistema en la medida en que entraban, salían o permanecían entidades al interior de éste. Otros ejemplos son los evidenciados en los trabajos realizados en 2008 por Gonga y Shouyang⁸ quienes utilizan las redes cerradas de Jackson para modelar una cadena de suministro en la que el número de servidores y estaciones debían ser determinados para hallar el equilibrio del sistema y no generar cuellos de botella.

Con lo desarrollado hasta ese punto en la teoría clásica de líneas de espera, su análisis se tornó descriptivo más que prescriptivo; en la práctica, esto significaba que dadas unas distribuciones de probabilidad de los tiempos entre llegadas y de los tiempos de servicio, las medias de desempeño se cuantificaban, pero éstas no se traducían en estrategias o decisiones relacionadas con el diseño óptimo del sistema⁹. Por este mismo hecho pareció lógico usar la teoría de colas no sólo como un método de análisis, sino como una herramienta de decisión, lo cual se hizo posible al integrar esta teoría con la optimización. Trabajos como el de J.S. Barras y D. J. Makowski¹⁰ realizado en 1985, usaron la teoría de colas para planear a largo plazo cuál sería la utilidad y productividad de un sistema de k colas; y usaron una optimización denominada μc -rule en donde minimizaban el costo

⁴DÍAZ REDONDO, Rebeca P; PAZOS ARIAS, José Juan y SUÁREZ GONZÁLEZ, Andrés. Teoría de colas y simulación de eventos discretos. Madrid: Pearson, 2003. p. 137.

⁵MAZUMDAR, Ravi; MASON, Lome G. y DOULIGERIS, Christos. Fairness in Network Optimal Control: Optimality of Product Flow Forms. En: IEEE Transactions On Communications, Vol. 39 (1991); p. 775 – 776.

⁶BURKE, P. J. Proof of a Conjecture on the Interarrival-Time Distribution in an MIMI1 Queue with Feedback. En: IEEE Transactions On Communications, Vol. 29 (1976); p. 575-589.

⁷DOWNB, Bonalda D. Stability of mixed generalized Jackson networks. En: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Vol. 7 (1998); p. 24 – 56.

⁸GONGA, Qiguo y SHOUYANG, Laib. Supply chain networks: Closed Jackson network models and properties. En: Int. J. Production Economics, Vol. 113 (2008); p. 567–574.

⁹NAHMIAS, Steven. Op. cit., p. 481- 497.

¹⁰BARRAS, J. S. y MAKOWSKI, D. J. K competing queues with geometric service requirements and linear costs: the μc - rule is always optimal. En: System and control letters. Vol. 6 (1985); p. 173 – 180.

para encontrar el número óptimo de servidores al variarlos a lo largo del tiempo. Por su parte, Ann Van Mieghem¹¹ en su aplicación orientada a los servicios, se concentró en el uso de la teoría de colas como insumo de la misma optimización de costos para determinar si se debía o no aumentar la capacidad del sistema, pues se sabía que independientemente del servicio, las colas resultaban más costosas en términos de calidad si en ellas había personas que necesitaban atención.

Algunas aplicaciones del campo empresarial, además de necesitar ser optimizadas, presentan una compleja estructura y configuración que no permite darles solución haciendo uso de métodos matemáticos tradicionales hasta aquí descritos. La gran mayoría de redes, por ejemplo, son sistemas complicados de difícil manejo, cuyas distribuciones de probabilidad usadas para modelar ciertos fenómenos no necesariamente siguen un comportamiento exponencial; o en donde ciertas restricciones, como la capacidad, hacen imposible su resolución matemática. Es allí donde surge y se integra la simulación como herramienta para dar soluciones a estas problemáticas y complementar el desarrollo teórico con el práctico.

Además, el perfeccionamiento y el uso de los computadores dieron un impulso decisivo al desarrollo y al empleo de la simulación. Jamhoura, Pennaa, Nabhenb y Pujolle¹² encontraron en la simulación la respuesta de cómo reducir las colas de datos en las redes informáticas. En este estudio se hizo la predicción de muchos escenarios del sistema para posteriormente comparar los resultados y encontrar el más eficiente. Este es uno de los tantos ejemplos en donde el híbrido optimización/simulación fue fundamental para el tratamiento de problemas de colas que hasta ese entonces por su dificultad en la modelación matemática no habían sido tratados. Casos como la investigación de Hyun Jeung, Chang y Taioun sobre proveedores de servicios logísticos (3PLs)¹³, muestra cómo a través de un modelo de optimización estos contratistas determinan la estructura de su red de atención en la prestación de servicios (mejora de tiempos, de distribución y de almacenamiento) a empresas con altos estándares de calidad. Posteriormente, en este mismo estudio 3PLs usa un modelo de simulación con el fin predecir el comportamiento y las medidas de desempeño del sistema de alguno de sus clientes.

Trabajos como el de Vanda, Felici e Impelluso¹⁴, exploran nuevas aplicaciones en cuanto al diseño y la gestión de un centro de salud. En este estudio un conjunto de servicios son requeridos por una o más categorías de usuarios y la definición de la cantidad de servidores asignados a cada uno de los servicios afecta la eficacia global del sistema y

¹¹ VAN MIEGHEM, Ann. Price and Service Discrimination in Queuing Systems: Incentive Compatibility of Gc Scheduling. En: Management Science, Vol. 46 (2000); p. 1249 – 1267.

¹² JAMHOURA, Edgard; PENNA, Manoel; NABHEN, Ricardo y PUJOLLE, Guy. Modeling a multi-queue network node with a fuzzy predictor. En: Fuzzy sets and systems. (2007); p. 1 – 12.

¹³ KO, Hyun Jeung; KO, Chang Seong y TAIOUN Kim. A hybrid optimization/simulation approach for a distribution network design of 3PLS. En: Computers & Industrial Engineering. Vol. 50 (2006); p. 440–449.

¹⁴ VANDA, Angelis; FELICI, Giovanni y IMPELLUSO, Paolo. Integrating simulation and optimization in health care centre management. En: European Journal of Operational Research. Vol. 150 (2003); p. 101–114.

sus indicadores. El principal aporte de la investigación es la metodología interactiva que utiliza la simulación para estimar la función objetivo que permite calcular y validar la configuración óptima de servidores. El desarrollo de esta herramienta sirve de apoyo a los administradores de la salud en la toma de decisiones para no dejar al azar el incremento en la cantidad de servidores y poder determinar los efectos que ello tiene en el desempeño del sistema y en el plan presupuestario.

Por su parte, M. Kuroda y A. Kawada¹⁵ consideraron la capacidad de un sistema generalizado de ventas que con ayuda de optimización logró minimizar el tiempo promedio de espera, y con la simulación probó la efectividad de las medidas tomadas bajo horizontes en donde la demanda o las tasas de llegada al sistema variaban inesperadamente.

Sobre este último tema, también se han realizado estudios en donde el comportamiento de las llegadas en la prestación de servicios no sigue un patrón homogéneo como el analizado hasta el momento, o que por el contrario presentan cambios en las llegadas, bajo distribuciones no necesariamente exponenciales¹⁶. En 2007 Jeongsim Kim y Bara Kim¹⁷ realizaron una investigación en sistemas con un único servidor en donde las llegadas estaban gobernadas por variaciones regulares de sus tasas. En estos sistemas, se evidenció cómo el tamaño de las colas aumentaba rápidamente en algunos periodos de tiempo mientras que en otros no se generaba ninguna espera. A su vez, Feldman, Mandelbaum, Massey y Whitt¹⁸ abordaron en 2008 la variación de las tasas de llegada a un *call center* y desarrollaron un método, vía simulación flexible basada en un *iterative-staffing algorithm*, para determinar los niveles apropiados de atención. El mayor aporte de esta investigación es el hallazgo de un punto estable de atención, a pesar de la incertidumbre sobre las tasas de llegada.

El modelamiento de sistemas a través de redes de colas y su solución por medio de herramientas de optimización y/o simulación es utilizado en diferentes campos de aplicación como control de sistemas hidroeléctricos¹⁹, diseño de cadenas de abastecimiento²⁰, optimización del tráfico vehicular²¹, estimación de ventas en tiendas²²,

¹⁵ KURODA M. y KAWADA A. Adaptive input control for job-shop type production systems with varying demands using inverse queueing network analysis. En: Int. J. Production Economics. Vol. 41 (1995); 217-225.

¹⁶ COCHRANA, Jeffery K. y ROCHEB, Kevin T. A multi-class queueing network analysis methodology for improving hospital emergency department performance. En: Computers & Operations Research. Vol. 36 (2009) p. 1497 – 1512.

¹⁷ KIM, Jeongsim y KIM, Bara. Regularly varying tails in a queue with discrete autoregressive arrivals of order p. En: Queueing Syst. Vol. 56 (2007); p. 93–102.

¹⁸ FELDMAN, Zohar; MANDELBAUM, Avishai; MASSEY, William A. y WHITT, Ward. Staffing of Time-Varying Queues to Achieve Time-Stable Performance. En: Management Science. Vol. 54 (2008); p. 324–338.

¹⁹ DEMBO, Ron S.; MULVEY, John M. y ZENIOS, Stavros A. Large-Scale Nonlinear Network Models And Their Application. En: Operations Research. Vol. 3 (1989); p. 353.

²⁰ KERBACHE, Laoucine y MACGREGOR, James. Queueing networks and the topological design of supply chain systems. En: J. Production Economics. Vol. 91 (2004); p. 251–272.

²¹ MERCHANT, D. K. y NEMHAUSER, G. L. A model and an algorithm for the dynamic traffic assignment problem. En: trans. Sci. (1978) p. 183-193.

control de tráfico aéreo²³, evaluación de desempeño en centros de servicio, entre otros; en donde se tienen estructuras complejas que requieren decisiones estratégicas que mejoren el desempeño del sistema y que impacten positivamente la calidad percibida del servicio. Estas aplicaciones industriales son algunos de los ejemplos utilizados por Dembo, Mulvey y Zenios como prototipos o modelos de redes generalizadas que por su versatilidad y flexibilidad podrían usarse para optimizar tiempos, costos, y en general, aumentar la productividad de diferentes aplicaciones empresariales²⁴. No obstante, aun no se han involucrado dentro del estudio de dichas aplicaciones variaciones en los procesos de llegadas, las cuales lejos de ser siempre aleatorios, pueden presentar fluctuaciones cíclicas, como lo hace la demanda por ejemplo, a lo largo del tiempo²⁵.

Además de los aspectos ya mencionados, para la optimización de aplicaciones empresariales también es necesario considerar la asignación eficiente de los recursos que agregan valor dentro de dichos procesos, pues este factor impacta el desempeño y la productividad de los sistemas. Relacionado con este aspecto, Hillier y So²⁶ analizaron el comportamiento de una red en serie con el fin de asignar eficientemente los recursos disponibles en el proceso. El objetivo de la asignación óptima de los servidores al interior de la red productiva era minimizar los tiempos y los costos de procesamiento de las piezas en la línea de fabricación. La metodología utilizada para tal fin partía de una asignación inicial, y sobre ésta modificaba la configuración de la fuerza de trabajo en cada nodo hasta encontrar aquella combinación en donde el incremento de una máquina mejoraba significativamente las medidas de desempeño establecidas para el sistema. A este procedimiento se le denominó *L-phenomenon*.

Otros autores interesados en minimizar los tiempos de espera al interior de un sistema, aplicaron modelos de optimización no lineal para realizar la asignación eficiente de servidores. En el caso de Berman, Larson y Chiu²⁷, basados en algunas propiedades de localización de servidores, buscaron el uso óptimo de los recursos en función del tipo de demanda: el tipo de demanda del servicio determinaba una asignación diferente de los servidores disponibles (localización y cantidad).

²² THOMPSON, G. L. y ZAWACK, D. J. A problem expanding parametric programming method for solving the job shop scheduling problem. En: Operation Research. (1986); p. 327-342.

²³ VELASCO, Enio Edgardo. Queueing theoretic approaches for the analysis and synthesis of air traffic control problems. En: Operation Research. Vol. 23 (1989); p. 336-337.

²⁴ DEMBO, Ron S.; MULVEY, John M. y ZENIOS, Stavros A. Op. cit., p. 357.

²⁵ GREEN, Linda y KOLESAR, Peter. The Pointwise Stationary Approximation for Queues with Non stationary Arrivals. En: Management Science; Vol.37 (1991); p. 84.

²⁶ HILLIER, Frederick S. y SO, Kut C. On the Simultaneous Optimization of Server and Work Allocations in Production Line Systems with Variable Processing Times. En: Operations Research, Vol. 44, No. 3 (1996); p. 435-443.

²⁷ BERMAN, Oded; LARSON, Richard C. y CHIU, Samuel S. Optimal Server Location on a Network Operating as an M/G/1 Queue. En: Operations Research, Vol. 33, No. 4 (1985); p. 746-771.

Por su parte, Marianov y Serra²⁸ en estudios más recientes analizaron una alternativa de cómo distribuir servidores de diferente tipo (p.e. oficinistas, cajeros y máquinas, entre otros) en redes de servicio. El principal objetivo del estudio era reducir los efectos de las colas sobre los usuarios, es decir, minimizar sus tiempos de espera en aplicaciones empresariales como bancos, hospitales y ventas de boletería. Para la definición de la óptima distribución se planteó el uso de un nuevo heurístico llamado *maximal covering*, con el cual, además de resolver el problema inicial planteado, se evidenció la importancia de la cantidad de servidores asignados en los primeros nodos de la red.

1.2 FORMULACIÓN DEL PROBLEMA

Tras analizar el comportamiento y el desempeño de redes de colas en procesos de servicio, ¿cómo minimizar los costos asociados con la espera y los tiempos de permanencia en el sistema al asignar eficientemente el recurso humano en organizaciones con estructuras de este tipo?

²⁸ MARIANOV, Vladimir y SERRA, Daniel. Location–Allocation of Multiple-Server Service Centers with Constrained Queues or Waiting Times. En: Operations Research. Vol. 111 (2002); p. 35–50.

2 JUSTIFICACIÓN DEL PROYECTO

El sector servicios cada día se consolida como uno de los ejes de desarrollo más importantes de la economía nacional²⁹, lo cual se evidencia en su aporte al Producto Interno Bruto (PIB) y en la generación de empleos en el mercado laboral, superando a cualquier otro sector de la economía colombiana³⁰. Estadísticas del DANE muestran que el sector en 1980 representaba el 47% del PIB, participación que subió a 59% en el 2007³¹ y que se mantuvo en el tercer trimestre de 2008 con un crecimiento de 20,1% con respecto al periodo anterior³². El crecimiento del sector impacta positivamente las contrataciones realizadas, lo cual se evidencia en el incremento de 7,8% durante el primer trimestre de 2008 frente a enero-marzo de 2007³³.

A la luz de estas cifras, es posible inferir que el sector está en uno de sus mejores momentos. Es vital analizar su comportamiento con el ánimo de generar ideas innovadoras que permitan mantener/potencializar su crecimiento, generar bienestar a la comunidad y consolidar la economía Colombiana³⁴. Bancos, hospitales, centros de atención al cliente y otras organizaciones que deben enfocar su atención en puntos como el desarrollo y diseño del producto, la estructura de precios, las estrategias de promoción y la definición de plazas y canales de distribución. Pero adicional a esto, las empresas de servicios deben enfocarse con mayor atención en el personal, en el proceso como tal y en la evidencia física del servicio, según lo expuesto por el marketing de servicios en el plan de mercadeo * extendido³⁵.

Sin embargo existe un problema común que impide la gestión eficiente de una organización a través de su talento humano, sus procesos y la evidencia física del servicio y son precisamente las colas. De acuerdo con Michael Fortino, de Priority Management en Pittsburgh, la mayoría de las personas pasan alrededor de cinco años de su vida en

²⁹ PORTAFOLIO. Disponible en internet:
<http://www.portafolio.com.co/economia/economiahoy/2008-11-24/ARTICULO-WEB-NOTA_INTERIOR_PORTA-468314>

³⁰ NOGUEIRA COBRA, Marcos Henrique. Marketing de servicios: estrategias para turismo, finanzas, salud y comunicación. Segunda edición. Bogotá; Buenos Aires: McGraw Hill, 2000. p. 3.

³¹ PORTAFOLIO. Op. cit. p.1.

³² DANE. Disponible en internet:

<http://www.dane.gov.co/index.php?option=com_content&task=category§ionid=15&id=54&Itemid=236>

³³ PORTAFOLIO. Op. cit.p.1.

³⁴ Ibid. p.1.

* Estrategia del Marketing que utiliza un paquete (mezcla) que no sólo satisfaga las necesidades de los clientes dentro de los mercados objetivo, sino que también maximice, simultáneamente, el desempeño de la organización.

³⁵ COHEN, William A. Plan de mercadotecnia. Primera edición. México: Compañía Editorial Continental: CECSA, 2004. p. 58 – 67.

filas de espera³⁶, fenómeno que se vuelve más frecuente en la medida que la demanda por un servicio se incrementa más que proporcionalmente con respecto a su capacidad para ofrecerlo, causando tedio, incomodidad, insatisfacción y sobrecostos³⁷. Ninguna organización es ajena a este problema, de hecho toda solicitud, pago, asesoría, reclamación o adquisición de productos son sólo ejemplos de cómo a diario consumidores deben esperar formando cola frente a un servicio.

En aeropuertos como Atlanta y Chicago, por ejemplo en 1995 los costos para las aerolíneas por las colas fueron cuantificados alrededor de 6 a 10 dólares por minuto, que acumulados en un año representan 10 millones de dólares³⁸, además de muchos otros costos intangibles y ocultos como el deterioro de la imagen corporativa y la calidad del servicio que son difíciles de dimensionar y cuantificar. El tiempo que la población de un país pierde al esperar en las colas es un factor importante tanto en la calidad de vida como en la eficiencia de su economía³⁹, pues al invertir los tiempos muertos de espera en actividades productivas se podrían generar beneficios económicos para el país y reducción de costos en las organizacionales. En potencias mundiales como Estados Unidos, en donde se ha estimado que toda su población pasa 37 mil millones de horas al año en líneas de espera, la eliminación de las colas significaría tener cerca de 20 millones de personas más trabajando al año.

A nivel mundial las empresas están pagando más en costos de personal debido a los incrementos en el valor de la mano y al exceso de personal en Europa, Estados Unidos, Japón y especialmente en economías en desarrollo como las latinoamericanas⁴⁰. De acuerdo a esto, saber con certeza cuántos empleados deben ser asignados a una labor para poder cumplir con los estándares de servicio se vuelve un factor primordial. Esta planeación debe hacerse teniendo en cuenta principalmente al cliente, pues es el principal generador de utilidades y es hacia quien se orienta la compañía. En las últimas décadas el servicio al cliente toma cada vez más importancia en las organizaciones y se considera como un factor de diferenciación que permite desarrollar ventajas competitivas sostenibles⁴¹. Con el conocimiento de la cantidad de recursos necesarios en los sistemas, las filas ya no serán más un obstáculo entre el cliente y las organizaciones, la calidad podrá aumentar, los costos diez veces menores de mantener a un cliente que atraerlo reducirán y se generará riqueza, empleo y satisfacción al entorno de las organizaciones.

³⁶ NAHMIAS, Steven. Op. cit., p. 453.

³⁷ LOVELOCK, Christopher H. Mercadotecnia de servicios. Primera edición. México: Prentice-Hall Hispanoamericana, 1997. p. 217-222.

³⁸ POTOMAC. New Study May Eliminate Queuing Costs. En: Airline Financial News. (Oct 21, 1996).

³⁹ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Investigación de operaciones. Séptima edición. México: McGraw-Hill, 2002. p. 880 - 889.

⁴⁰ Fondo Monetario Internacional. Disponible en internet:
<<http://www.docstoc.com/docs/1014168/Vuelve-Japn---Daniel-Citrin-y-Alexander-Wolfson---Finanzas-y-Desarrollo---Junio-de-2006>>

⁴¹ Universidad de los Andes. Disponible en internet:

< https://ctp.uniandes.edu.co/Empresas/Servicios/Articulos/el_papel_del_recurso_human.php>

Luego, es fundamental gestionar las filas para garantizar entre otras cosas, la calidad y el bienestar que como organización debe proporcionar a sus clientes y empleados con el ánimo de generar productividad, maximizar utilidades y minimizar costos. Es entonces en ese punto donde la teoría de colas entra a jugar un papel determinante en el análisis de los sistemas, en la medición de su desempeño y en la cuantificación de los costos asociados con su operación. Su uso además es necesario debido a la incertidumbre en los sistemas, la cual se potencializa aun mas al involucrar humanos a las operaciones allí realizadas. El comportamiento humano hace que el estado de un sistema varíe de forma no muy predecible, pero que a pesar de esto, es posible regir ese comportamiento por leyes estadísticas, que de una forma clara o aproximada, se encuadran en distribuciones probabilísticas⁴².

Adicional a esto, surge una dificultad que radica en la complejidad matemática que todo lo anterior sugiere, pues los sistemas a analizar deben reflejar la situación real y la operación diaria de organizaciones compuestas por múltiples puntos de atención y diferentes configuraciones. Hasta el 70% de los problemas de programación matemática en el sector empresarial se pueden representar como modelos relacionados con redes⁴³, haciéndose necesario desarrollar herramientas que extiendan la aplicabilidad de la teoría de colas a estos ámbitos empresariales. Al analizar sistemas modelados con redes es posible tener una visión más amplia del interior del sistema y tener en cuenta la sinergia* que existe entre subsistemas. Los resultados del análisis de un modelo de redes tendrán la exactitud necesaria de un problema real sin caer en el error de la aproximación en donde no es posible vislumbrar, controlar y evaluar un sistema de manera eficiente⁴⁴.

Por ello, es en las redes y sus colas en donde deben centrarse esfuerzos considerables. Más sin embargo, se debe ir más allá de la utilidad misma de la teoría de colas, al no sólo medir desempeño, sino al buscar mejoras usando herramientas de optimización que permitan calcular medidas eficientes de recursos, costos, utilidades, tiempos y otros indicadores propios de cada organización. Adicionalmente, se hace necesario el uso de otra herramienta de análisis como la simulación: una herramienta poderosa, flexible e intuitiva que es sumamente útil cuando se dificulta o imposibilita la resolución del modelo analítico o numérico requerido.⁴⁵ Este híbrido permitirá analizar sistemas complejos en donde se presentan diferentes tasas de llegada y de servicio, y cambios en la cantidad de servidores y en los recursos disponibles de un sistema.

⁴² ESCUDERO, Laureano F. Aplicaciones de la teoría de colas: análisis y solución de los problemas de cuellos de botellas en almacén, producción y ventas con referencias breves a Business Games y teoría de juegos. Primera edición. Bilbao: Deusto, 1972. p. 33 – 34.

⁴³ NAHMIAS, Steven. Op. cit., p. 453.

* Sinergia: la acción de dos o más sustancias tienen un efecto mayor y diferente que la suma individual de ellas.

⁴⁴ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 880 – 889.

⁴⁵ PRAWDA WITENBERG, Juan. Métodos y modelos de investigación de operaciones. México: Limusa, 1976. p. 245 – 256.

Lo expuesto anteriormente evidencia la necesidad en las organizaciones de analizar, conocer y explorar sus sistemas de atención en donde se generan líneas de espera, con el fin de mermar el impacto que ellas tienen sobre su productividad. Las redes de colas y la visión más amplia de los sistemas, junto con la mayor aproximación de las distribuciones no exponenciales y las nuevas metodologías de solución, son variaciones que hacen que este estudio sea de gran utilidad al sector servicios, aproximándose cada vez más a su situación real. Además, la realización de este estudio desde la perspectiva multidisciplinar que brinda la ingeniería industrial, hace que las soluciones abarquen aspectos tecnológicos, organizacionales, comerciales, económicos, financieros y operativos que son indispensables para alcanzar la rentabilidad y la productividad que señala Richard C. Vaughn como el pilar de esta carrera.⁴⁶

Adicional a los beneficios percibidos por las organizaciones, este desarrollo será un aporte investigativo a la teoría de colas, al trabajar, analizar y comprender sistemas complejos de redes que pueden usarse en un sinnúmero de aplicaciones empresariales, dadas unas variables y una metodología estandarizada de solución.

⁴⁶ VAUGHN, Richard C. Introducción a la ingeniería industrial. Segunda edición. Barcelona: reverte,1993. p. 30-32.

3 OBJETIVOS

3.1 OBJETIVO GENERAL

Analizar el comportamiento y desempeño de redes de colas en procesos de servicio para que minimizando los costos asociados con la espera y los tiempos de permanencia en el sistema se asigne eficientemente el recurso humano en organizaciones con estructuras de este tipo.

3.2 OBJETIVOS ESPECÍFICOS

- Definir una red de colas sobre la cual sea posible realizar modificaciones de ciertos factores de las líneas de espera, como tasas de llegada al servicio, distribuciones de probabilidad de los tiempos entre llegadas, cantidad de servidores disponibles y carga de trabajo, entre otros.
- Determinar vía simulación el desempeño de los posibles sistemas que se pueden generar en la red de colas al combinar los tratamientos definidos en el experimento diseñado y evaluar cómo la variación de dichos parámetros en el sistema impacta su eficiencia.
- Desarrollar un modelo de optimización multi-objetivo para minimizar los tiempos de espera en la red y los costos asociados con dicha espera al asignar eficientemente los recursos mediante la aplicación de algoritmos heurísticos.
- Construir una herramienta computacional orientada a organizaciones del sector servicios que determine el desempeño del sistema a partir de los resultados de la simulación y asigne el recurso en la red a partir de la solución del modelo de optimización.

4 MARCO TEÓRICO

4.1 PROCESOS ESTOCÁSTICOS

Un proceso estocástico se define como una colección indexada de variables aleatorias $\{X_t\}$, donde el índice t toma valores de un conjunto T dado. Con frecuencia T se toma como el conjunto de enteros no negativos y X_t representa una característica de interés medible en el tiempo t . Por ejemplo, X_t puede representar el número de personas en un sistema al final del periodo t .⁴⁷

El interés de los procesos estocásticos es describir el comportamiento de un sistema en operación durante algunos periodos. Para saber su comportamiento, existen muchas aplicaciones y es usado también como sustento para muchas teorías probabilísticas una de ellas es la teoría de colas.

4.2 TEORÍA DE COLAS

Muchas industrias de servicios tienen un sistema de colas, en el que los productos o clientes llegan a una estación y esperan en una fila, obtienen algún tipo de servicio y luego salen del sistema⁴⁸. Este tipo de fenómeno se origina cuando los usuarios de un determinado servicio llegan con mayor rapidez a la que este tiene capacidad de despachar, y por tal motivo se acumula personas u objetos que deben esperar para ser atendidos.⁴⁹

La teoría de colas es el estudio de dicha espera en las sus diferentes formas, usando los modelos de colas se representan los sistemas de líneas de espera que surgen en la práctica. Las formulas para cada modelo indican cual debería ser desempeño del sistema correspondiente y señalan la cantidad esperada de tiempo y personas en una cola, en una gama de circunstancias.⁵⁰

El carácter aleatorio de los fenómenos de colas implica evidentemente que el arma principal de este análisis es el cálculo de probabilidades. El conjunto de formulaciones y relaciones que usa a los datos a una determinada distribución probabilística constituye un

⁴⁷ ESCUDERO, Laureano F. Op. cit., p. 33.

⁴⁸ KAMLESH, Mathur y SOLOW, Daniel. Investigación de operaciones: el arte de la toma de decisiones. Primera edición. México; New York: Prentice-Hall Hispanoamericana, 1996. p. 710 -711.

⁴⁹ DÍAZ REDONDO, Rebeca P; PAZOS ARIAS, José Juan y SUÁREZ GONZÁLEZ, Andrés. Op. cit., p. 135 – 139

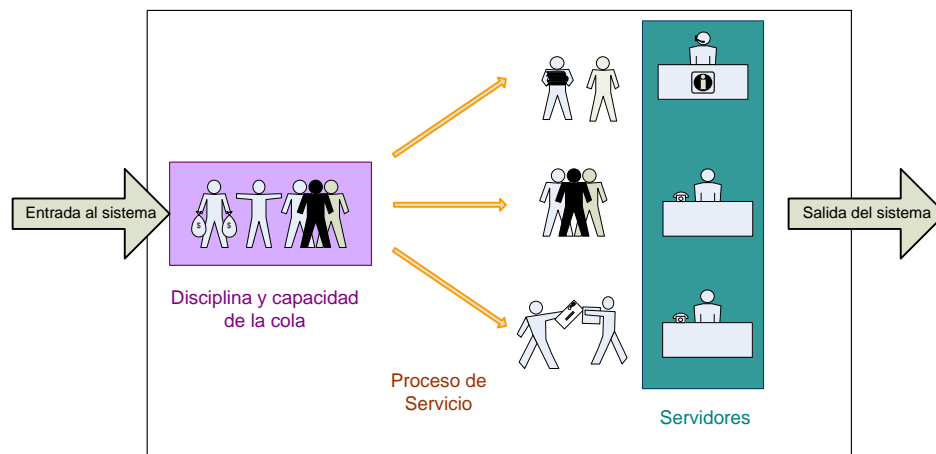
⁵⁰ GROSS, Donald. y HARRIS, Carl M. Fundamentals of queuing theory. Tercera edición. New York ; Chichester : John Wiley, 1998. p. 20- 34.

modelo matemático que se denomina proceso estocástico en el cual una o varias magnitudes varían en forma aleatoria en función del tiempo.⁵¹ Estructura básica de los modelos de colas

Una línea de espera está constituida por un cliente que requiere de un servicio que es proporcionado por un servidor en un determinado periodo. Los clientes entran aleatoriamente al sistema y forman una o varias colas para ser atendidos. Si el servidor está desocupado, de acuerdo a ciertas reglas preestablecidas con el nombre de disciplina del servicio, se proporciona el servicio a los elementos de la cola en un periodo determinado de tiempo, llamado tiempo de servicio y luego abandonan el sistema⁵². Las líneas de espera se pueden clasificar de acuerdo a:⁵³

- a) El número de clientes que pueden esperar en la cola y población.
- b) Intervalo de tiempo transcurrido entre un cliente y otro.
- c) El tiempo de servicio.
- d) La disciplina de la cola. (Peps, Ueps, etc.)
- e) El número de servidores.
- f) La estructura de las estaciones de servicio.
- g) La estabilidad del sistema.

Figura 1. Estructura de sistemas de colas



Fuente: Elaboración propia.

⁵¹ ESCUDERO, Laureano F. Op. cit., p. 33.

⁵² KAMLESH, Mathur y SOLOW, Daniel. Op. cit., p. 710.

⁵³ Ibid., p. 711.

4.2.1 Sistema de clasificación

El sistema de clasificación más conocido es el de Kendall y Lee que en 1953 elaboraron con el fin de estandarizar características similares en algunos modelos y de esta forma analizarlos por igual. Esto no solo es benéfico en términos de la notación sino que abre la posibilidad de asociar cierto tipo de fórmulas ya calculadas a sistemas con características similares en el caso de las distribuciones exponenciales⁵⁴.

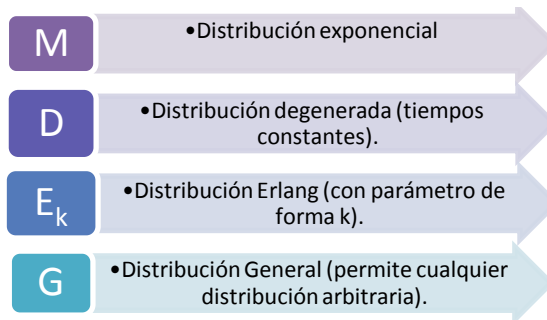
Figura 2. Clasificación Kendall y Lee



Fuente: Elaboración propia.

1. Distribución de probabilidad del tiempo entre llegadas de las entidades
2. Distribución de probabilidad del tiempo de servicio
3. Número de servidores
4. Disciplina de atención
5. Número máximo de entidades que soporta el sistema en un periodo t
6. Población

Figura 3. Tipos de distribuciones



Fuente: Elaboración propia.

Por ejemplo un sistema $M/M/1$ significa unos tiempos de llegada y atención con distribución exponencial y un solo servidor atendiendo dicha colas; como no se especifican las demás características se asume la disciplina es general y los otros parámetros infinitos. En el caso de un $G/M/S$ estaríamos hablando entonces de una llegada con una distribución general, atención exponencial y S servidores. Debido al

⁵⁴ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 840 – 845.

complejo manejo de las distribuciones no exponenciales en las colas para encontrar los valores a las incógnitas es necesaria la simulación⁵⁵.

4.2.2 Medidas de desempeño

En la revisión analítica de las colas hay dos tipos de medidas de desempeño que son las más importantes y usadas, las dos medidas por lo común se expresan en términos de sus valores esperados y son:

L = número esperado de clientes en el sistema

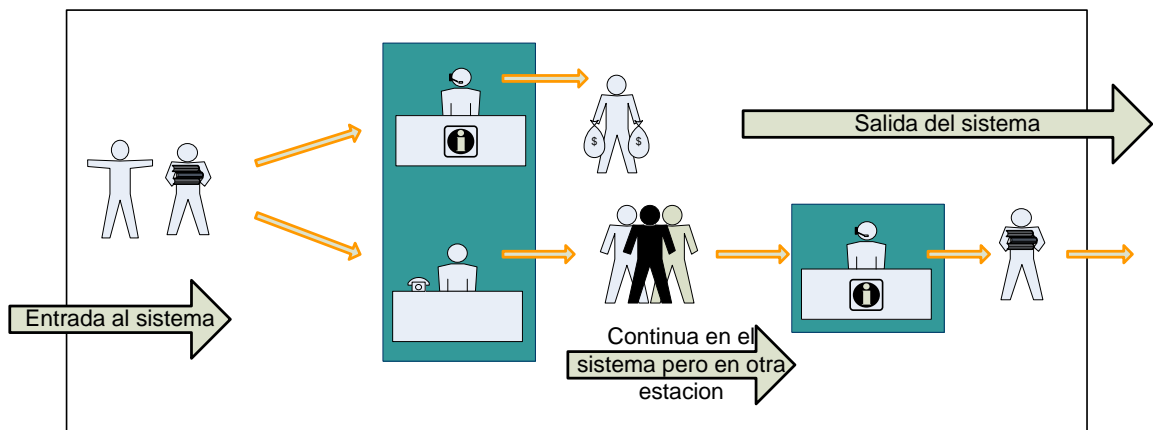
W = tiempo de espera esperado en el sistema

Ahora, si no se cuenta a las personas que se atienden sino solo las de la cola se tiene las mismas dos fórmulas pero con subíndice q (L_q, W_q)⁵⁶

4.3 REDES DE COLAS

Hasta ahora se han tomado en cuenta nada más los sistemas de colas que tienen una instalación de servicio con uno o más servidores, pero los sistemas de colas que se encuentran en los estudios de investigación de operaciones, en realidad a veces son redes de colas, es decir, redes de instalaciones de servicio en las que los clientes solicitan un servicio de algunas o todas ellas.⁵⁷

Figura 4. Redes de colas



Fuente: Elaboración propia.

⁵⁵ ESCUDERO, Laureano F. Op. cit., p. 133 – 140.

⁵⁶ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Métodos cuantitativos para administración: un enfoque de modelos y casos de estudio, con hoja de cálculo. México; Buenos Aires : McGraw-Hill/Irwin, 2002. p. 695 – 698.

⁵⁷ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Métodos cuantitativos para administración: un enfoque de modelos y casos de estudio, con hoja de cálculo. México; Buenos Aires : McGraw-Hill/Irwin, 2002. p. 885.

Una red de colas es un conjunto de sistemas de colas conectados entre sí. En este tipo de estructuras los clientes llegan un nodo o punto de servicio y esperan su turno hasta ser atendidos como en un sistema de colas convencional; la diferencia es que al instante mismo de abandonar un nodo pueden o no dirigirse a otro para recibir un servicio adicional o continuar un servicio en una dependencia diferente, esto sucede indefinidamente hasta que todos los requerimientos del servicio queden atendidos.⁵⁸

4.3.1 Clasificación de las redes de colas

- **Redes abiertas:** Reciben entrada de clientes procedentes de una o varias poblaciones externas y tienen salidas procedentes hacia el exterior. Son las usadas en centros de servicio y otras más aplicaciones. Sus dos subtipos son:
 - Acíclicas: Un trabajo nunca puede volver a la misma cola.
 - Cíclicas: Hay bucles en la red.
- **Redes cerradas:** no reciben entradas de poblaciones externas ni tienen salidas al exterior. Número constante de clientes circulando dentro de la red.⁵⁹

1.1.1.1.1

Figura 5. Clasificación de las redes de colas.



Fuente: Departamento de sistemas informáticos. EPSA. Avda. de España.

⁵⁸ ODoni, Amedeo R. Disponible en internet:
< <http://web.mit.edu/aeroastro/www/people/odoni/bio.html> >

⁵⁹ JACKSON, J. R. Op. cit., p. 518–521

4.3.2 Redes de Jackson

Una red de Jackson es un sistema de m instalaciones donde la instalación i ($i=1,2,\dots,m$) que tiene:

1. Una cola de capacidad infinita.
2. Clientes que llegan de fuera del sistema de acuerdo a un proceso de entrada Poisson de parámetro a_i .
3. Un número de servidores s_i , con la misma distribución exponencial con parámetro μ_i , para los tiempos de servicio.
4. Un cliente que deja la instalación i , puede salir del sistema o bien puede ir a otra instalación j ($j=1,2,\dots,m$ y $j \neq i$), con probabilidad P_{ij} . La probabilidad de salir del sistema es:

$$q_i = 1 - \sum_{j=1}^m p_{ij} \text{ }^{60}$$

Las Redes de Jackson reciben ese nombre debido a que fue Jackson quien descubrió una propiedad que es vital para el análisis: Bajo condiciones de estado estable, cada instalación j ($j=1,2,\dots,m$) de una red, se comporta como si fuera un sistema de colas M/M/s independiente, con tasa de llegadas λ_j

$$\lambda_j = a_j + \sum_{i=1}^m \lambda_i p_{ij} \quad \text{donde, } s_j \mu_j > \lambda_j$$

4.3.2.1 Medidas de desempeño

Número medio de trabajos en el sistema, L_{red} , que es la suma de los números medios de trabajos en cada uno de los nodos:

$$L_{red} = \sum_{i=1}^K L_i$$

Tiempo medio en el sistema, W_{red} , que es el tiempo medio que pasa una tarea desde que entra en la red hasta que sale de ella:

$$W_{red} = \frac{L_{red}}{\lambda_{red}}$$

⁶⁰ JACKSON, J. R. Jobshop-like queuing systems. En: management science, Vol. 10 (1963); p. 131-142.

Propiedad de Equivalencia: Supóngase que una instalación de servicio tiene S servidores, un proceso de entrada Poisson con parámetro λ , y la misma distribución de los tiempos de servicio para cada servidor con parámetro μ , en donde $\rho < 1$. Entonces, la salida en estado estable de esta instalación de servicio también es un proceso de Poisson de media λ ⁶¹.

4.4 APLICACIONES DE TEORÍA DE COLAS

La teoría de colas ha tenido un lugar sobresaliente en las técnicas analíticas modernas, pero hasta aquí el enfoque se ha limitado a la formulación de una teoría matemática descriptiva. Así, no concierne en forma directa a la teoría de colas alcanzar la meta de la investigación de operaciones: la toma de decisiones óptimas.⁶² No obstante en un insumo necesario para este fin que puede alcanzarse con la simulación y optimización.

4.5 SIMULACIÓN

Simulación es una técnica numérica para conducir experimentos que comprenden ciertos tipos de relaciones matemáticas y lógicas, las cuales son necesarias para describir el comportamiento y la estructura de sistemas complejos del mundo real a través de largos periodos de tiempo.

Dentro de la gran variedad de paquetes de simulación disponibles en el mercado, los más conocidos y usados en la actualidad son:

- **Arena:** Simulador que se usa para modelar procesos, informar y documentar. Simular la respuesta futura del sistema de cara a comprender relaciones complejas e identificar posibilidades de mejora. Visualizar las operaciones con gráficos dinámicos animados. Analizar como el sistema funciona en su configuración “tal cual” y bajo un conjunto de posibles alternativas posibles de forma que se pueda elegir con certeza la mejor de ellas.⁶³

⁶¹ BURKE, P. J. The Output of a queuing system, *En: Operations research*, Vol. 4 (1956); p. 699-704.

⁶² HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p., c2002. P .907 – 909.

⁶³ KELTON, W. David. Simulación con software Arena. 4a ed. Publicación: México; Bogotá : McGraw-Hill Interamericana Editores, 2008. p. 25-32.

4.6 OPTIMIZACIÓN

Es el uso de modelos matemáticos, con el objetivo de brindar un conocimiento para la realización de la toma de decisiones. Busca mejorar el rendimiento de los procesos, para así determinar cómo se pueden maximizar o minimizar los beneficios.⁶⁴

4.6.1 Tipos de Optimización

Existen varios tipos de optimización usados para diferentes situaciones que por su complejidad se constituyen en un problema. Algunos de esos métodos son:

- **Programación lineal:** Procedimiento de resolución de problemas mediante formulación a través de ecuaciones lineales, optimizando la función objetivo teniendo que está restringida por ecuaciones lineales.
- **Programación entera:** Al igual que la lineal utiliza un procedimiento o algoritmo para resolver problemas. En estos modelos algunas o todas sus variables deben ser valores enteros.⁶⁵
- **Programación dinámica:** Es un enfoque general para la solución de problemas en los que es necesario tomar decisiones en etapas sucesivas. Las decisiones tomadas en una etapa condicionan la evolución futura del sistema y afectan a las situaciones en las que el sistema se encontrará en el futuro (denominadas estados), y a las decisiones que se plantearán en ese momento⁶⁶
- **Meta-Heurísticas:** Es un método para resolver un tipo de problema general, usando los parámetros dados por el usuario sobre unos procedimientos genéricos y abstractos de una manera que se espera eficiente. Las meta-heurísticas generalmente se aplican a problemas que no tienen un algoritmo o heurística específica que dé una solución satisfactoria; o bien cuando no es posible implementar ese método óptimo.

⁶⁴ HOOKER, John Norman. Integrated methods for optimization. New York: Springer, 2007. P. 20-54.

⁶⁵ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 7 – 145.

⁶⁶ Análisis Cuantitativo con WINQSB. Disponible en la web:
< <http://www.eumed.net/libros/2006c/216/1j.htm>>

Algunas meta-heurísticas de mayor utilización y por consiguiente más conocidas son:

- Optimización aleatoria
- Búsqueda local
- Búsqueda primero el mejor
- Optimización basada en colonias de hormigas
- Búsqueda Tabú
- Algoritmos Genéticos
- Algoritmos Matemáticos
- Inteligencia enjambre
- Búsqueda por difusión estocástica
- Optimización extrema

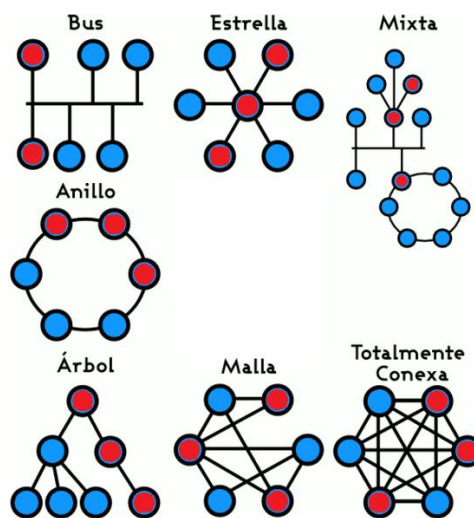
5 REDES DE COLAS EN CENTROS DE SERVICIO

Los centros de servicios han evolucionado y, hoy en día, poseen características físicas, tecnológicas y procedimentales muy específicas que van más allá de lo que la teoría de colas ha intentado modelar y explicar hasta el momento. La gran mayoría de esfuerzos y hallazgos en esta teoría se han centrado en el análisis de líneas de espera sencillas, dejando a un lado el estudio de estructuras más complejas como las redes de colas, las cuales representan de manera más acertada centros de servicio actuales. En este capítulo se diseñará un experimento que permitirá analizar y determinar cómo la variación de los factores propios de las líneas de espera impacta el comportamiento y desempeño de redes de colas en procesos de servicio.

5.1 DEFINICIÓN DE LA RED DE COLAS

Los procesos de servicio en general son estructuras complicadas cuya composición depende de los flujos de llegada o salida hacia los puntos de atención o hacia los nodos de servicio. Debido a esto, cada centro de atención dispone según su conveniencia de una distribución en forma de red para atender al cliente. Dentro de este tipo de estructuras se encuentran las redes tipo estrella, bus, cíclica, mixta, árbol, maya y totalmente conexas.

Figura 6. Tipos de redes



Fuente: miguelven.files.wordpress.com

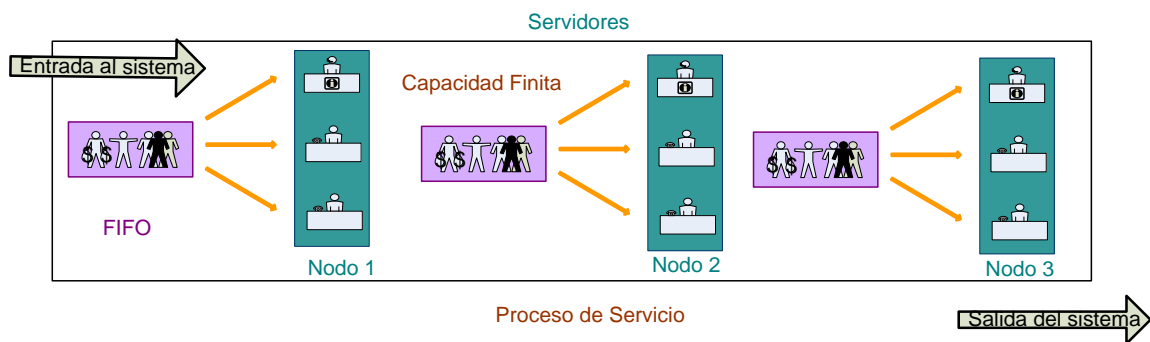
La figura 6 muestra algunas de las estructuras en red más usadas y comunes en el sector empresarial. Cada nodo o círculo azul representa un punto de atención al interior de la red; los nodos de color rojo son puntos de atención que identifican nodos secuenciales o redes en serie al interior de las demás estructuras.

Los nodos en serie son el tipo de red más simple y ocurrente; representan una base estructural dentro de todas las demás redes y por ello la importancia que cobra su uso en el presente estudio. El uso de redes en serie es la mejor manera de construir procesos consecutivos sin limitarse a casos particulares y con la posibilidad de modelar diversas estructuras de atención, mediante procesos de expansión, unión o combinación. Debido a esto, se trabajará sobre una red en serie con cantidad de nodos constantes y sin ciclos entre uno o más nodos.

La selección de red acíclica en serie permite simplificar el funcionamiento de sistemas reales mediante el fraccionamiento de todo el proceso de atención en subprocesos en serie. De acuerdo a esto, los clientes podrán entrar por un único punto, ser atendidos en cada nodo según la capacidad de los servidores dispuestos y finalmente salir del sistema.

Adicionalmente, se asume que a dicha red no llegarán entidades de forma masiva, no habrá prioridades percibidas para la atención de entidades, la disciplina de la fila será FIFO, se generará una cola diferente en cada nodo, la cantidad de servidores cambiará por nodo y cada nodo tendrá una capacidad restringida distinta.

Figura 7. Estructura de la red de colas definida



Fuente: Elaboración propia.

5.2 FACTORES DE INFLUENCIA

Según David Anderson en su libro *Métodos Cuantitativos para Negocios*⁶⁷, el comportamiento de las líneas de espera está determinado por seis variables o características principales que Kendall y Lee reunieron para identificar y clasificar los diversos modelos de colas. La distribución de probabilidad de los tiempos entre llegadas y de los tiempos de servicio, el número de servidores, la capacidad del sistema, la disciplina de la cola y el tamaño de la población o de la fuente de origen, hacen parte de dicha notación que permiten construir diferentes modelos y determinar el comportamiento de las medidas de desempeño de los sistemas mediante la variación de sus características.

5.2.1 Distribuciones de probabilidad de los tiempos entre llegadas

Las redes de colas, y en general la teoría de colas, han sido estudiadas en su mayoría bajo la distribución de probabilidad exponencial debido a su facilidad matemática. Su aplicación se volvió muy popular, mas sin embargo, con las herramientas computacionales que hoy en día existen es posible realizar estudios mucho más complejos de los que en su momento Jackson y Burke concebían, cambiando el enfoque hacia nuevas distribuciones de probabilidad para modelar los tiempos entre llegadas. Aunque en el pasado la distribución exponencial resultó muy útil, en el presente trabajo su uso significa una simplificación matemática de lo que realmente ocurre en un sistema, generando resultados incongruentes⁶⁸ en la gran mayoría de casos.

La distribución exponencial descarta la posibilidad de entradas al sistema mediante el muestreo de tiempos entre llegadas, pues es una distribución equidispersa cuya media y desviación de los tiempos de arribo es la misma, suposición que se cumple en muy pocas oportunidades en el entorno empresarial. Adicional a esto, el hecho de poseer un único parámetro no le permite modelar eficazmente el comportamiento de un sistema en operación, pues entre más parámetros contenga una distribución es más probable que el modelo pueda explicar la realidad⁶⁹.

Debido a lo expuesto anteriormente, se hace necesario el uso de otras distribuciones de probabilidad diferentes a la exponencial para modelar los tiempos entre llegadas, con el fin de suplir sus deficiencias y tener representaciones cada vez mas aproximadas a la realidad. Se deben buscar distribuciones de probabilidad más versátiles, más robustas, con más aplicabilidad al contexto de atención empresarial que es donde se necesitan.

⁶⁷ ANDERSON David R. *Métodos cuantitativos para los negocios. Soluciones empresariales.* Cengage Learning Editores, 2004. p. 618.

⁶⁸ MALHOTRA, Manish. y REIBMAN, Andrew. Selecting and implementing phase approximations for semi-markov models. *En: Stochastic Models*, Vol. 9, No. 3 (1993); p. 473-506.

⁶⁹ KECECIOGLU, Dimitri. *Reliability and Life Testing Handbook.* Prentice Hall, Inc. Volume I, New Jersey, (1993).

Recientemente, las distribuciones de probabilidad obtenidas mediante mezclas o adiciones de parámetros han mostrado ser las más apropiadas para cumplir con dicho fin, pues son sobredispersas* y presentan colas más largas que la distribución de partida, dando lugar a mejores ajustes sobre los datos obtenidos en un muestreo de tiempos.⁷⁰

Dentro de este tipo de distribuciones, la Híper-Erlang o mezcla de Erlang, se destaca por darle mayor variabilidad al tiempo entre arribos, debido a que posee mayor cantidad de parámetros. Adicional a sus parámetros, esta distribución tiene sub-fases, cada una de las cuales posee la propiedad de no memoria, lo cual aporta parte de la simplificación matemática de la distribución exponencial. Adicionalmente, ofrece la posibilidad de mostrar de mejor manera problemas de dispersión de colas con un menor tiempo de procesamiento con respecto a sus semejantes.⁷¹

La distribución Híper-Erlang resulta de la interrelación de procesos Poisson en secuencia o en fases, haciendo del comportamiento de entrada al sistema un proceso más robusto y completo. La modelación de redes de tráfico con distribuciones tipo fase ha demostrado en el pasado ser un poderoso medio para facilitar la evaluación del desempeño de un sistema de colas y llegar a conclusiones contundentes⁷².

Una de las más grandes ventajas de la distribución Híper-Erlang es que ha probado ser capaz de aproximarse a cualquier variable aleatoria positiva según datos recogidos en un muestreo⁷³, da la flexibilidad necesaria para usarse en diferentes contextos y aporta la exactitud en el montaje del comportamiento de una cola⁷⁴. Por ejemplo, Feitelson y Rudolph⁷⁵ en su trabajo "Job scheduling strategies for parallel processing" usan dicha distribución para el modelamiento de sistemas en donde hay que realizar una asignación eficiente de trabajos y es necesario que dicha asignación posea una probabilidad de ocurrencia según un proceso Poisson. El objetivo de este estudio era llegar a una asignación eficiente de los trabajos según la variabilidad de las entradas a lo largo del proceso.

Por otro lado, esta misma distribución ha sido aplicada en el modelamiento de redes de comunicación móvil en todo el mundo. Yeoy y Chi⁷⁶ identificaron procesos de llegadas variables a dichas redes que la distribución Híper-Erlang modelaba mejor que cualquier otra, debido al comportamiento impredecible de las líneas celulares estudiadas.

* Equidispersa: posee media y varianza diferente.

⁷⁰ GÓMEZ, Emilio y SARABIA, José María. La distribución binomial-exponencial truncada con aplicaciones en el sector del seguro de automóviles. En: IAE, Vol. 14, (2008); p. 3-20.

⁷¹ HONGXIA, Zhou. WANG, Junfeng. XU, Fanjiang y LI, Lei. Hyper-Erlang Based Model for Network Traffic Approximation. En: China Postdoctoral Science Foundation, (2005).

⁷² Ibid. p. 296 – 312.

⁷³ ZHANG Yan , HU Honglin , FUJISE Masayuki. Resource, mobility, and security management in wireless networks and mobile communications. CRC Press, (2006). p. 160-162

⁷⁴ WANG, Junfeng. LIU, Jin y CHUNDONG, She. Segment-based adaptive hyper-Erlang model for long-tailed network traffic approximation. En: J Supercomput, Vol. 45 (2008); p. 296 – 312.

⁷⁵ FEITELSON Dror G. , RUDOLPH Larry . Job scheduling strategies for parallel processing. En: Springer (1997). p. 298.

⁷⁶ YEO, K. M. y CHI-Hyuck Jun, "Teletraffic analysis of cellular communication systems with general mobility based on hyper-erlang characterization", Computers & Industrial Engineering, Vol. 42(2002). p. 507-520.

Bajo esta misma línea, se desarrollaron una serie de estudios a cargo de Zhang y Hu encaminados al análisis de la capacidad de una red inalámbrica para soportar la conexión masiva de ordenadores, evitando simultáneamente problemas de seguridad, lentitud y saturación⁷⁷. El modelo desarrollado por Zhang y Hu estaba basado en distribuciones Híper Erlang para garantizar un modelamiento más acertado de las entradas al sistema y realizar cálculos estimativos acerca de la necesidad de ampliar o no la cobertura de la red inalámbrica evitando problemas de saturación.⁷⁸

Finalmente, Wang y Zhou comprobaron las ventajas que posee la distribución Híper Erlang con respecto a sus semejantes, como la distribución Weibull y Pareto. En este experimento modelaron el tráfico web, caracterizado por poseer alta variabilidad en sus datos, con distribuciones Hiper-Erlang, Weibull y Pareo. El resultado principal evidenciado fue el menor tiempo de procesamiento de la Híper Erlang debido a su simplicidad y su mayor aproximación a lo visto en un sistema real de tráfico de internet.⁷⁹

La distribución Híper-Erlang es una distribución de probabilidad continua con parámetros $\lambda > 0$, $l > 0$ (número de fases), y $C \geq l$ (parámetro de forma).

Su función de densidad es:

$$f(x) = \sum_{i=1}^l \alpha_i \frac{\lambda_i^C}{\Gamma(C)} x^{C-1} e^{-\lambda_i x}$$

Donde α_i es la probabilidad de ocurrencia de cada fase ($\sum_{i=1}^l \alpha_i = 1$, $\alpha_i \geq 0$)

Su función acumulada de probabilidad es:

$$F(X) = 1 - \sum_{i=1}^l \alpha_i e^{-\lambda_i x} \sum_{j=0}^{C_i-1} \frac{(\lambda_i x)^j}{\Gamma(j+1)}$$

La función generadora de momentos de una variable aleatoria X con distribución Híper-Erlang es:

$$E(X^n) = \sum_{i=1}^l \alpha_i \lambda_i^{-n} \frac{\Gamma(C_i + n)}{\Gamma(C_i)}$$

⁷⁷ FANG H. y W. Ma, Mobility management for wireless networks: modeling and analysis, En: Wireless Communications Systems and Networking, edited by M. Guizani, Kluwer Academic/Plenum Publishers, New York, 2004. p.473-512.

⁷⁸ ZHANG Yan , HU Honglin , FUJISE Masayuki . Opcit. p. 160.

⁷⁹ HONGXIA, Zhou. WANG, Junfeng. XU, Fanjiang y LI, Lei. Opcit. 540.

5.2.2 Distribuciones de probabilidad de tiempos entre llegadas a contrastar

De acuerdo a lo anterior, se van a contrastar dos distribuciones de probabilidad para modelar los tiempos entre llegadas: la distribución exponencial y la distribución Híper-Erlang. Para hacer comparables sus resultados es necesario tomar medidas equivalentes entre ellas y estimarse parámetros semejantes de las distribuciones de probabilidad seleccionadas. Para ello se supone que el parámetro lambda de la exponencial y de la Híper-Erlang es el mismo.

El Lambda o tasa de llegadas al sistema es el número esperado de llegadas por unidad de tiempo y se determinará a partir de la fórmula de utilización del sistema de la siguiente manera:

$$\rho = \frac{E(X)}{E(t)}$$

- Para la Distribución Exponencial

$$\rho = \frac{1/s\mu}{1/\lambda}$$

$$\lambda = \rho s\mu$$

- Para la Híper-Erlang

$$\rho = \frac{1/s\mu}{Ic/\lambda}$$

$$\lambda = \rho cs\mu$$

En donde μ es la tasa de atención, s el número de servidores, c el orden de la Híper-Erlang e I el número de sus fases.

La distribución de probabilidad Híper-Erlang posee un parámetro de forma que debe ser estimado, así como un número de fases y una probabilidad de ocurrencia de cada una de éstas. Para fijarle un valor a estos parámetros, se toma el valor esperado de la distribución de probabilidad y se suponen lambdas iguales para cada una de las fases, como se muestra a continuación:

$$E[X] = \sum_{i=1}^I \alpha_i \lambda_i^{-1} \frac{(C_i - 1 + 1)!}{(C_i - 1)!}$$

$$E[X] = \alpha_i \lambda^{-1} \frac{(C)!}{(C - 1)!} + \alpha_{i+1} \lambda^{-1} \frac{(C)!}{(C - 1)!}$$

$$E[X] = (\alpha_i \lambda^{-1} + \alpha_{i+1} \lambda^{-1}) \frac{(C)!}{(C-1)!}$$

$$E[X] = \left(\frac{1}{\lambda}\right) \frac{(C)!}{(C-1)!}$$

$$E[X] = \frac{C}{\lambda}$$

Como resultado de este procedimiento se obtiene que si los parámetros en cada fase de la distribución Híper-Erlang son iguales y C=1, se generará una distribución exponencial; y si C=2 se tendrá una distribución Erlang. Esto evidencia, que a partir de valores iguales o superiores a 3 se tendrían variables aleatorias Híper-Erlang.

Tomando como referencia este hallazgo, se fijan los parámetros faltantes de la siguiente forma:

- C (orden de la distribución Erlang) = 3. A partir de este valor se garantiza la existencia de una distribución Híper-Erlang.
- l (número de fases) = 2. Con este parámetro se puede ver la interacción entre las fases y aprovechar la propiedad tipo fase, la cual permite determinar un modelamiento con mayor precisión y con un rango más amplio de sucesos⁸⁰.
- Alfa = 1/l = 1/2. De esta manera se asigna la misma probabilidad de ocurrencia para ambas fases.

En conclusión, se contrastará el desempeño del sistema con tiempos de llegada exponenciales con parámetro lambda; y tiempos de llegada Híper-Erlang con parámetros lambda, C=3 y dos fases con la misma probabilidad de ocurrencia (1/2).

Tabla 1. Resumen parámetros de distribuciones de llegada

Distribución	Exponencial	Híper-Erlang
Lambda	$\lambda = \rho s \mu$	$\lambda = l \rho c s \mu$
Orden (forma)	1	C = 3
Fases	1	l = 2
Probabilidad (fases)	1	$\alpha = 1/2$

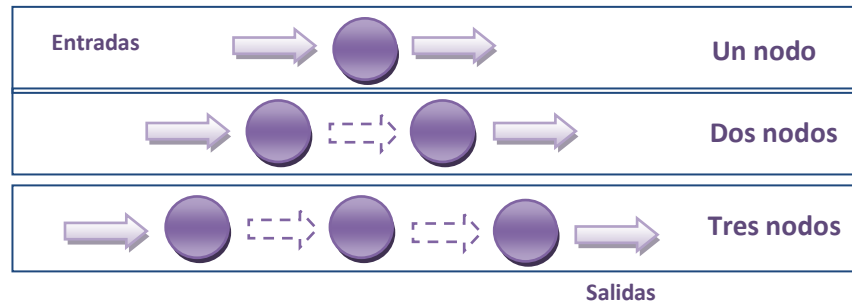
Fuente: Elaboración propia.

⁸⁰ WANG, Junfeng. LIU, Jin y CHUNDONG, She. Op. cit. p. 296 – 312.

5.2.3 Cantidad de Nodos

Los sistemas de colas actuales pueden ser muy diversos en cuanto a su forma y tamaño. Debido a esto, no basta definir una red en serie para abordar el objetivo planteado, también es necesario determinar su longitud.

Figura 8. Cantidad de nodos en serie



Fuente: Elaboración Propia

La figura 8 muestra las redes a modelar con la cantidad de nodos que se contemplarán en este estudio. De acuerdo con esto, se construirán tres redes diferentes que puedan ser un marco de referencia para desarrollar un estudio estadístico y estocástico de colas; y además un medio para representar sistemas de filas reales.

5.2.4 Cantidad de servidores por nodo

Dentro de cada nodo o punto de atención existe un número de servidores que atienden los flujos de clientes y evacuan las filas. Para la evaluación de la cantidad de servidores necesarios en cada nodo y su impacto en el sistema, se variará dicha cantidad tomando como referencia las gráficas y tablas que muestran Hillier y Liberman en el capítulo de Teoría de Colas del libro *Introducción a la Investigación de Operaciones*⁸¹. En dichas gráficas, se muestra la variación del número de clientes promedio en cola en un modelo M/M/S con respecto al factor de utilización para 1, 2, 3, 5, 7 y 10 servidores.

Esta cantidad de servidores disponibles por nodo variará en cada uno de éstos, ya que en muchos procesos no existe una homogeneidad en el servicio de nodo a nodo, es decir, que el número de personas que atienden por ejemplo en la recepción es diferente al número de servidores que atiende más adelante en el mismo proceso. Por lo tanto, dichas cantidades variarán en cada uno de los nodos para formar combinaciones de atención entre cada punto en serie.

⁸¹ HILLIER y LIEBERMAN. Op. cit., p. 858 – 859.

5.2.5 Porcentaje de utilización por nodo

Dentro de cada nodo existe una relación denominada factor de utilización del sistema que determina qué tan congestionado está el sistema, qué impacto genera ello en el correcto funcionamiento y si se alcanzará el estado estable de operación.

Rho o factor de utilización resulta de la división entre el tiempo esperado de atención, sobre el tiempo esperado entre llegadas al sistema y significa la fracción esperada de tiempo que los servidores están ocupados⁸²; por esta razón, el factor de utilización tiene repercusión directa en el estado de las colas a pesar de no ser una de las seis características consideradas por Kendall y Lee.

$$\rho = \frac{E(x)}{E(t)}$$

En el libro *Métodos Cuantitativos para Administradores* de Hillier y Liberman⁸³, se desarrolla un ejemplo que ilustra este comportamiento y que se presenta a continuación:

Tabla 2. Utilización del sistema.

P	Longitud de la cola	F Relativa %	F acumulada %	% Causas	Corte	Valor Absoluto
<u>0,9</u>	<u>9</u>	<u>42,5</u>	<u>42,5</u>	<u>12,5</u>	<u>55,027</u>	<u>45,0</u>
<u>0,8</u>	<u>4</u>	<u>18,9</u>	<u>61,4</u>	<u>25,0</u>	<u>86,428</u>	<u>13,6</u>
<u>0,75</u>	<u>3</u>	<u>14,2</u>	<u>75,6</u>	<u>37,5</u>	<u>113,104</u>	<u>13,1</u>
0,7	2,33	11,0	86,6	50,0	136,613	36,6
0,6	1,5	7,1	93,7	62,5	156,201	56,2
0,5	1	4,7	98,4	75,0	173,426	73,4
0,25	0,333	1,6	100,0	87,5	187,500	87,5
0	0	0,0	100,0	100,0	200,000	100,0

Fuente: Elaboración propia.

La tabla anterior presenta el factor de utilización y la longitud de la cola de un sistema M/M/1 con lambda igual a rho, un único servidor y tasa de servicio igual a la unidad. Adicionalmente, esta tabla muestra la frecuencia relativa y acumulada sobre el total de personas en cola. Esta información es utilizada para aplicar el principio Pareto o regla 80/20⁸⁴, el cual dice que el 20% de una acción producirá el 80% de los efectos, mientras que el 80% restante sólo originará el 20% de los efectos. Al aplicar este principio es posible delimitar a partir de qué porcentaje de utilización se genera la mayor concentración de personas en cola.

⁸² Ibid. p. 770.

⁸³ Ibid. p. 727.

⁸⁴ J. M. Juran. *Juran y la calidad por el diseño*. (1996). P. 75.

El resultado muestra que el 37,5% de dichos ρ representan el 75,6% de la longitud de la cola. Por lo tanto, se tomarán utilidades del sistema dentro del rango de 0,7 a 0,99, haciendo saltos constantes de una unidad decimal. No se tienen en cuenta niveles de utilización iguales o superiores a 1 porque el modelo permanece en su estado estable* hasta 0.99⁸⁵; de ahí en adelante el sistema estaría en condición transitoria y no sería posible realizar el análisis que se quiere para la investigación.

5.2.6 Distribución de tiempos de servicio y Tasa de servicio por nodo

Uno de los objetivos de este estudio es determinar cómo la variación de los tiempos entre llegadas impacta el desempeño del sistema, por este motivo se deciden plantear y analizar dos escenarios de comparación con distribuciones de probabilidad diferentes de tiempos entre llegadas. De acuerdo con esto, en todos los escenarios a generarse, se tendrán distribuciones de probabilidad diferentes para los tiempos entre llegadas, bien sea exponencial o hyper erlang; y se manejarán tiempos de servicio iguales con tasas de atención constantes.

Para la definición de la distribución de probabilidad a utilizarse para modelar los tiempos de servicio, se toma como referencia la distribución Exponencial, ya que ésta es la más común dentro de la teoría de colas y permite modelar procesos independientes por su propiedad de no memoria, haciendo posible que el tiempo de atención de un cliente que llego en un momento determinado se distribuya exponencialmente al igual que el tiempo de otro cliente que llegue un periodo después, siendo uno independiente del otro.⁸⁶

Una vez definida la distribución de probabilidad para los tiempos de servicio, se debe determinar entonces su tasa media de atención, la cual representa el número de esperado de personas atendidas por unidad de tiempo en el sistema.

Para establecer esta tasa de servicio, se tomará como referencia el trabajo de grado realizado por Sandra Alvarado Vera de la Universidad Javeriana titulado “Diseño de una estrategia para el mejoramiento de la calidad del servicio de talleres por medio del entrenamiento técnico tomando como base de análisis y estudio a General Motors Colmotores”⁸⁷. En este estudio se presenta un taller automotriz con varios puntos de atención o sistemas de colas conectados en serie con una distribución de probabilidad entre tiempo de servicio exponencial y una cantidad variable de servidores por nodo.

Mediante un muestreo, en dicho trabajo se calculó una tasa de atención variable para cada punto. De las tasas de servicio calculadas, se selecciona la tasa más baja, pues al

* Estado del sistema en el que es independiente del tiempo transcurrido y estado inicial del mismo.

⁸⁵ HILLIER y LIEBERMAN. Op. cit., p. 854

⁸⁶ Ibid. 830

⁸⁷ ALVARADO Vera, Sandra. Diseño de una estrategia para el mejoramiento de la calidad del servicio de talleres por medio del entrenamiento técnico tomando como base de análisis y estudio a General Motors Colmotores. Trabajo de grado.(2004). p. 60-70.

ser dicho punto parte de una serie de procesos para la reparación de un automóvil, esta tasa es la que determina el nivel de atención del sistema. Es importante recordar que según la teoría de restricciones en los procesos en serie el nivel de atención será determinado por el eslabón más lento⁸⁸.

Debido al análisis anterior, la tasa μ de servicio de 0,16 clientes por minuto calculado en el trabajo de Alvarado Vera se toma como insumo para el presente estudio, no sólo por la compatibilidad entre los estudios sino por ser el referente más importante al marcar el ritmo de atención de todo un sistema, aun cuando existen otras tasas medias de diferentes proporciones.

5.2.7 Capacidad del sistema

La capacidad del sistema es uno de los factores más influyentes dentro de la percepción de un cliente hacia la calidad del servicio. Si la capacidad es insuficiente, un buen número de clientes quedaría desatendido al ni siquiera poder ingresar al lugar de atención. Por otro lado, los que estuvieran en él cuando se encuentre al tope de su capacidad seguramente se sentirán incómodos e insatisfechos.

Dicha capacidad por lo tanto debe estar acorde con la demanda y la oferta del servicio, y limitada por el espacio físico disponible. En el presente estudio la capacidad de cada nodo será determinada tomando como referencia centros de servicios de diferentes sectores y tamaños ubicados en Bogotá. Se tomarán en cuenta tres niveles diferentes de la siguiente manera:

Sistema Pequeño: Banco Santander⁸⁹ - Oficina Javeriana.



La oficina del Banco Santander ubicada en la Universidad Javeriana cuenta con tres puntos de servicio de diferentes especialidades con capacidad promedio de 10 personas.

⁸⁸ RENDER Barry, HEIZER Jay H . Principios de administración de operaciones. Pearson Educación, (2004). p. 575

⁸⁹ Observación directa dentro de sus instalaciones.

Sistema Mediano: Hospital San Ignacio - Urgencias⁹⁰.



La unidad de urgencias del Hospital San Ignacio tiene capacidad de 20 camas para la atención de pacientes críticos.

Sistema Grande: Súper-CADE⁹¹.



Una de las más grandes redes de colas de Bogotá en la calle 26 con carrera 30, en donde se atienden más de 300 tipos de solicitudes con capacidad de albergar en cada módulo aproximadamente a 30 personas.

5.3 DISEÑO DEL EXPERIMENTO

Cada uno de los factores que inciden en el comportamiento de las redes de colas y que se han descrito previamente dan lugar a un experimento que proporcionará la información necesaria para realizar las pruebas, generar los resultados y posteriormente analizar el impacto sobre las medidas de desempeño de las redes que se van a generar.

A continuación se presenta el diseño de dicho experimento siguiendo la metodología planteada en el libro Principios Estadísticos para el Diseño y Análisis de Investigaciones

⁹⁰ HUSI. Disponible en la web: http://www.husi.org.co/publicaciones.php?pub_id=317&abrir=2

⁹¹ Radio Santafe. Disponible en la web: <http://www.radiosantafe.com/index.php?s=supercade>

de Robert Kuehl⁹², bajo un protocolo específico que da la posibilidad de observar las implicaciones de los resultados importantes.

5.3.1 Objetivo del estudio

Determinar el desempeño de las diferentes redes de colas que se generan al combinar características de las líneas de espera como distribuciones de probabilidad de los tiempos entre llegadas, cantidad de nodos, servidores por nodo en el sistema, nivel de utilización y capacidad del sistema.

Adicionalmente, el análisis de los sistemas generados buscará encontrar valores para mejorar el desempeño de los sistemas analizados.

5.3.2 Variables de respuesta

Las medidas claves de desempeño de un sistema de colas son el número esperado de clientes en el sistema (L), el número de entidades en cola (Lq), el tiempo esperado en el sistema (W) y el tiempo esperado en cola (Wq)⁹³. De estas medidas de desempeño, se seleccionarán como variable respuesta Lq y Wq , ya que éstas son las que muestran como tal el comportamiento de las colas.

Debido a la ausencia de fórmulas acotadas que permitan calcular fácilmente las medidas de desempeño de las redes de colas es necesario recurrir a técnicas avanzadas de cálculo. Para esto se hará uso de la simulación en el software Rockwell Arena, el cual genera reportes consolidados para todas las corridas y reportes detallados por cada réplica.

Para conocer las medidas de respuesta de los modelos, se tomarán los datos del reporte detallado que se genera en Microsoft Access, como se muestra a continuación:

Figura 9. Indicadores dentro de los reportes de simulación

	ID	ReportID	Name
+	8	1	Servicio.Queue.WaitingTime
+	101	1	Servicio.Queue.NumberInQueue

Fuente: Elaboración Propia

⁹² KUEHL, Robert. Principios estadísticos para el diseño y análisis de investigaciones. Segunda Edición Thomson, (2000). p. 20-225.

⁹³ HILLIER y LIEBERMAN. Op. cit., p. 770

La información que se obtiene de cada modelo de simulación, será promediada y consolidada en Microsoft Excel, en donde se generarán los intervalos de confianza que permitirán hacer el posterior análisis.

El intervalo de confianza o Half width se halla bajo los supuestos del teorema del límite central, ya que el número de observaciones es suficientemente grande como para que los datos tengan un comportamiento normal.

La fórmula para el cálculo de es:

$$\text{Half Width} = Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

5.3.3 Factores controlables

En resumen, para ver el comportamiento de las redes de colas se variarán los siguientes factores controlables:

1. Número de nodos de la red
2. Número de servidores por nodo
3. Distribución de probabilidad de tiempos entre llegadas al sistema
4. Factor de utilización por nodo
5. Capacidad de cada nodo

Con ellos se conformarán dos sistemas principales: uno con distribución de probabilidad exponencial para los tiempos entre llegadas, distribución de los tiempos de atención exponencial, disciplina de cola FIFO, capacidad finita y población objetivo infinita (M/M/s/k); y otro con distribución Híper-Erlang de los tiempos entre llegadas, distribución de los tiempos de atención exponencial, disciplina de cola FIFO, capacidad finita y población objetivo infinita (G/M/s/k).

5.3.4 Tipo de mediciones/observaciones

Las observaciones se concentraran específicamente en los indicadores de medición Lq y Wq. Debido a que estas medidas corresponden a promedios de tiempo y entidades en el sistema de colas, son de tipo continuo y mayores que cero.

5.3.5 Niveles de los factores

Tabla 3. Resumen de niveles por factores.

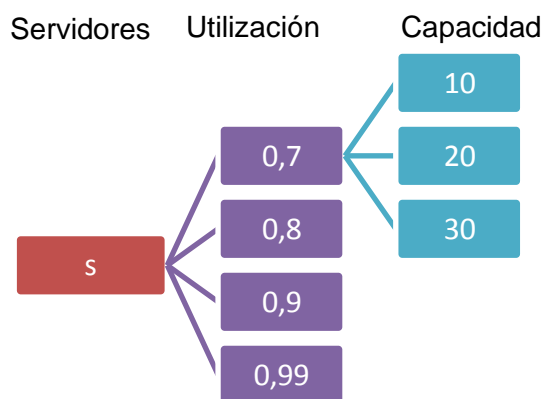
Factor	Niveles	Descripción
Número de nodos de la red	3	1, 2,3.
Número de servidores por nodo	6	1,2,3,5,7,10
Distribuciones de probabilidad de los tiempos entre llegadas	2	Exponencial, Híper-Erlang
Factor de utilización	4	0.7, 0.8, 0.9, 0.99
Capacidad de cada nodo	3	10, 20, 30

Fuente: elaboración Propia.

5.3.6 Tamaño del experimento

La naturaleza de las redes y la cantidad de factores e interacciones de los mismos hacen posible que las comparaciones entre los tratamientos se vean afectadas de manera sustancial por las condiciones en las que ocurren. Debido a esto, un diseño factorial es la metodología y el tratamiento adecuado para este experimento, ya que cada observación proporciona información sobre todos los factores y es posible ver las respuestas de un factor en diferentes niveles de otro factor en el mismo experimento⁹⁴.

Figura 10. Factores Anidados



Fuente: Elaboración Propia.

⁹⁴ KUEHL, Robert. Principios estadísticos para el diseño y análisis de investigaciones. Segunda Edición Thomson, (2000). p. 175.

Además de ser un diseño factorial, se tiene un experimento anidado, lo cual supone que cada nivel particular de uno de los factores ocurre en todos los niveles de otro factor anidado, como muestra la figura 9. De esta manera la cantidad de combinaciones del diseño anidado es la multiplicación de los factores según su interacción en cada uno de los nodos.

Tabla 4. Número de combinaciones del experimento

Modelo	Cantidad de Distribución	Cantidad de Utilizaciones	Cantidad de Capacidades	Cantidad de Servidores	Combinaciones
1 Nodo	2	4	3	6	144
2 Nodos	2	4	3 ²	6 ²	2592
3 Nodos	2	4	3 ³	6 ³	46656

Fuente: Elaboración Propia

En total se tienen 49.392 combinaciones. A pesar del alto número de combinaciones se procederá a hacer el factorial completo, ya que no hay homogeneidad entre niveles de los factores para construir bloques que permitan generar un modelo factorial fraccionado; además, no se conocen de antemano interacciones que no tengan sentido o interés analizar para ser eliminadas del factorial completo.

Con esto se da por terminado el diseño del experimento, luego de definir cuidadosamente los factores que inciden en un modelo de colas en red es posible ejecutar la simulación con el fin de producir resultados concluyentes del comportamiento de las colas y efectuar un análisis para descifrar algún método de mejora para dichos sistemas.

6 SIMULACIÓN DEL EXPERIMENTO

La determinación de los factores de incidencia al interior de las colas realizado anteriormente producen un enfoque orientado a la simulación debido a la elevada dificultad matemática de los modelos. Este capítulo explicará el procedimiento, tamaño y configuración que ésta debe poseer para obtener los resultados esperados.

6.1 CONFIGURACIÓN DE LA SIMULACIÓN

La simulación se realizará haciendo uso del software Rockwell Arena en tres distintos modelos según la cantidad de nodos que se quiera simular. Debido al elevado número de sistemas a simular, se realiza una macro para correr los modelos de forma automática, a partir de las diferentes combinaciones de factores que se encuentran organizadas en un archivo de Microsoft Excel (ver Anexo A). Esta macro automatiza el proceso al tomar la información de entrada desde Excel y llevarla a Arena para correr cada una de las 49.392 posibles combinaciones.

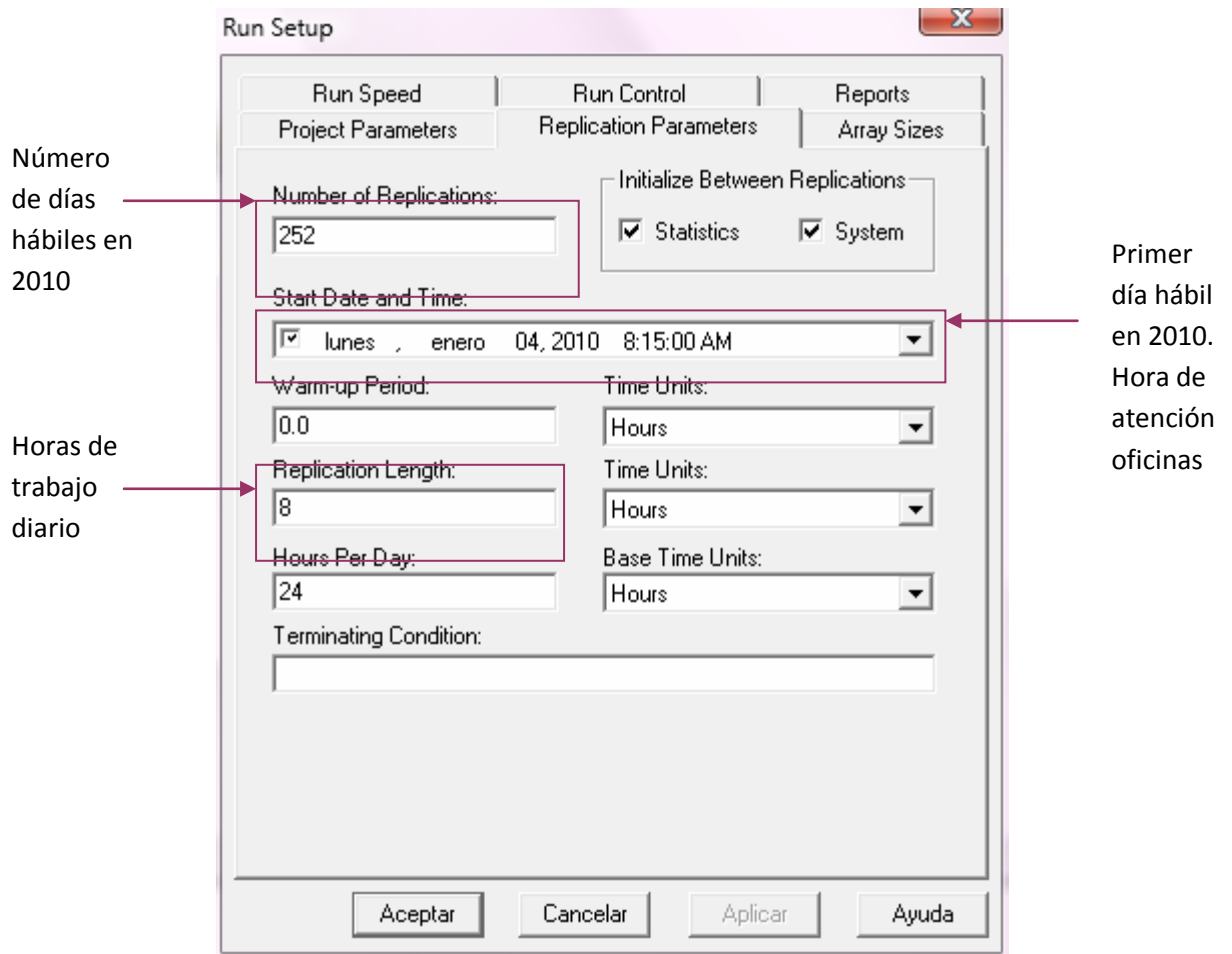
6.1.1 Definición de réplicas y duración

El número de réplicas se determina con el ánimo de simular un año de trabajo en el sector de servicios. Las empresas dentro de este sector trabajan de lunes a viernes ocho horas diarias según la ley laboral colombiana⁹⁵; y los días hábiles en Colombia para el año 2010 según el artículo 62 de la Ley 4ª de 1913 (Art. 70 C.C.)⁹⁶ son 252.

⁹⁵ Disponible en internet: www.proexport.gov.co
Reglamento Del Decreto Legislativo N° 854 Modificado Por La Ley N° 27671, Sobre Jornada De Trabajo, Horario Y Trabajo En Sobre Tiempo

⁹⁶ Disponible en internet: Alcandía de Bogotá
<http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=8684>

Figura 11. Configuración de réplicas en Arena



Fuente: Elaboración Propia del Autor

6.1.2 Técnicas de reducción de varianza

Para buscar la eficiencia experimental del diseño y aumentar su precisión, existen técnicas que buscan reducir la varianza y dan un mejor uso a la información disponible.⁹⁷ Dichas técnicas son usadas en estudios estadísticos donde es necesario llegar a resultados con la menor variabilidad posible para hacer comparaciones precisas y concluir acertadamente.

⁹⁷ CAO Ricardo, ABAD Ricardo, Introducción a la Simulación y a la Teoría de Colas. Editor Net biblo, (2002). P.103 -110

Una de dichas técnicas aplicables a este estudio es el método de Monte Carlo, ya que posibilita la realización de experimentos con el uso de números pseudo-aleatorios⁹⁸. La ventaja del método de Monte Carlo para este estudio es que tiene un error absoluto de la estimación que decrece en la medida en que se incrementa el tamaño de muestra del experimento (teorema del límite central: $\sigma = \frac{1}{\sqrt{n}}$), es decir, que entre mayor sea el número de réplicas menor será la varianza del mismo.

Para poder aplicar este método en el presente estudio, se recurre a procedimientos de generación de variables aleatorias con el ánimo de expresar las distribuciones de probabilidad en función de números aleatorios uniformes.

Utilizando el método de la transformada inversa para la distribución exponencial se tiene que:

$$F(x) = 1 - e^{-\lambda x}$$

$$1 - e^{-\lambda x} = R$$

$$x_e = -\frac{1}{\lambda} \ln R$$

Donde x_e son variables aleatorias exponenciales.

Para el caso de la distribución Hyper-Erlang, el método de la transformada inversa, a pesar de ser el procedimiento más común y sencillo, no puede ser utilizado pues no es posible realizar un despeje de la variable aleatoria. Debido a esto, se recurre al método de convolución y a las propiedades mismas de la distribución para generar valores de la variable aleatoria.

Por la definición de la distribución se sabe que ésta es la suma de i variables aleatorias Erlang cada una con valor esperado C/λ . Por consiguiente, como en el caso de la distribución Erlang es posible generar valores aleatorios con la suma de n variables aleatorias exponenciales⁹⁹, la Híper-Erlang puede ser generada con la suma de i variables aleatorias Erlang.

Se sabe por el método de convolución que una variable aleatoria Erlang¹⁰⁰ está dada por:

$$x = -\frac{1}{\lambda} \ln \prod_{i=1}^n R_i$$

Luego, la suma de dichas variables sería:

98 EVANS, Michael J. Probabilidad y estadística Editor Reverte, (2005) .p.675.

99 R. COSS, Bu. Simulación Enfoque Práctico. Méjico: LIMUSA, (1982) p. 60 – 65.

100 HAMDY A. Taha. Investigación de operaciones. Méjico: Prentice Hall. ed. 7. (2004). P. 649 - 652.

$$\begin{aligned}
x_{he} &= \sum_{i=1}^I x_i \\
x_{he} &= -\frac{1}{\lambda} \sum_{i=1}^I \ln \prod_{i=1}^n R_i \\
x_{he} &= -\frac{1}{\lambda} \ln \prod_{i=1}^I \prod_{i=1}^n R_i \\
x_{he} &= (-1/\lambda) * (\ln(R_i * R_i * R_i) * (R_i * R_i * R_i))
\end{aligned}$$

Donde x_i son variables aleatorias Erlang

Según esto, X_{he} es el generador de variables aleatorias Hiper-Erlang.

Ahora, con el objetivo de aumentar la precisión del experimento para el caso de esta última distribución, se decide aplicar otro método de reducción de varianza que es compatible con el ya usado, llamado Variables Antitéticas. Este procedimiento se basa en introducir correlación negativa entre dos muestras consecutivas con el fin de reducir la varianza total.

Para el caso de la distribución Hiper-Erlang definida en el experimento, se tiene que al ser λ constante para cada fase, la varianza del generador X_{he} calculado anteriormente está dada por:

$$Var(X_{he}) = \frac{1}{N^2} \left(\sum_{i=1}^N Var(R_i) + 2 \sum_{i,j=1, i < j}^n Cov(R, R_j) \right)$$

De tal forma que si es posible que la covarianza entre las variables aleatorias $R_{i,j}$ sea negativa, se reducirá la variabilidad total de la réplica.

Ya que $R_{i,j}$ son números uniformemente distribuidos entre cero y uno, tal que $R_{i,j} \sim U(0,1)$ entonces $(1-R)$ también se distribuye uniformemente. Como ambos son estimadores insesgados, una es el complemento de la otra y su covarianza es negativa se les denomina variables antitéticas.

Para este caso, la forma más sencilla de conseguir la covarianza negativa entre los seis números pseudo-aleatorios uniformes incluidos en el estimador X_{he} es que las variables R_{2i-1} tomen el valor de R , y las variables R_{2i} tomen el valor $(1-R)$.

De acuerdo a esto, se tiene como resultado del generador de variables aleatorias Hiper-Erlang la siguiente expresión:

$$x_{he} = \left(-\frac{1}{\lambda} \right) * [\ln(R * (1 - R) * R) * ((1 - R) * R * (1 - R))]$$

6.2 TIEMPO DE EJECUCIÓN DE LA SIMULACIÓN

Para poder estimar el tiempo de ejecución de la simulación, se toma como referencia la duración de ejecución de cada réplica, la cual es de aproximadamente un segundo. Si se realizaran 252 réplicas para completar un año de prestación de servicio se tendría un tiempo total por modelo de 3,08 minutos. Si el número total de modelos a correr es 49.392, entonces se tiene una duración de 2476.7 horas, es decir, 103.19 días.

6.3 REQUERIMIENTOS DE SISTEMA

La corrida de cada modelo de simulación genera un archivo en Microsoft Access en donde consolida la información. Este archivo pesa en promedio 5.23 MB, para un total de 258.160 MB por todos los modelos. Adicionalmente, cada modelo guarda un archivo de texto de tiempos entre salidas de cada nodo –para un total de 144.648 archivos- cada uno de los cuales tiene un peso en promedio de 410 KB (total: 58.785.680 KB).

En general, todo el proceso de la simulación generará información con un peso aproximado de 306,5 GB. De acuerdo con esto se necesita equipos con un procesador de velocidad igual o superior a 1.3 GHz, memoria mayor a 1 GB y disco duro de capacidad mínima de 306,5 GB. En cuanto al software, el sistema operativo debe ser Windows, y son necesarios programas como Rockwell Arena, Microsoft Access, Microsoft Excel y Visual Basic con el complemento ADO (ActiveX Data Objects) y DAO (Data Access object) para la ejecución de la macro que unifica toda la información (ver Anexo B).

6.4 CONSTRUCCIÓN DEL LOS MODELOS DE SIMULACIÓN

Para la simulación de las redes de colas se construirán modelos que representen el experimento y que a su vez arrojen los resultados buscados. Dichos modelos están compuestos en su totalidad por los siguientes módulos:

- Create para la llegada
- Dispose para las salidas
- Process para cada nodo o punto de atención
- Decide para limitar la capacidad de los nodos.

Adicionalmente existirá un módulo ReadWrite para llevar un control de las salidas y entradas a cada uno de los nodos de la red de colas. Aunque el estudio de este comportamiento no se encuentra considerado en el alcance de la actual propuesta, puede ser retomado para un estudio futuro.

6.4.1 Configuración de los módulos

6.4.1.1 Create: La red de colas estará compuesta por una única entrada en donde los usuarios llegan de acuerdo a un tiempo entre llegadas que se distribuye Exponencial o Híper-Erlang. En el caso de esta última, como Arena no posee esta distribución en sus librerías, es necesario generar variables aleatorias que sigan dicha distribución a partir de números pseudo-aleatorios. Para dicho fin se usa la expresión que resulta de las técnicas de reducción de varianza para los tiempos de llegada según una distribución Híper-Erlang. Para el caso de la distribución Exponencial, aunque la distribución exponencial si existe en las librerías del software de simulación, también se usará el generador calculado en el numeral de técnicas de reducción de varianza.

Figura 12. Configuración del modulo Create

Fuente: Elaboración Propia.

Híper-Erlang:

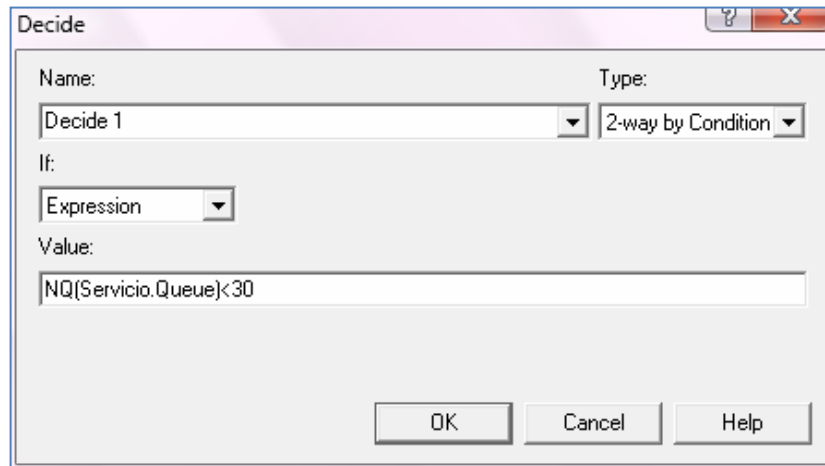
$$x_{he} = (-1/\lambda) * (\ln(UNIF(0, 1) * (1 - UNIF(0, 1)) * UNIF(0, 1)) * ((1 - UNIF(0, 1)) * UNIF(0, 1) * (1 - UNIF(0, 1))))$$

Exponencial:

$$x_e = (-1/\lambda) * \ln UNIF(0, 1)$$

6.4.1.2 Decide: Luego del proceso de entrada se encuentra un módulo de decisión en donde se define la capacidad del sistema, se monitorea el estado de la cola y se direcciona el flujo de entidades en el modelo, bien sea hacia el inicio del proceso o hacia su retiro del sistema. Mediante el comando $NQ(\text{Servicio.Queue}) < \text{Capacidad del Nodo}$ se limita la entrada de usuarios al proceso de tal manera que al alcanzarse la capacidad establecida no se permite la entrada de nuevos usuarios. En caso que los usuarios no puedan acceder al mismo, este módulo los direcciona hacia una salida del sistema representada con el módulo Dispose.

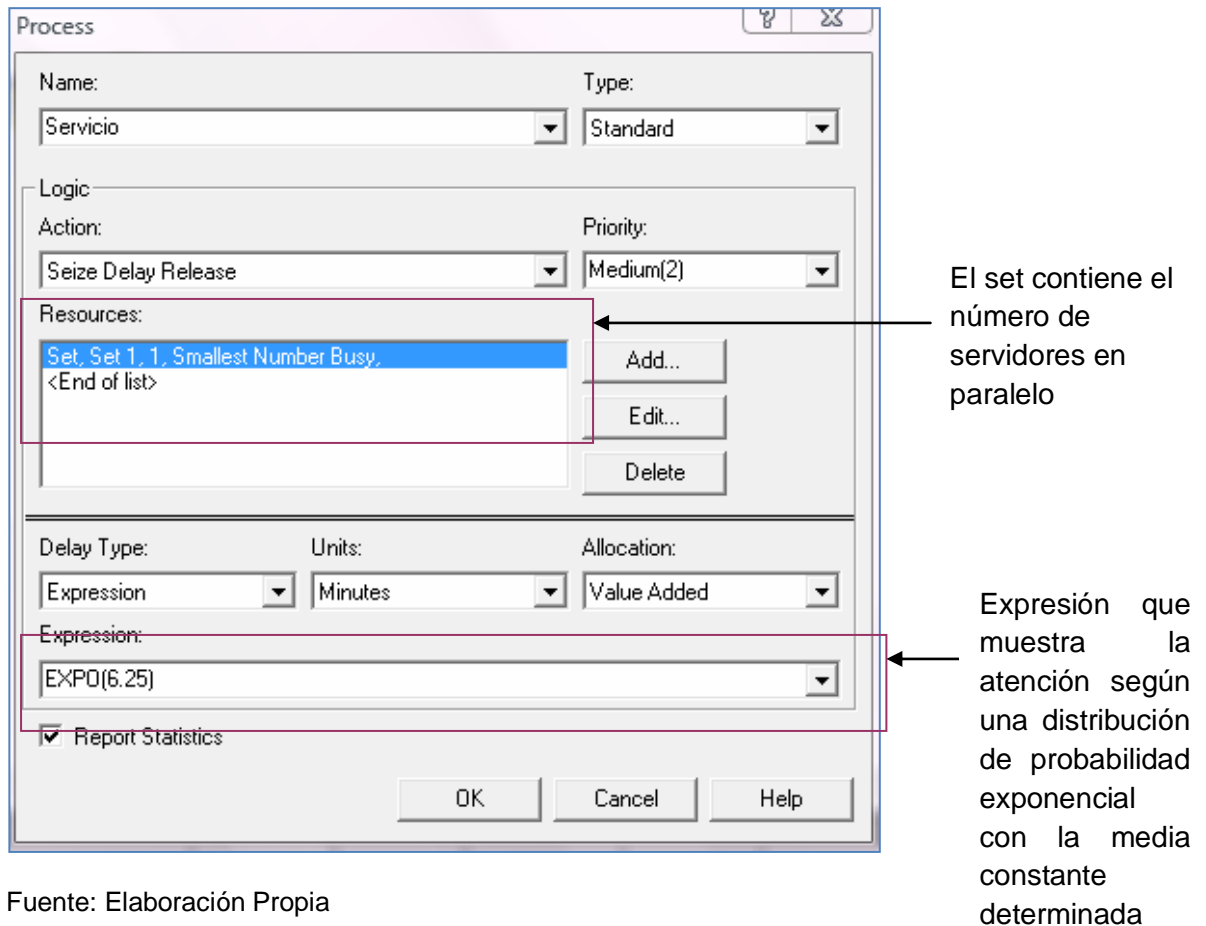
Figura 13. Configuración del módulo Decide



Fuente: Elaboración Propia.

6.4.1.3 Process: Representa cada uno de los nodos de servicio del sistema. Es aquí en donde se realiza el procesamiento de las entidades y en donde hay una cantidad de servidores en paralelo que son asignados a los clientes de acuerdo a la regla *Smallest Number Busy*. El tiempo de servicio sigue la distribución exponencial con parámetro μ que permanece constante en cada uno de los nodos.

Figura 14. Configuración del módulo Process



Fuente: Elaboración Propia

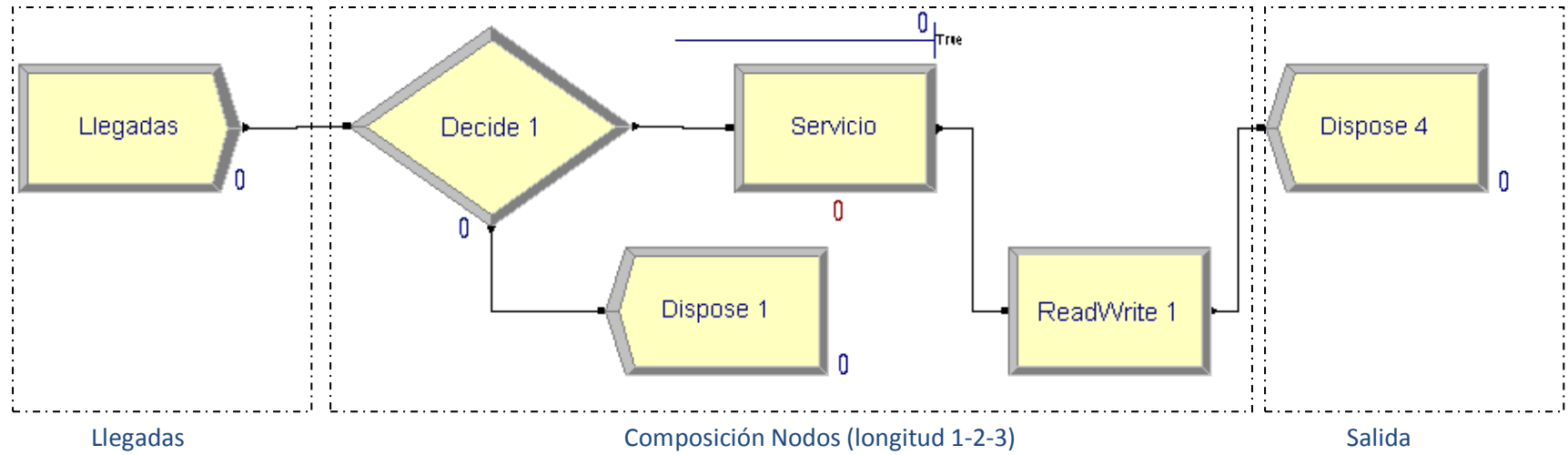
6.4.1.4 ReadWrite: Finalmente existe un módulo ReadWrite el cual se encarga de grabar el tiempo actual de cada salida del proceso y el inicio del servicio de la siguiente transacción mediante el atributo Tnow. El objetivo de este módulo es llevar un registro del número de salidas y de los tiempos entre éstas de cada uno de los nodos a lo largo de la simulación.

6.4.1.5 Dispose: Es el módulo que representa la salida de las entidades del sistema de colas. Las salidas pueden presentarse en dos momentos diferentes: el primero de ellos es cuando la capacidad de los nodos o puntos de atención es insuficiente para el número de personas que quieren acceder a él. En ese momento el sistema estará lleno y el modulo Decide direccionará las entidades hacia algún Dispose. La segunda salida se presenta cuando una entidad que logró entrar al sistema llega al final de la red de colas y abandona el sistema.

6.4.2 Modelos de simulación

Cada modelo de simulación está compuesto por la llegada, los nodos según la longitud del modelo (1,2 y 3) y finalmente una salida.

Figura 15. Modelos de simulación.



Fuente: Elaboración del Autor.

7 RESULTADOS DEL EXPERIMENTO

Después de la realización de la simulación y el cálculo de los indicadores de desempeño de los modelos planteados, es necesario realizar un análisis de los sistemas y determinar cómo la variación de las características de la red de colas aquí consideradas permite lograr mejores resultados. Con este análisis se podrán vislumbrar las diferencias principales entre las dos distribuciones seleccionadas y determinar el cambio y los efectos entre los escenarios estudiados.

Para este análisis se procede a aplicar pruebas estadísticas con la ayuda del programa SPSS sobre la información consolidada de los 49.393 modelos del experimento (ver base de datos consolidada en Anexo C). Por medio de este análisis se podrá comprobar las relaciones evidenciadas mediante un análisis descriptivo.

Dentro de las pruebas a realizarse esta la homogeneidad de varianzas o ANOVA que mide la influencia de las variables categóricas (factores) sobre la variable dependiente cuantitativa¹⁰¹, en este caso sobre Lq y Wq; y valora los efectos principales de los factores e interrelaciones de éstos entre sí¹⁰². Adicional a esta prueba, se complementa el análisis con un estudio de los conjuntos homogéneos y los intervalos de confianza entre las interacciones de los factores.

A continuación se presenta el análisis de la simulación, el cual busca determinar mediante procedimientos estadísticos cómo se comporta un modelo de colas en red según los factores expuestos en el diseño del experimento y extraer de sus resultados información clave en busca de mejorar la eficiencia general de los sistemas aquí desarrollados.

7.1 GENERALIDADES: INTERACCIONES ENTRE LOS FACTORES

Con los resultados es posible ver la diferencia entre las medias marginales de los factores y dentro de ellas ver la relaciones individuales (primer grado) y combinadas (segundo y tercer grado) de los factores sobre los índices de desempeño. Resultados inferiores al 5% del nivel de significancia o *p value* confirmará la interrelación de factores.

Dentro de dicha interacción las relaciones de segundo orden para Lq y Wq son:

¹⁰¹ Walpole Ronald E., Myers Raymond H. y Myers Sharon L. Probabilidad y estadística para ingenieros. Sexta Edición, Editor Pearson Educación. P 347-398

¹⁰² Ibid. p.347.

- Factor de utilización y número de servidores (únicamente para Wq)
- Factor de utilización y capacidad del sistema
- Número de servidores y capacidad del sistema
- Distribución de los tiempos entre llegadas y factor de utilización
- Distribución de los tiempos entre llegadas y capacidad del sistema
- Distribución de los tiempos entre llegadas y número de servidores

Las interacciones de tercer orden para Lq y Wq son:

- Factor de utilización, distribución y capacidad
- Factor de utilización, capacidad y servidores

Interacciones entre los factores como niveles de utilización, distribución y servidores; o nivel de utilización, distribución, servidores y capacidad no tienen una relación significativa con los resultados, esto quiere decir que si bien individualmente dichos factores impactan los resultados de las variables dependientes (Lq y Wq), cuando se aumenta el nivel de interacción con la combinación de mas factores su efecto pierde importancia.

7.2 ANÁLISIS DEL NÚMERO PROMEDIO DE PERSONAS EN COLA

El número esperado de personas en cola es uno de los dos indicadores de desempeño seleccionados para medir el comportamiento de las redes de colas. Dentro del diseño del experimento se señalan todas las variables que serán estudiadas por afectar la fluctuación de este indicador. A continuación, mediante la organización y estudio de la información, se realizará un análisis descriptivo del comportamiento evidenciado.

7.2.1 Relación entre número esperado de personas en cola y nodo de atención

El número esperado de personas bajo parámetros de entrada y atención semejantes y capacidades infinitas tendría que ser igual en cada uno de los nodos según el teorema de Jackson y Burke¹⁰³ en sistemas con tiempos de llegada y servicios exponenciales; sin embargo, en sistemas con capacidades finitas la tasa media de salida de un nodo a otro difiere de la tasa de entrada. Este comportamiento se presenta debido a que no todas las entidades que quisieron acceder al servicio pudieron hacerlo, y en consecuencia, abandonaron el sistema disminuyendo la tasa al nodo siguiente.

¹⁰³ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 810 – 845.

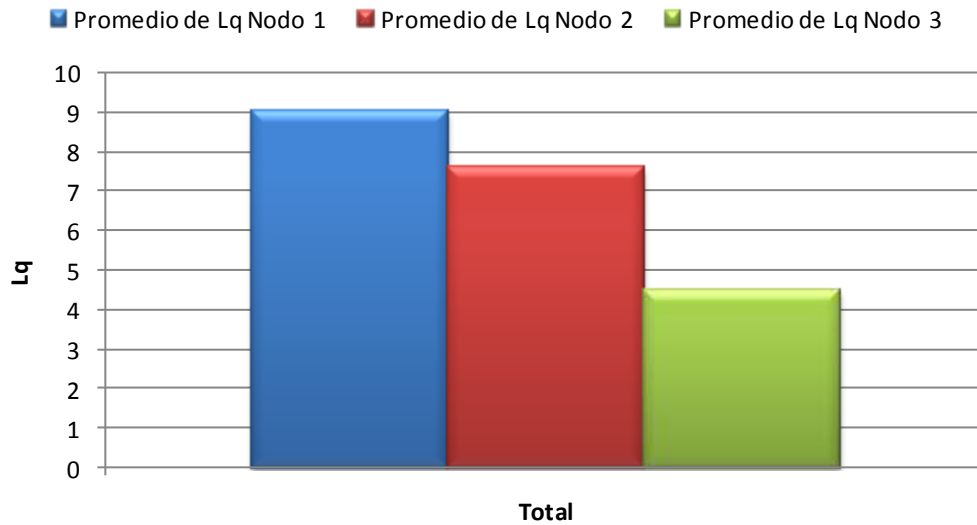
Figura 16. Redes de colas con capacidad finita



Fuente: Elaboración Propia.

Esta situación hace que la tasa de llegada a los nodos siguientes en la red sea menor y el número esperado de personas en cola disminuya.

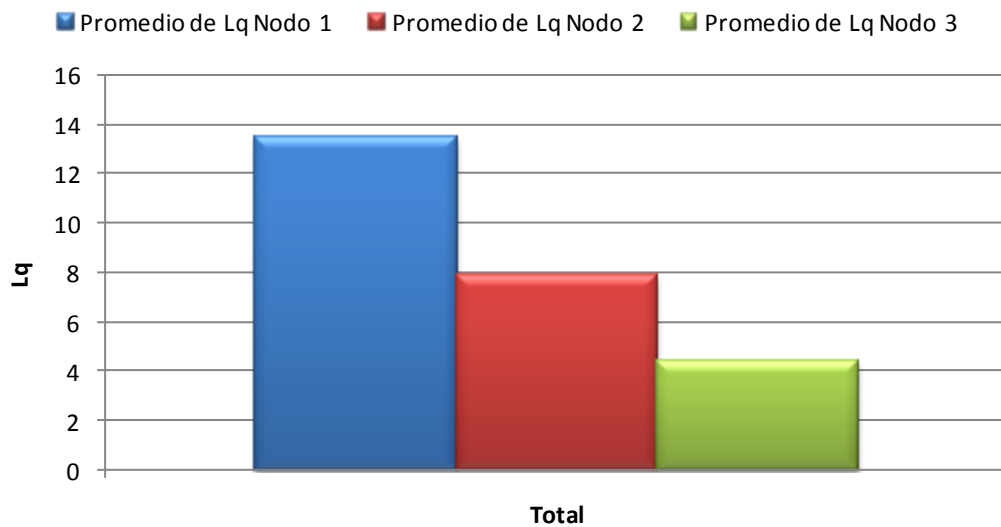
Figura 17. Resultados de L_q según el nodo de atención con tiempos entre llegadas exponenciales



Fuente: Elaboración Propia.

Al igual que en los sistemas con tiempos entre llegadas exponenciales, para el caso de la distribución Híper-Erlang, la tasa de llegada cambia y hacen que el primer nodo tienda a congestionarse. En la medida en que el nodo este más alejado de la fuente de entrada, los valores de L_q serán inferiores. La figura 18 evidencia este comportamiento entre nodo y nodo de la red.

Figura 18. Resultados de L_q según el nodo bajo la distribución Híper-Erlang.



Fuente: Elaboración Propia.

La diferencia entre el primer nodo y los nodos subsiguientes se debe a que la tasa de llegadas y la distribución de los tiempos entre llegadas hacen que en el primer punto de atención se formen grandes filas que impiden el paso a los siguientes nodos, formado un cuello de botella. El ritmo de salida del primer nodo al siguiente disminuye sustancialmente y ese efecto continúa en cadena haciéndose aun más evidente en la medida que la saturación aumente en la primera estación de llegada y la red sea más larga.

Realizando una comparación de los resultados del L_q para ambas distribuciones encontramos cambios superiores al 44% de la Híper-Erlang con respecto de la exponencial, cuya variación es más representativa en el nodo uno debido a la mayor variabilidad en los resultados de la Híper-Erlang en el punto de entrada del sistema frente a al exponencial.

Tabla 5. Cambio entre distribuciones según la longitud de la red (L_q)

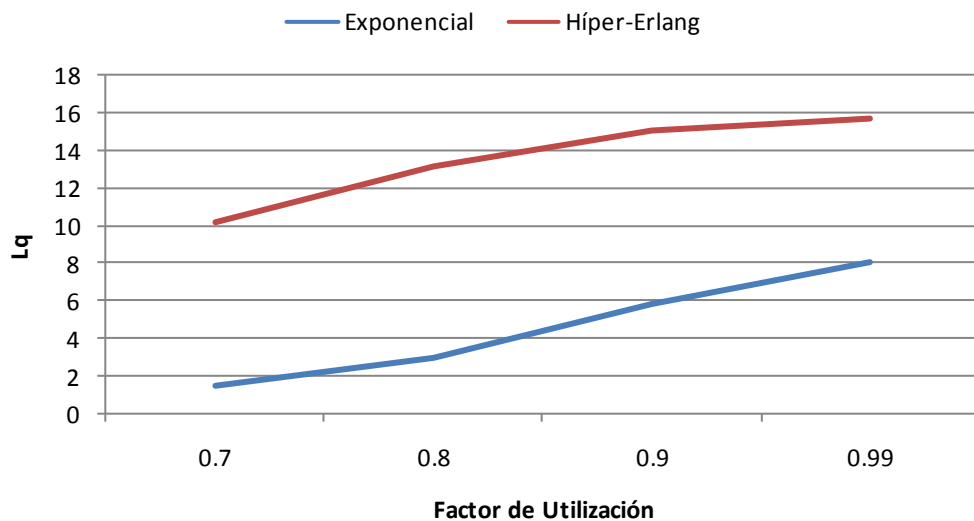
Longitud	ΔL_q
1 Nodo	44.4%
2 Nodos	7.83%
3 Nodos	5.40%
Prom. Total	19.21%

Fuente: Elaboración del Autor.

7.2.2 Relación entre número esperado de personas en cola y factor de utilización

El factor de utilización describe cuantitativamente que tan saturado está el sistema; por consiguiente, es natural ver que su aumento también impacta el número de personas en cola (L_q). Como se ve en la figura 19, el indicador de desempeño L_q crece conforme el factor de utilización se acerca a uno únicamente para el primer nodo que es donde dicho valor es conocido.

Figura 19. Resultados de L_q según el factor de utilización bajo la distribución exponencial en el primer nodo.



Fuente: Elaboración Propia.

Para sistemas con distribuciones exponenciales, el crecimiento del L_q con relación al ρ es lineal, con un ligero cambio en la pendiente que hace que se aumente el valor del indicador al acercarse el factor de utilización a uno. Para la distribución Híper-Erlang, el

Lq es mucho mayor con una curva cóncava que igualmente se incrementa cuando el Rho tiende a uno.

Los cambios entre distribuciones son mucho más tangibles cuando el factor de utilización es menor y presentan disminuciones importantes conforme el Rho tiende a uno. La saturación del sistema hace que la Híper-Erlang pierda aceleración en el crecimiento del Lq, mientras que la exponencial sigue aumentando linealmente.

Tabla 6. Cambio entre distribuciones según el factor de utilización (Lq)

Factor de Utilización	ΔLq
0.7	426.35%
0.8	401.43%
0.9	146.22%
0.99	25.75%
Prom. Total	249.37%

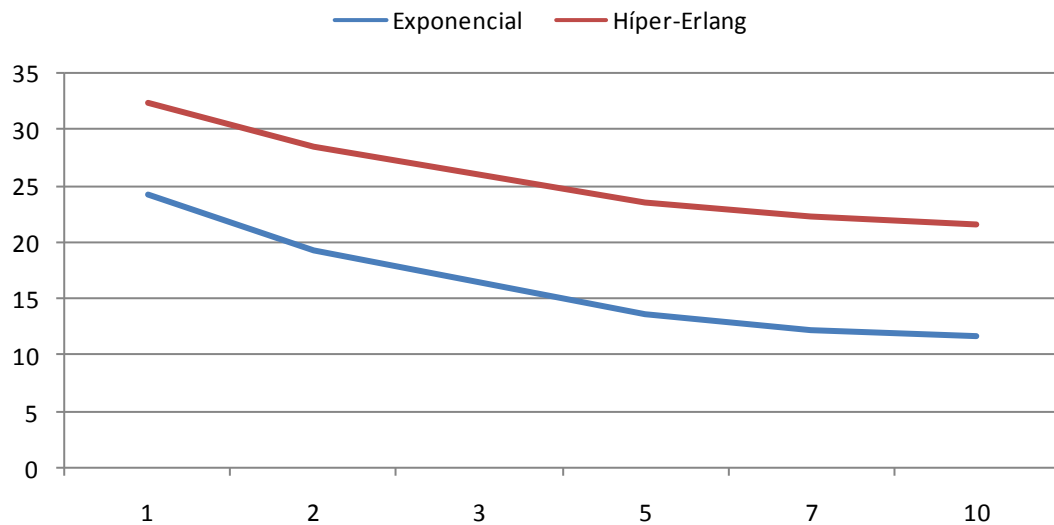
Fuente: Elaboración del Autor.

7.2.3 Relación entre número esperado de personas en cola y número de servidores

En la medida que el número de servidores aumenta en el sistema, la capacidad de desalojar las filas se hace más tangible, especialmente cuando la variación de servidores pasa de uno a dos. Esto significa que el impacto de la variación de un servidor en una fila es mayor entre menos servidores haya en dicho momento en el nodo de atención.

La figura 20 muestra los cambios que existen entre las dos distribuciones comparadas. Evidentemente los valores que toma el indicador Lq con la distribución Híper-Erlang son mucho mayores. Para efectos prácticos, el significado de esta variación indica que tan equivoco podría estar un modelo si se escogiese una distribución que en realidad no represente lo que pasa en un centro de atención; por ejemplo, en sistemas con un solo servidor la selección de la distribución de los tiempos entre llegadas genera un cambio del Lq de 33,4%; y aun más drástico es el cambio cuando son 10 los servidores en el sistema, en donde Lq se incrementaría un 84,11%.

Figura 20. Comparativo distribuciones según el número de servidores



Fuente: Elaboración del Autor.

Tabla 7. Cambio entre distribuciones según el número de servidores (L_q)

Servidores	ΔL_q
1	33.40%
2	47.10%
3	56.88%
5	71.32%
7	80.28%
10	84.11%
Prom. Total	57.34%

Fuente: Elaboración Propia.

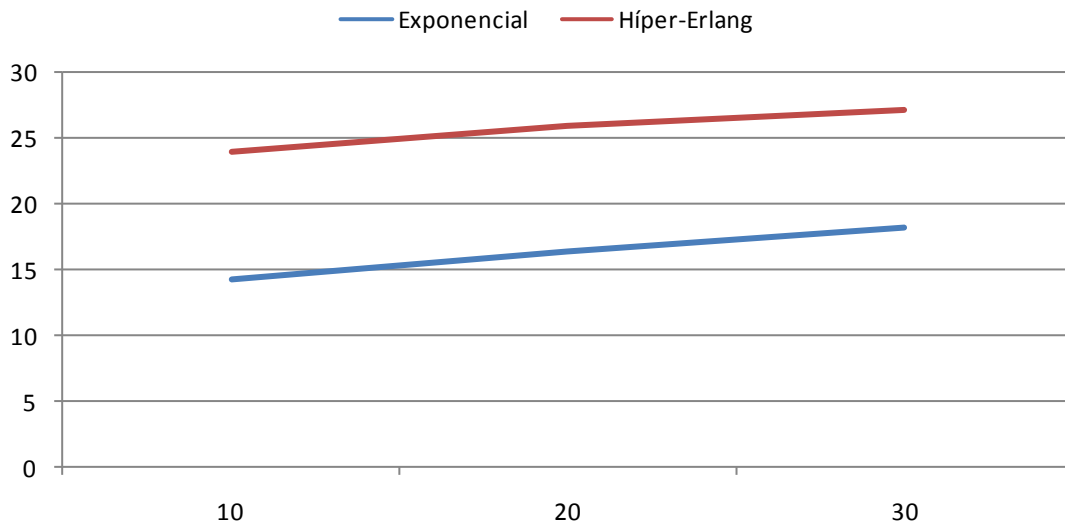
7.2.4 Relación entre número esperado de personas en cola y capacidad del nodo

La capacidad de los nodos limita el número de personas que en él esperan, por lo tanto, entre mayor capacidad, mayor será el número promedio de personas en el sistema. Sin embargo, cuando la utilización es de 0,7, el impacto de la capacidad es mínimo y en ciertos modelos es indiferente al número de personas en cola, es decir, que sin importar cuál sea la capacidad que se tenga en el punto de atención el número de personas que ahí esperan será similar entre todos los sistemas. Si esto sucede con un Rho de 0,7, es

posible concluir que para niveles de utilización inferiores este comportamiento debe ser similar e incluso más evidente, por lo tanto, no tendría importancia estudiar factores de utilización inferiores a 0,7 ya que el L_q sería indiferente bajo cualquier capacidad.

Los cambios entre ambas distribuciones tienden a mantenerse constantes como se ve en la figura 21, alrededor del 57,34% en promedio. La diferencia más significativa se da con la capacidad de 10 entidades en donde el valor de L_q para la distribución Híper-Erlang cambia con respecto a la exponencial en un 67.48%, y para las capacidades de 20 y 30 con una variación de 57.80% y 48.97% respectivamente.

Figura 21 Comparativo distribuciones según la capacidad (L_q)



Fuente: Elaboración Propia.

Tabla 8. Comparativo distribuciones según la capacidad (L_q)

Capacidad	ΔL_q
10	67.48%
20	57.80%
30	48.97%
Prom. Total	57.34%

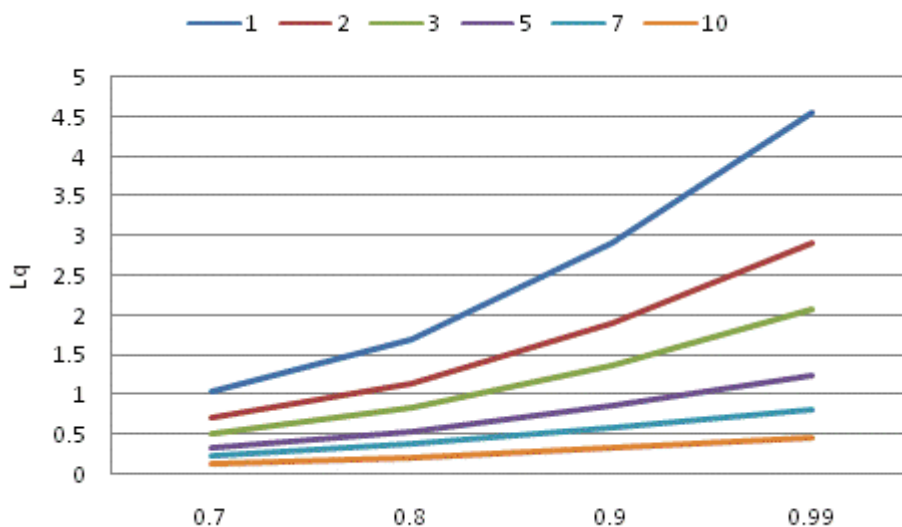
Fuente: Elaboración del Autor.

7.2.5 Relación entre número esperado de personas en cola, cantidad de servidores y nivel de utilización

En términos generales, la variación del L_q respecto a la cantidad de servidores y el nivel de utilización no presenta una interacción directa con los valores del L_q , sin embargo su relación dentro del primer nodo si permite evidenciar los efectos de los factores dentro del inicio de la red de colas.

Para el caso de los modelos con distribución exponencial, el L_q crece con un ritmo acelerado conforme ρ aumenta cuando hay hasta 3 servidores, y linealmente de ahí en adelante para cada uno de los nodos, como lo muestra la figura 22.

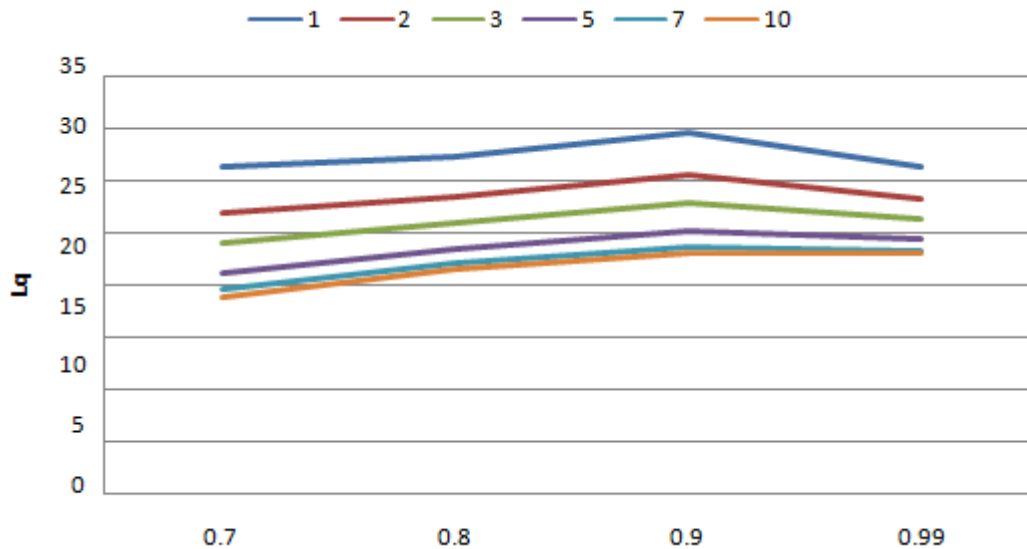
Figura 22. Resultados de L_q en el primer nodo según el número de servidores bajo la distribución Exponencial



Fuente: Elaboración del Autor.

Los resultados con la Híper-Erlang ratifican el menor impacto que tiene el aumento de servidores al depurar las filas conforme exista una mayor cantidad de los mismos en los puntos de atención.

Figura 23. Resultados de L_q en el primer nodo según el número de servidores bajo la distribución Híper-Erlang.



Fuente: Elaboración del Autor.

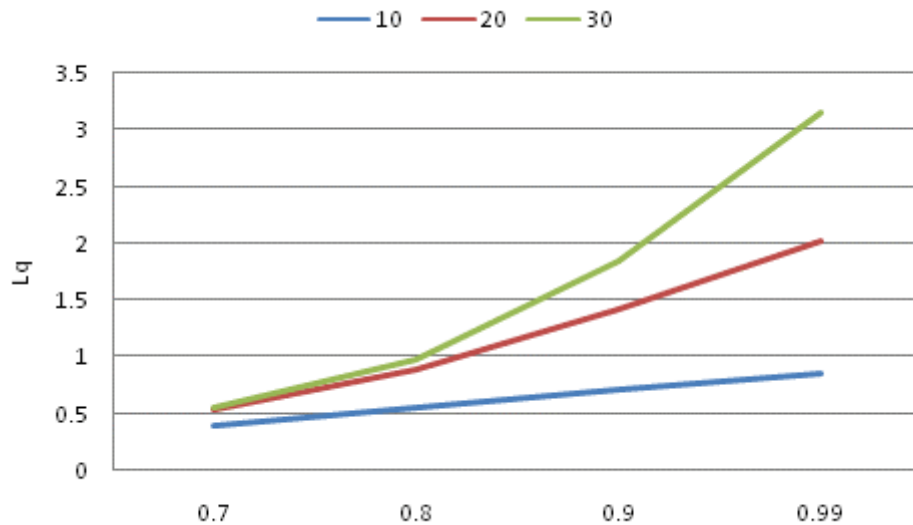
Adicionalmente, los resultados con la distribución Híper-Erlang presentan un punto de inflexión a la altura de niveles de utilización de 0,9. Este evento se debe a la saturación del sistema al alcanzar el límite de 30 personas por nodo. En ese momento aunque la tendencia diga que el L_q debe seguir en aumento, éste se estanca pues no se dejan ingresar o continuar dentro del sistema a todas aquellas entidades generadas por el modelo. Después de esto, L_q desciende desalojando las filas y atendiendo a las personas que permanecen en el sistema.

7.2.6 Relación entre número esperado de personas en cola, capacidad del nodo y nivel de utilización

La variación del L_q respecto a la capacidad y el nivel de utilización muestran la conveniencia o no de poseer mayor capacidad para albergar clientes según la carga de trabajo que posee una red de puntos de servicio.

En los modelos con distribución exponencial, las capacidades de 20 y 30 personas por nodo son casi indiferentes en el resultado del L_q bajo la utilización de 0,7 y 0,8. En la medida que la utilización crezca las curvas que representan L_q en diferentes capacidades tienden a volverse convexas entre mayor sea la capacidad del sistema.

Figura 24. Resultados de Lq según la capacidad en primer nodo bajo la distribución Exponencial

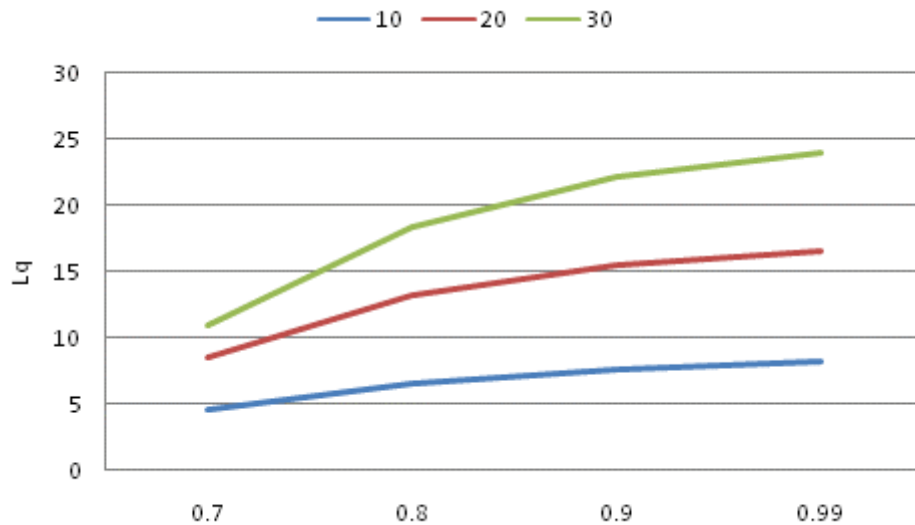


Fuente: Elaboración Propia.

Aunque el efecto descrito anteriormente es similar para las dos distribuciones de probabilidad, existe una diferencia relacionada con la velocidad de crecimiento, ya que el comportamiento de las curvas bajo la distribución Híper-Erlang tienden a ser cóncavas (figura 25). Este fenómeno no aplica cuando la capacidad es de 10 personas, en donde el Lq crece linealmente sin importar la distribución.

Las figuras 53 y 54 del Anexo D son evidencia de situaciones inusuales, que muestran comportamientos erráticos a partir de ρ iguales o superiores a 0,9. La explicación a ello es el mismo fenómeno que se explicó en la figura 23, en donde colapsa la atención, y se producen abandonos masivos. Esta situación genera el punto de inflexión cuando la capacidad es igual 30 entidades.

Figura 25. Resultados de L_q según la capacidad en el primer nodo bajo la distribución Híper-Erlang.



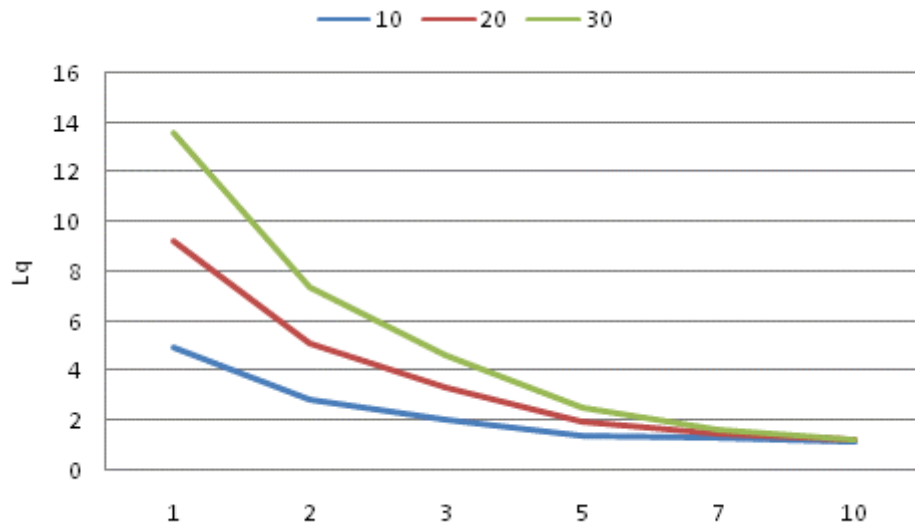
Fuente: Elaboración Propia.

7.2.7 Relación entre número esperado de personas en cola, capacidad del nodo y cantidad de servidores

En términos generales, la variación del L_q respecto a la cantidad de servidores disminuye conforme esta última aumenta para cada una de las capacidades. Las curvas que representan las capacidades tienden a acercarse conforme existan más servidores hasta que se unen cerca de los 7 servidores; de ahí en adelante es indiferente la variación de L_q , con lo cual se demuestra que el aumento de servidores por encima de diez, en este caso no tendría ningún efecto sobre las colas.

Para los sistemas con distribución exponencial, se observa cómo en la medida en que los servidores se aproximan a siete es indiferente la capacidad del sistema; mientras que cuando sólo existe un servidor, entre mayor sea la capacidad del sistema L_q aumentará drásticamente como se observa en la figura 26.

Figura 26. Resultados de L_q según el número de servidores y la capacidad el primer nodo bajo la distribución Exponencial

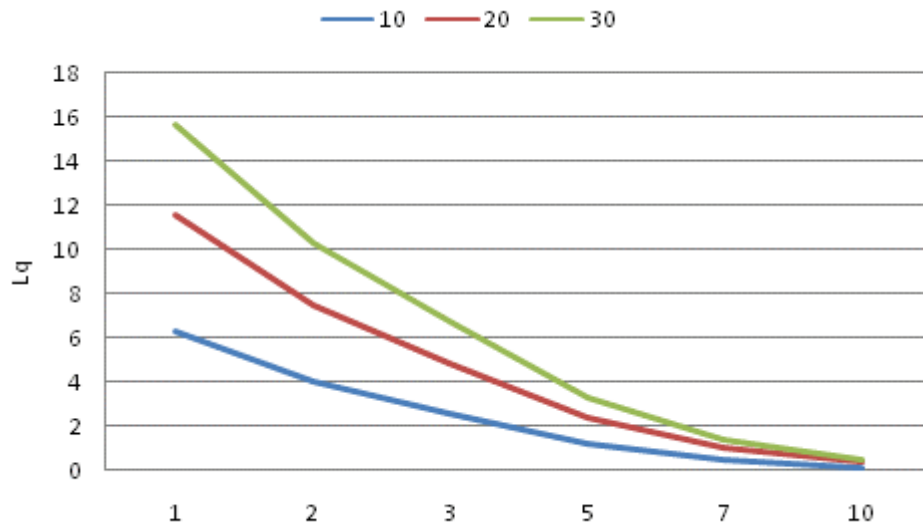


Fuente: Elaboración Propia.

Para el modelo Híper-Erlang, el comportamiento es idéntico pero con intersección entre las curvas que representan de capacidad a la altura de 10 servidores, lo que evidencia que la capacidad influye sobre el valor de L_q para todos los niveles de los servidores con excepción del último.

Respecto a los valores del L_q el cambio porcentual de los resultados de la Híper-Erlang respecto a la exponencial es en promedio del 57.34% para cada uno de los servidores según su capacidad.

Figura 27. Resultados de L_q según el número de servidores y la capacidad en el primer nodo bajo la distribución Híper-Erlang



Fuente: Elaboración Propia.

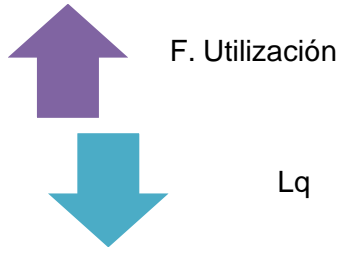
7.2.8 Resumen de interacción de factores sobre el número esperado de personas en cola

Tabla 9. Resumen relación entre número esperado de personas en cola y nodo de atención

Relación	Distribución Exponencial	Distribución Híper-Erlang
<p>Nodos</p> <p>L_q</p>	<p>Nodos:</p> <p>Disminución del L_q reduciendo la velocidad conforme el nodo está más alejado del origen de entrada. Variación del nodo uno a dos menor que del dos al tercer nodo.</p>	<p>Nodos:</p> <p>Disminución del L_q reduciendo la velocidad conforme el nodo está más alejado del origen de entrada. mayor variación entre los nodos y especialmente del primer al segundo nodo</p>

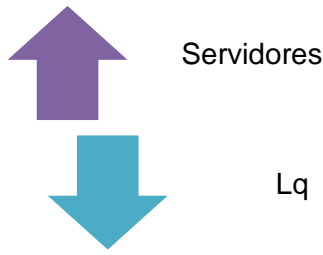
Fuente: Elaboración Propia.

Tabla 10. Resumen relación número esperado de personas en cola y factor de utilización

Relación	Distribución Exponencial	Distribución Híper-Erlang
	<p>Nodo 1: Aumento lineal del Lq con tendencia a acentuarse cuando el Rho se acerca a 1. Número esperado máximo promedio de personas en fila de 8.</p>	<p>Nodo 1: Aumento del Lq reduciendo la velocidad conforme el Rho se acerca a 1. Número mínimo de personas en fila 10.</p>

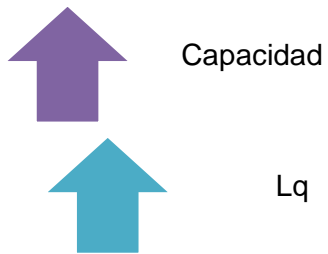
Fuente: Elaboración Propia.

Tabla 11. Resumen relación número esperado de personas en cola y número de servidores

Relación	Distribución Exponencial	Distribución Híper-Erlang
	<p>Nodo 1: Disminución del Lq conforme existan más servidores, en especial el cambio de 1 a 2 servidores. Lq similar con Rho 0,7 de 2 a 10 servidores.</p> <p>Nodo 2 y 3: Resultados similares con menor Lq en cada uno de los puntos.</p>	<p>Nodo 1: Disminución del Lq conforme existan más servidores, en especial el cambio de 1 a 2 servidores. Se producen anomalías en la tendencia cuando el Lq está cercano a la capacidad máxima del sistema</p> <p>Nodo 2: Lq menores con aumentos a menor velocidad que en el nodo uno y aumentando conforme el Rho se acerca a 1. Ver anexo D. Figura 57, 58, 59, 60.</p> <p>Nodo 3: Lq menores a los anteriores nodos con un crecimiento totalmente lineal. Anexo C figuras 61, 62, 63.</p>

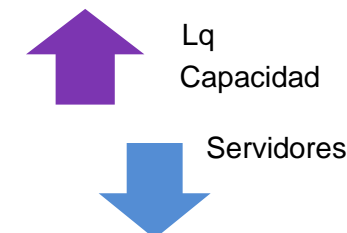
Fuente: Elaboración Propia.

Tabla 12. Resumen relación número esperado de personas en cola y capacidad del nodo

Relación	Distribución Exponencial	Distribución Híper-Erlang
 <p>Capacidad</p> <p>Lq</p>	<p>Nodo 1: Lq aumenta conforme la capacidad lo hace, con excepción de la utilización de 0,7. El crecimiento se torna convexo en las capacidades de 20 y 30; y totalmente lineal en la capacidad de 10 personas.</p> <p>Nodo 2 y 3: Lq con aumentos lineales poco pronunciados y con tendencia a mantenerse constante. Ver Anexo D.</p>	<p>Nodo 1: Lq aumenta conforme la capacidad también lo hace según una curva logarítmica.</p> <p>Nodo 2 y 3: Comportamiento lineal de Lq en todas las capacidades con inflexiones drásticas cuando el sistema llega a su límite de atención. Ver Anexo D. figuras 53, 57 y 61.</p>

Fuente: Elaboración Propia.

Tabla 13. Resumen Relación número esperado de personas en cola, capacidad del nodo y cantidad de servidores

Relación	Distribución Exponencial	Distribución Híper-Erlang
 <p>Lq</p> <p>Capacidad</p> <p>Servidores</p>	<p>Nodo 1: Disminución acelerada del Lq al pasar de uno a dos servidores. La capacidad genera mayor impacto a menor sea la cantidad de servidores hasta volverse indiferentes a partir de 7 servidores.</p> <p>Nodo 2 y 3: Lq con aumentos lineales poco pronunciados y con tendencia a mantenerse constante.</p>	<p>Nodo 1: Disminución del Lq al aumentar los servidores. La capacidad afecta el indicador menos cuando los servidores son diez.</p> <p>Nodo 2: Aumento lineal del Lq entre menos servidores existan. Las capacidades son indiferentes cuando hay 10 servidores.</p> <p>Nodo 3: Aumento acelerado del Lq entre menos servidores existan. Las capacidades con más de 7 servidores son indiferentes</p>

Fuente: Elaboración Propia.

7.3 TIEMPO PROMEDIO DE ESPERA EN LA COLA

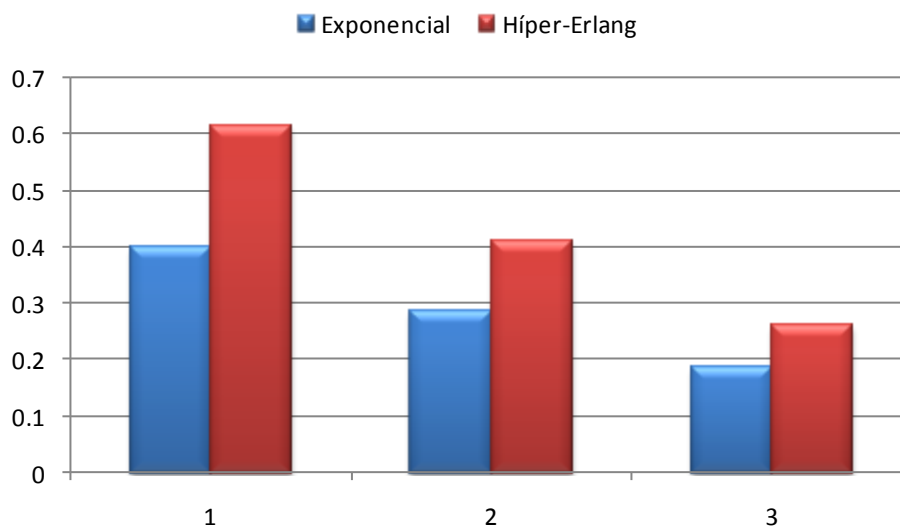
El indicador percibido más importante para el cliente es el tiempo de espera antes de ser atendido¹⁰⁴. Este indicador está ligado a la calidad del servicio y es el reflejo de qué tan bien estructurado está el proceso de atención al público. El comportamiento de Wq es similar al del indicador Lq descrito anteriormente. A continuación se mostrarán nuevas interacciones entre los factores, así como conclusiones acerca del funcionamiento de las redes de colas.

7.3.1 Relación entre tiempo esperado en cola y nodo de atención

El tiempo promedio de espera según el nodo de atención presenta el mismo comportamiento que con el indicador Lq , la gráfica 28, evidencia los cambios que si bien presentan la misma tendencia difieren en su valor.

Explicar esta relación

Figura 28. Comparativo cambio del Wq a lo largo de los nodos entre las distribuciones



Fuente: Elaboración del Autor.

¹⁰⁴ C. Gronroos, Service Management and Marketing: Customer Management in Service. Competition, 3rd Edition, Wiley, 2007.

Tabla 14. Comparativo distribuciones según la longitud (Wq)

Longitud	ΔWq
1 Nodo	54,32%
2 Nodos	43,98%
3 Nodos	40,56%
Prom. Total	45.62

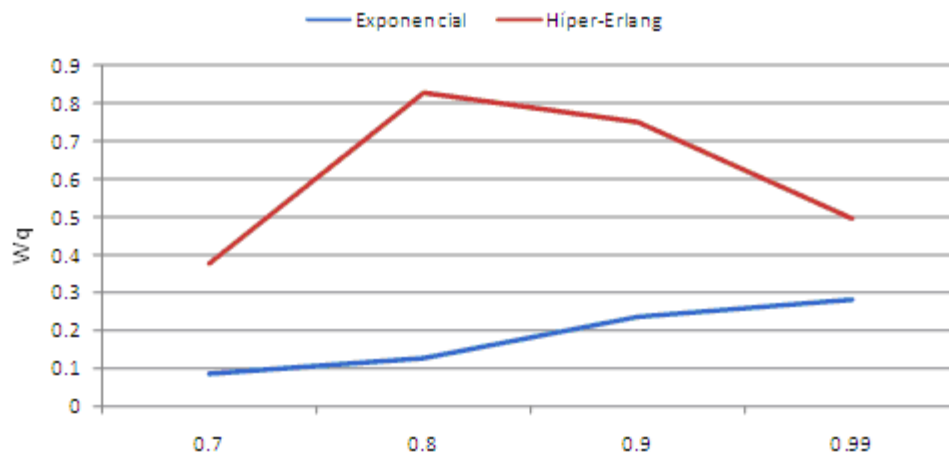
Fuente: Elaboración del Autor.

7.3.2 Relación entre tiempo esperado en cola y el factor de utilización

El tiempo promedio de espera en cola está directamente relacionado con el factor de utilización, ya que en la medida que el sistema este mas lleno las personas esperarán mayor tiempo para ser atendidas.

La figura 29 muestra el comportamiento de Wq según el tipo de distribución, dejando claro que el uso de alguna de las dos distribuciones determinará estados diferentes del sistema. Para el caso de la distribución exponencial, el tiempo de espera en el primer nodo crece linealmente. Por su parte, en la Híper-Erlang hay una variación pronunciada de Wq con respecto al nivel de utilización del sistema. Se percibe saturación en ρ igual a 0.8 en donde hay salidas del sistema.

Figura 29. Resultados de Wq en el primer nodo según el factor de utilización.

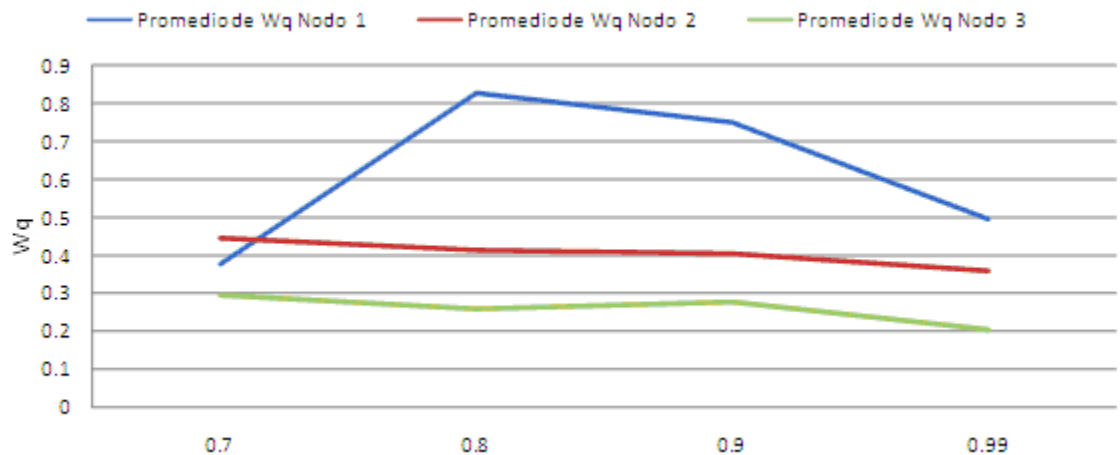


Fuente: Elaboración Propia.

En la distribución Híper-Erlang se observa que cuando se llega a la saturación, el sistema parece comportarse atípicamente debido a la congestión que presenta el punto de

atención. Las salidas del sistema son masivas, lo que hace que Wq disminuya drásticamente.

Figura 30. Resultados de Wq según el factor de utilización bajo la distribución Híper-Erlang.



Fuente: Elaboración del Autor.

El cambio porcentual de la Híper-Erlang con respecto a la exponencial alcanza un máximo en el nivel de utilización de 0.8, con una variación de 575.91% y un mínimo en 0.99 con un cambio de un 69.08%.

Tabla 15. Comparativo Distribuciones según el factor de utilización (Wq)

Factor de Utilización	ΔWq
0.7	287.93%
0.8	575.91%
0.9	204.09%
0.99	69.08%
Prom. Total	284.47%

Fuente: Elaboración Propia.

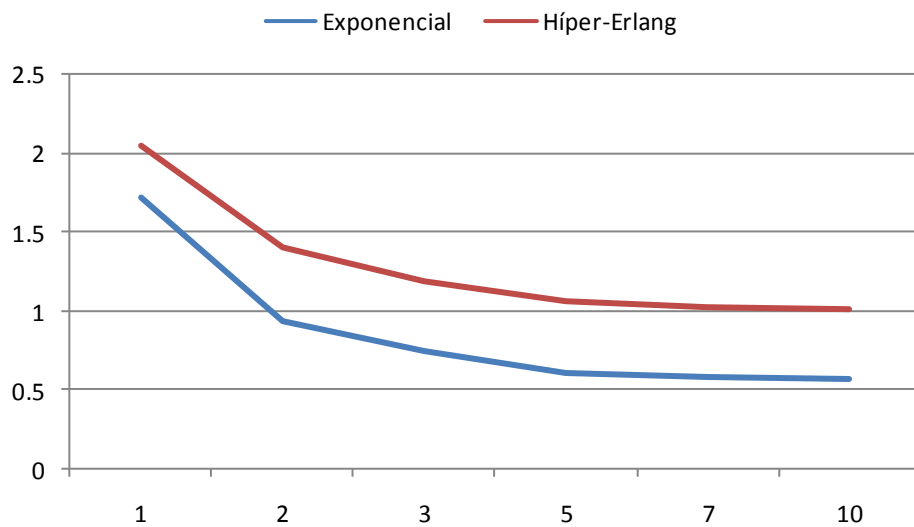
7.3.3 Relación entre tiempo esperado en cola y número de servidores

El número de servidores en los puntos de atención determina el tiempo promedio de espera en cola, entre mayor sea la cantidad de servidores el tiempo disminuye a lo largo

de todos los factores de utilización. Este comportamiento se hace aun más evidente cuando el sistema tiene una utilización cercana a uno. Por otro lado, la cantidad de servidores tiene un efecto positivo en la depuración de las colas sobre todo cuando existen menos servidores en el sistema.

Como se evidencio en la relación de L_q , ambas distribuciones difieren en sus resultados y esta diferencia aumenta conforme incrementa el número de servidores; sin embargo, la diferencia con tan solo un servidor de una distribución con respecto a la otra es mucho menor que lo visto anteriormente.

Figura 31. Comparativo distribuciones según la capacidad



Fuente: Elaboración Propia.

Tabla 16. Comparativo distribuciones según el numero de servidores (W_q)

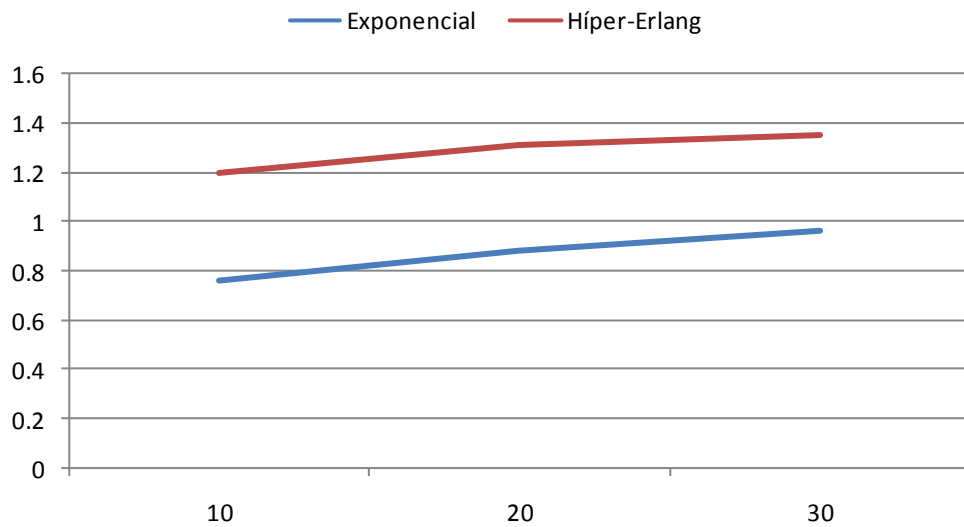
Servidores	ΔW_q
1	18.62%
2	47.63%
3	56.61%
5	71.83%
7	75.50%
10	75.59%
Prom. Total	48.40%

Fuente: Elaboración del Autor.

7.3.4 Relación entre tiempo esperado en cola y capacidad del nodo

El tiempo de espera aumenta conforme existe más capacidad en los nodos, lo que indica que la capacidad es determinante en el momento de atender a un mayor número de personas. Aunque los valores de L_q y W_q aumenten conforme la capacidad crece, en algunos casos la insuficiencia de capacidad hace que el número de personas que intentarían acceder al sistema no aumente. Para W_q se evidencia que las diferencias entre ambas distribuciones son mucho menores (alrededor de la mitad que lo visto en el análisis de L_q).

Figura 32. Comparativo distribuciones según capacidad (W_q)



Fuente: Elaboración del Autor.

Tabla 17. Comparativo distribuciones según capacidad

Capacidad	ΔW_q
10	58.43%
20	48.50%
30	40.43%
Prom. Total	48.40%

Fuente: Elaboración del Autor.

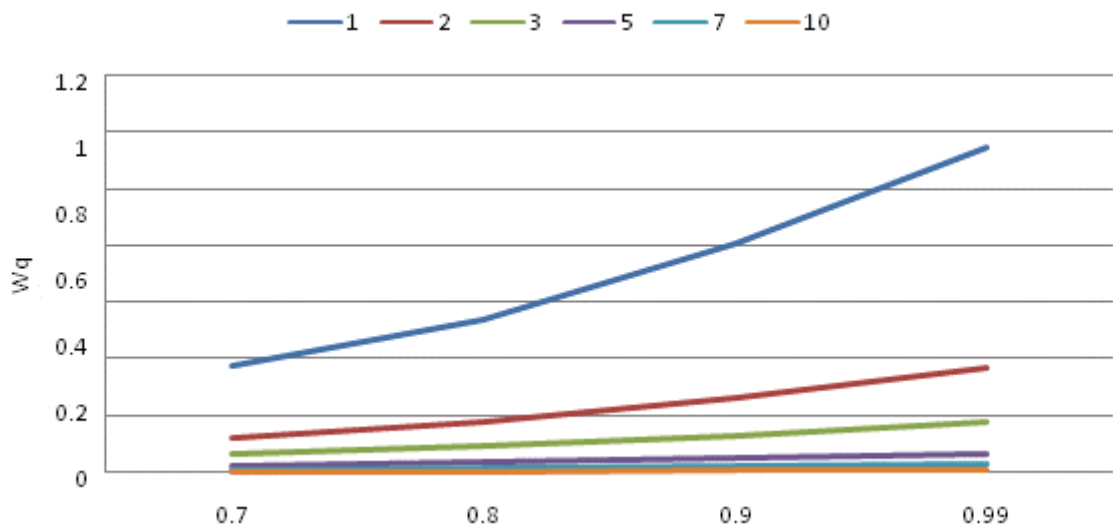
7.3.5 Relación entre tiempo esperado en cola, cantidad de servidores y nivel de utilización

El tiempo de espera en cola reduce significativamente conforme el número de servidores sea mayor, no obstante dicho impacto tiende a ser más o menos notorio dependiendo del uso del sistema.

En la distribución exponencial, el tiempo de espera es inferior para Rhos iguales a 0,7 y aumenta aceleradamente conforme la utilización se acerca a uno. Con un Rho de 0,7, no existen variaciones de Wq en la medida que aumentan los servidores. Si se tomaran factores de utilización inferiores a 0,7 seguramente las curvas se acercarían mucho más hasta que se hace indiferente la cantidad de servidores con respecto al Wq .

Adicionalmente, con esta distribución se confirma el comportamiento evidenciado en el estudio de Lq : el impacto en el tiempo de espera cuando se aumenta un servidor es mayor si existen menos servidores en el punto de atención.

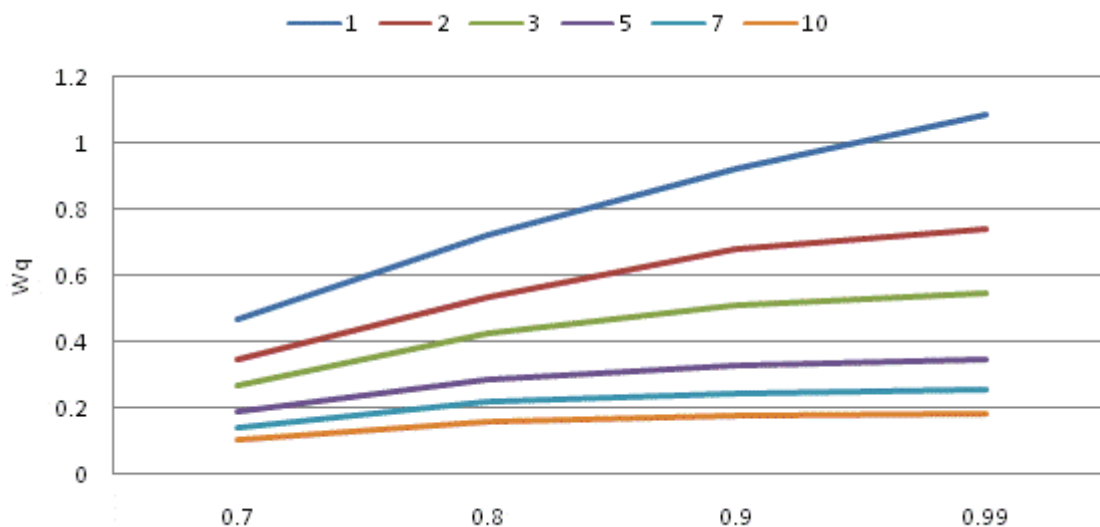
Figura 33. Resultados de Wq en todos los nodos según el número de servidores bajo la distribución Exponencial



Fuente: Elaboración Propia.

En el caso de la Híper-Erlang, el impacto del cambio entre servidores es mucho menor para uno y dos servidores, y mayor para el resto de ellos con respecto a la exponencial; además a lo largo de los Rho la diferencia entre Wq varía menos y la tendencia de las curvas según el número de servidores es cóncava.

Figura 34. Resultados de Wq en el primer nodo según el número de servidores bajo la distribución Híper-Erlang.



Fuente: Elaboración Propia.

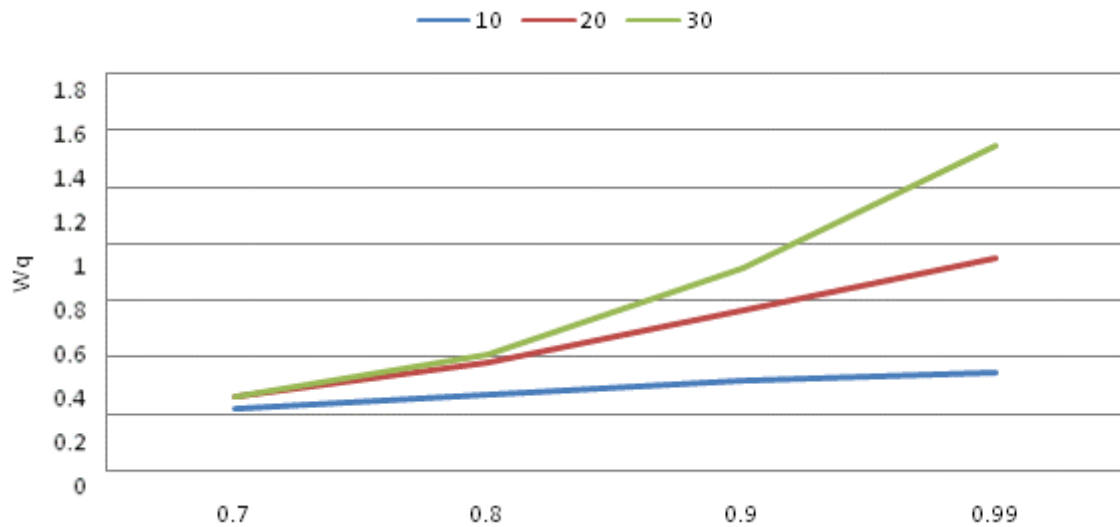
En el nodo dos, el cambio del Wq es indiferente cuando la cantidad de servidores se aproxima a diez, mostrando que no hay una relación entre el número de servidores y la utilización del sistema, es decir, que los resultados del tiempo de espera para un número de servidores mayores a 10 serían iguales sin importar cuántos servidores más se incluyan.

7.3.6 Relación entre tiempo esperado en cola, capacidad del nodo y nivel de utilización

Estas tres variables interactúan con el fin de ilustrar el efecto que tiene la capacidad en el tiempo esperado en cola según el factor de utilización. Su efecto es similar para ambas distribuciones ya que existen puntos en donde los cambios en la capacidad no tienen efecto en las media de desempeño.

Al interior de la distribución exponencial, es indiferente la capacidad respecto a Wq con factores de utilización menores a 0,7. Para sistemas con niveles de utilización de 0,8 y con capacidades de 20 y 30 personas, se pueden evidenciar comportamientos similares de Wq . A partir de niveles superiores a 0,8, el aumento del Wq vs. la capacidad a lo largo del factor de utilización crece linealmente.

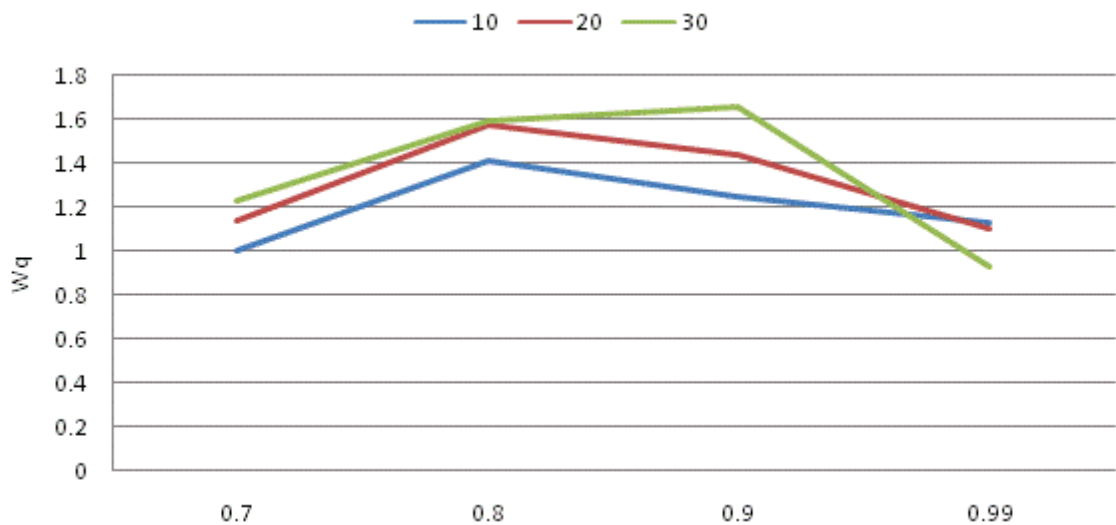
Figura 35. Resultados de Wq según la capacidad en el primer nodo bajo la distribución Exponencial



Fuente: Elaboración Propia.

Bajo la distribución Híper-Erlang, Wq tiene un comportamiento cóncavo, el cual se acentúa cada vez más en la medida que el factor de utilización se acerca a uno. Con Rhos menores a 0,7, las rectas tenderían a unirse, mostrando que sin importar cuál sea la capacidad, el tiempo de espera es constante. Los movimientos atípicos se deben a la llegada de los límites de capacidad.

Figura 36. Resultados de Wq según la capacidad en el primer nodo bajo la distribución Híper-Erlang



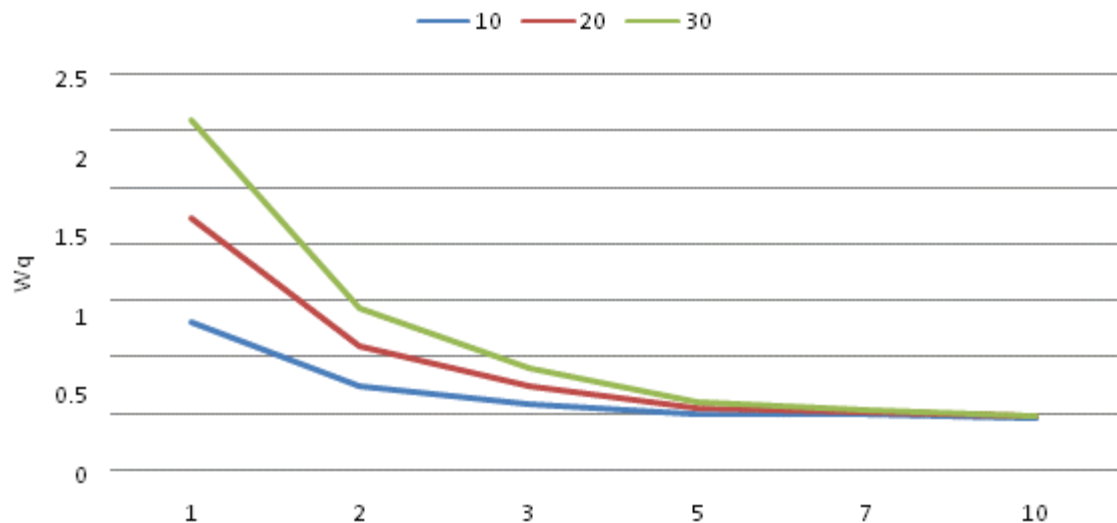
Fuente: Elaboración Propia.

7.3.7 Relación entre tiempo esperado en cola, capacidad del nodo y cantidad de servidores

El aumento del número de servidores disminuye el tiempo de espera, mientras que el aumento de la capacidad hace que la espera aumente debido al mayor número de personas con la posibilidad de hacer fila. Sin embargo, la diferencia en los valores de W_q con relación a la capacidad va disminuyendo y tienden a ser los mismos cuando el sistema posee 10 servidores en el punto de atención. Este efecto se presenta debido a que en el momento en que el sistema este relativamente despejado la capacidad deja de ser una limitación, es decir, que el tiempo esperado en cola es indiferente a la variación de la capacidad del sistema.

En sistemas con tiempos exponenciales, la variación entre 7 y 10 servidores para todas las capacidades es indiferente frente a W_q . Cuando el número de servidores es menor, la capacidad tiene un efecto en el tiempo de espera que tiende a ser más fehaciente sobre todo cuando existe un solo servidor.

1.1.1.1.2 *Figura 37. Resultados de W_q según el número de servidores y la capacidad en todos los nodos bajo la distribución Exponencial*

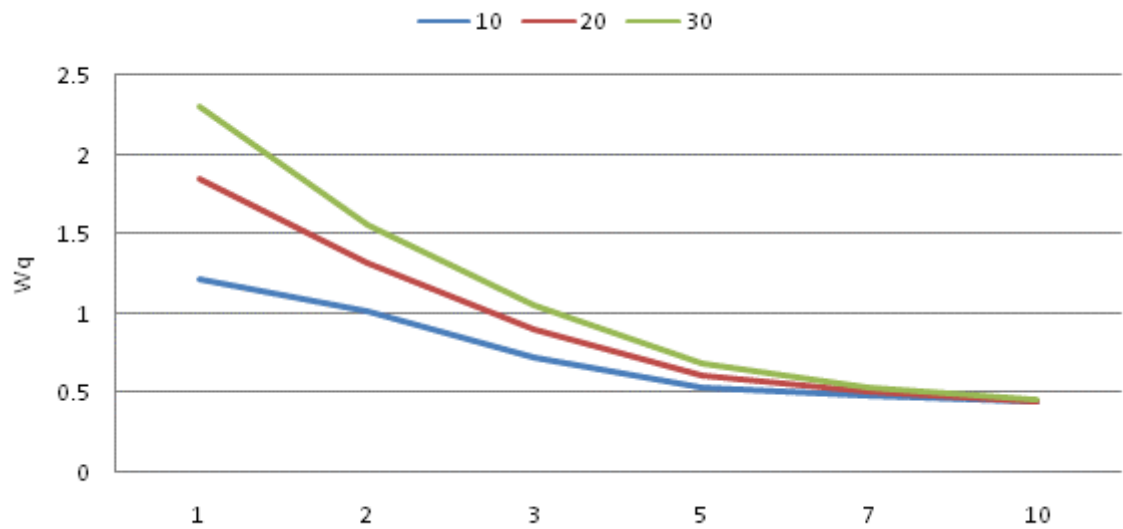


Fuente: Elaboración Propia.

En la distribución Híper-Erlang, se observa que en el nodo uno los valores de W_q disminuyen conforme aumenta el número de servidores. La capacidad 20 y 30 se aproximan bastante a largo de los servidores, sin embargo solo hasta 7-10 servidores la capacidad deja de tener importancia en el resultado del W_q .

Los resultados de Wq en los nodos dos y tres, tienden a comportarse según una curva convexa conforme los servidores aumentan. La diferencia de Wq con uno y dos servidores no es tan grande como lo fue en el caso de la exponencial, si ocurre que la capacidad no afecta el rendimiento de la cola cuando existen más de 7 servidores.

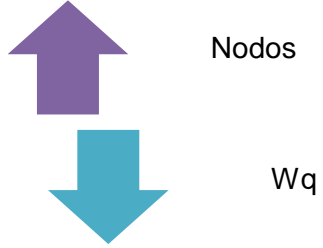
Figura 38. Resultados de Wq según el número de servidores y la capacidad en el primer nodo bajo la distribución Híper-Erlang



Fuente: Elaboración Propia.

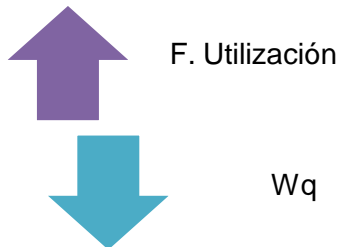
7.3.8 Resumen de interacción de factores sobre el tiempo esperado en cola

Tabla 18. Resumen relación tiempo esperado en cola y nodo de atención

Relación	Distribución Exponencial	Distribución Híper-Erlang
	<p>Nodos:</p> <p>Disminución del Wq reduciendo la velocidad conforme el nodo está más alejado del origen de entrada.</p>	<p>Nodos:</p> <p>Disminución del Wq reduciendo la velocidad conforme el nodo está más alejado del origen de entrada. mayor variación entre los nodos y especialmente del primer al segundo nodo</p>

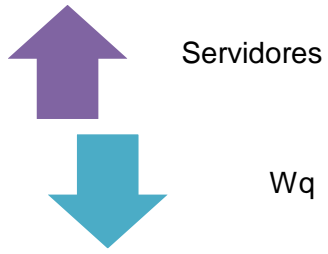
Fuente: Elaboración Propia.

Tabla 19. Resumen relación tiempo esperado en cola y el factor de utilización

Relación	Distribución Exponencial	Distribución Híper-Erlang
	<p>Nodo 1:</p> <p>Aumento lineal de Wq que tiende a acelerarse después de ρ igual a 0,8 en todos los nodos.</p> <p>Nodo 2 y 3:</p> <p>Estabilidad en los resultados del Wq con pequeños movimientos según el factor de utilización.</p>	<p>Nodo 1:</p> <p>Aumento del Wq aumentando con cambios drásticos a partir de ρ 0.8.</p> <p>Nodo 2 y 3:</p> <p>Wq inferiores al primer nodo debido a la saturación de este y la caída de la tasa de salida al interior de los nodos.</p>

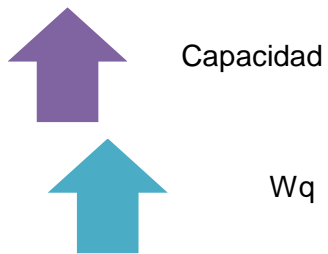
Fuente: Elaboración del Autor.

Tabla 20. Resumen relación tiempo esperado en cola y número de servidores

Relación	Distribución Exponencial	Distribución Híper-Erlang
	<p>Nodo 1,2,3: Disminución del W_q conforme existan más servidores, en especial el cambio de 1 a 2 servidores. W_q similar con ρ 0,7 de 2 a 10 servidores.</p> <p>Nodos 2 y 3: Resultados similares con menor W_q en cada uno de los puntos. Ver anexo E. Figuras 65, 66, 67.</p>	<p>Nodo 1: Aumento del W_q reduciendo la velocidad conforme el ρ se acerca a 1. Entre menos servidores existan el cambio del W_q es más evidente.</p> <p>Nodo 2: W_q menores con aumentos a menor velocidad que en el nodo uno y reduciéndose conforme el ρ se acerca a 1.</p> <p>Nodo 3: W_q menores a los anteriores nodos con un crecimiento lineal y comportamientos atípicos debido los resultados sobre el límite de la capacidad del sistema</p>

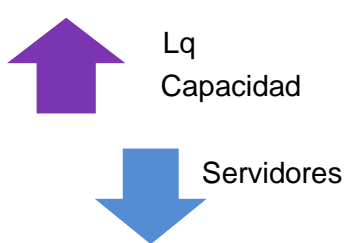
Fuente: Elaboración Propia.

Tabla 21. Resumen relación tiempo esperado en cola y capacidad del nodo

Relación	Distribución Exponencial	Distribución Híper-Erlang
	<p>Nodo 1: W_q aumenta conforme la capacidad lo hace, con excepción de la utilización de 0,7. El crecimiento se torna convexo en las capacidades de 20 y 30; y totalmente lineal en la capacidad de 10 personas.</p> <p>Nodos 2 y 3: Resultados similares con menor W_q en cada uno de los puntos.</p>	<p>Nodo 1: W_q aumenta conforme la capacidad también lo hace según una curva cóncava, con inflexiones cuando el sistema llega a su capacidad máxima.</p> <p>Nodo 2: Comportamiento lineal de W_q en todas las capacidades.</p> <p>Nodo 3: Comportamiento aproximado a lineal mucho menor que en los anteriores nodos.</p>

Fuente: Elaboración Propia.

Tabla 22. Resumen relación número esperado de personas en cola, capacidad del nodo y cantidad de servidores

Relación	Distribución Exponencial	Distribución Híper-Erlang
	<p>Nodo 1,2,3: Aumento acelerado de W_q entre menos servidores existan. Las capacidades generan mayor impacto en el W_q a menor número de servidores y son indiferentes a partir de 7 servidores. La brecha del resultado de W_q de uno a dos servidores es mayor que la diferencia de dos a más servidores.</p> <p>Nodos 2 y 3: Resultados similares con menor W_q en cada uno de los puntos. Ver anexo E. 65 – 71.</p>	<p>Nodo 1: Aumento acelerado del W_q entre menos servidores existan. Las curvas que representan las capacidades son similares cuando poseen el valor de 20 y 30.</p> <p>Nodo 2: Las capacidades son indiferentes cuando hay 10 servidores.</p> <p>Nodo 3: Efecto similar al del nodo dos, pero con curvas menos convexas. Ver anexo E. 59, 60, 61.</p>

Fuente: Elaboración Propia.

7.4 ANÁLISIS GENERAL DEL SISTEMA

Hasta el momento se han analizado las colas nodo a nodo, sin embargo cobra aun mas sentido analizar el comportamiento de las variables de respuesta, L_q y W_q , en el sistema completo. Este análisis permite estudiar el efecto de los factores anidados, ya que es posible ver configuraciones de un nodo de acuerdo a un nivel diferente al de los nodos siguientes; además se puede observar la relación integral de los factores dentro de la sinergia existente en una red y comprobar su efecto e incidencia a nivel individual en los nodos y general en toda la red.

7.4.1 Subconjuntos homogéneos

Para identificar subconjuntos homogéneos se aplica la prueba de Scheffé^{*} para cada factor. De comprobarse que no existen diferencias significativas entre las medias se rechaza la hipótesis nula con una significancia del 5%.

^{*} Prueba para hacer comparaciones múltiples de las medias de grupos. Realización en SPSS.

Según la capacidad:

Cuando las capacidades generales del sistema son superiores a 30 personas, es decir, cuando la capacidad en cada nodo es superior a 10 entidades se identifican conjuntos sin diferencia en sus medias para sistemas con capacidad para (40, 50, 70, 90) entidades.

Según el número de servidores:

Para el caso del número de servidores, existen 11 subconjuntos homogéneos, lo cual indica que ninguna configuración de servidores es independiente y todas ellas están relacionadas de algún modo.

La aparición de múltiples subconjuntos homogéneos entre los niveles de capacidad y servidores prueba que dentro de un sistema existen uno o varias configuraciones con igual resultado en los indicadores L_q y W_q pero impacto económico, procedimental o físico diferente.

7.5 RESUMEN MEDIDAS DE DESEMPEÑO

El análisis que hasta aquí se ha desarrollado se basa en los resultados que arrojó la simulación después de la combinación de los factores y niveles definidos para este estudio. Las distribuciones de probabilidad definidas para los tiempos entre llegadas permitieron realizar comparaciones sobre las medidas de desempeño de las interacciones del sistema, siendo siempre superiores las medidas de desempeño para los tiempos Hiper-Erlang. Adicionalmente, fue posible evidenciar similitudes con respecto a los cambios, patrones y tendencias en los dos escenarios analizados, lo cual hace pensar que las conclusiones de este experimento pueden extenderse a otras distribuciones de probabilidad, ya que el comportamiento para ambas distribuciones fue homogéneo respecto al efecto de los factores. A pesar de esto, queda abierta la posibilidad de modificar y/o adicionar otra distribución de probabilidad al experimento para evidenciar y corroborar la hipótesis aquí planteada.

7.5.1 Generalidades

- Entre mayor capacidad, mayor L_q y W_q .
- Entre mayor nivel de utilización, mayor es el valor de L_q y W_q .
- Si el número de servidores aumenta, los valores de L_q y W_q decrecen.
- Si se elige la distribución Hiper-Erlang los resultados del L_q y W_q son mucho mayores con respecto a la exponencial.

- Wq y Lq cambia a lo largo de los nodos de la red en serie. Dicha variación se acentúa en el primer nodo y pierde afecto conforme se avanza en la red.
- Exceptuando el nivel de utilización, la variación de un factor del primer al segundo nivel del mismo genera mayor cambio en los resultados del Lq y el Wq que la variación de los niveles sub siguientes.

7.5.2 Distribución Exponencial

- Los valores de los indicadores tienden a presentar menor variabilidad en los resultados.
- Los mayores cambios del Lq y Wq se dan en niveles de utilización superiores a 0.9.
- La capacidad del sistema tiene un efecto sobre Wq y Lq para menos de 7 servidores.
- El número de servidores afecta el rendimiento del sistema cuando se pasa de uno a dos servidores, en el resto de variaciones solo tiene efecto en los indicadores cuando el ρ tiende a uno.

7.5.3 Distribución Híper-Erlang

- Wq y Lq cambia a lo largo de los nodos de la red en serie. Dicha variación es del 76% del primer al segundo nodo y del 12% del segundo al tercer nodo. Luego, el aumento en la atención en los primeros nodos tendrá más impacto en el Lq total del sistema.
- Los cambios de Lq y Wq que se presentan en niveles de utilización de 0.9 y en ocasiones de 0,8 presentan situaciones extremas donde el sistema no es estable.
- La capacidad del sistema tiene un efecto sobre Wq y Lq cuando el factor de utilización es cercano a uno o la cantidad de servidores es superior a 7.
- El número de servidores afecta el rendimiento del sistema, especialmente al pasar de uno a dos servidores.

8 DISEÑO DEL MODELO DE OPTIMIZACIÓN

Los resultados de los modelos de simulación esbozan líneas de espera de diferentes características y particulares conclusiones; sin embargo, el estudio hasta ahora sólo ha presentado material informativo y no planes de acción o desarrollos que permitan mejorar los indicadores estadísticos ya calculados de las líneas de espera. El verdadero impacto de dicha información debe traducirse en el diseño y mejoramiento de las redes de colas existentes para así mitigar el efecto negativo que éstas pueden tener en la percepción del cliente.

El tiempo de espera de un cliente en fila antes de ser atendido es considerado el indicador de calidad más significativo en la percepción del cliente¹⁰⁵, ya que muestra qué tan bien está diseñado el proceso y cuál es la capacidad real que una compañía tiene para satisfacer sus necesidades.¹⁰⁶ En un mundo globalizado las empresas deben tener en cuenta la calidad de su servicio con el fin de atraer y mantener nuevos clientes y no asumir los costos de perder los que ya posee. Por lo tanto brindar la mejor atención es hoy en día la política número uno de calidad en cualquier empresa.

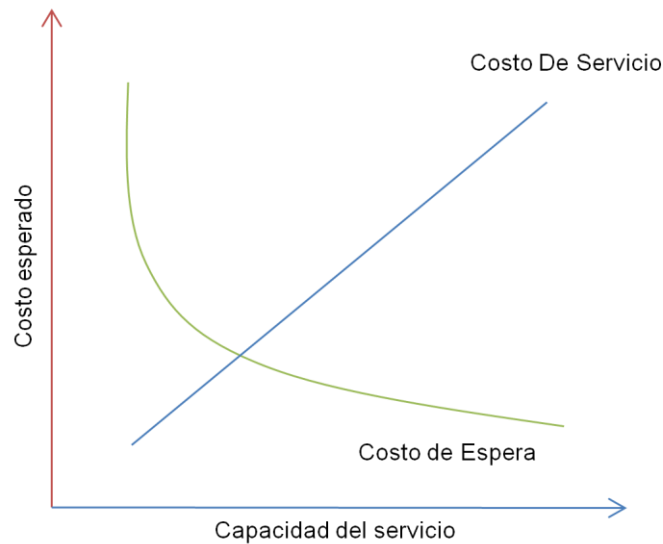
No obstante, dentro de esta nueva orientación existen objeciones de tipo económico y productivo, pues brindar un estado óptimo de calidad en donde el cliente no deba esperar por ser atendido es una opción costosa e ineficiente. Los recursos necesarios para soportar dicho escenario implica demasiados tiempos ociosos, por lo tanto, las empresas descartan de facto dicha opción y optan por ignorar o aplazar el estudio de esta problemática.

La figura 39 muestra la paradoja que tienen los directivos a cargo de gestionar las líneas de espera: un incremento en la capacidad de atención significa sobrecostos para la empresa (costo de servicio); no obstante, si no existe un nivel de atención suficiente se generaría un detrimento de la calidad del servicio (tiempo de espera) y una pérdida económica de oportunidad y de imagen, denominadas costos de espera. Ante esta paradoja, se debe buscar un equilibrio que satisfaga las necesidades de los clientes y se ajuste a las posibilidades de la empresa.

¹⁰⁵ C. Gronroos, *Service Management and Marketing: Customer Management in Service. Competition*, 3rd Edition, Wiley, 2007.

¹⁰⁶ M. Davis, and J. Heineke, How Disconfirmation, Perception and Actual Waiting Time Impact Customer Satisfaction. *En: International Journal of Service Industry Management*, 9(1), 1998, 64-73.

Figura 39. Costos del servicio.



Fuente: Elaboración Propia.

Para lograr esto, es fundamental entender los factores que inciden en un sistema de colas y conocer cómo éstos impactan los indicadores de desempeño del sistema, lo cual se realizó en el capítulo anterior. Con este entendimiento será posible definir variables, parámetros y restricciones que permitan encontrar herramientas y soluciones que apoyen el proceso de toma de decisiones y den la posibilidad de mejorar las condiciones actuales de servicio, para aumentar sus índices de calidad.

A continuación, se introducirán las variables y parámetros que se usarán para diseñar un modelo de optimización que cumpla con las características de las redes de colas aquí estudiadas, y que además brinde un aporte a los métodos de solución que normalmente se usan para abordar este tipo de problemas. Para ello se tomará como base el costo que implica cada variación del sistema, ya que este es el resultado con mayor impacto en las decisiones de los directivos.

8.1 COSTOS ASOCIADOS CON LAS REDES DE COLAS

El objetivo principal de un modelo de optimización orientado a las colas es la minimización del tiempo de espera y el número de entidades en las colas; no obstante, la minimización de estos indicadores implica grandes inversiones de capital por parte de la empresa, los cuales no pueden descartarse y se deben incluir en el proceso de optimización. Tanto la contratación de personal con el fin de depurar las colas y mejorar los niveles de servicio,

como la adecuación de las instalaciones representan costos que deben ser igualmente minimizados para brindar a las empresas soluciones eficientes y viables económicamente.

Dentro de la definición de los parámetros que afectan un sistema de colas, debe existir claridad acerca de cuáles son los costos implícitos en él, su naturaleza y causa independiente de la aplicación empresarial que se tenga. Dichos costos serán trabajados con una periodicidad mensual, ya que los registros contables se manejan con esta periodicidad y esto facilita la determinación de su valor.

8.1.1 Costos de mano de obra

La mano de obra es sin duda el recurso más valioso en la prestación de un servicio y también el más costoso. A nivel mundial, el 63% de las empresas del sector privado están pagando más en costos de personal que hace un año¹⁰⁷. La carga prestacional de poseer muchos empleados puede llevar a la quiebra a cualquier negocio, ya que dentro del costo unitario de producir un bien, alrededor del 80% representan la materia prima y la mano de obra¹⁰⁸.

Además de la compensación salarial, los costos asociados con la carga prestacional, los aportes parafiscales y los aportes a seguridad social hacen que un trabajador represente un alto costo para la empresa, de hecho cuando en cinco países latinoamericanos el costo de mano de obra subió 10%, su participación en el mercado de exportación se redujo en 2.5 puntos porcentuales¹⁰⁹,

Por esta razón, la mano de obra debe tenerse muy en cuenta a la hora de diseñar un sistema de servicio, pues sobredimensionar su cantidad genera tiempos ociosos y sobrecostos; y caer por debajo de lo necesario, aumenta el tiempo de espera, las cargas de trabajo y la eficiencia del sistema.

Teniendo en cuenta esto, uno de los principales parámetros a mejorar debe orientarse a la asignación eficiente del recurso humano. En la red de colas en serie definida, se considera una mano de obra homogénea y sin distinción en la compensación salarial de los empleados o servidores. El costo total de ésta es igual a:

$$C_{mo} = C_s S$$

¹⁰⁷ International Business Report (IBR). Disponible en internet:
<http://www.internationalbusinessreport.com/Reports/Focus-reports/2008_global_overview.asp>

¹⁰⁸ Banco Interamericano De Desarrollo. Disponible en internet:
www.iadb.org/res/publications/pubfiles/pubbb-2001s_7814.pdf

¹⁰⁹ Ibid.

En donde, C_s es la compensación salarial de los servidores y S el total de ellos en el sistema. Es importante aclarar que la compensación salarial incluye: salario básico, comisiones si aplica, prestaciones sociales, seguridad social, aportes parafiscales, afiliación a riesgos profesionales, entre otros.

8.1.2 Costos de utilización del sistema

Uno de los costos más importante a la hora de definir cómo prestar un servicio son las instalaciones físicas. Ray Croc, promotor de McDonalds, afirma que la ubicación de un establecimiento no solo es importante sino que es el factor que más influye en el éxito de un negocio¹¹⁰. Junto con el lugar se asocian muchos más costos que son derivados del mismo y que son necesarios para mantener el lugar de atención en las condiciones pertinentes y aptas para desarrollar la actividad que éste prevista.

Debido a la no caracterización de una oficina o punto en específico de atención en este estudio, se tomará un costo total mensual, que puede ser consultado en recibos, facturas etc., definido de la siguiente forma:

$$C_{utot} = C_{ar} + C_{adm} + C_{man} + C_{rh} + C_{as} + C_{im} + C_{sp}$$

Donde, C_{ar} es el costo de arrendamiento del lugar, C_{adm} representa el costo de administración, C_{man} el costo de mantenimiento, C_{rh} el costo de seguridad, C_{as} aseo, C_{im} impuestos y C_{sp} servicios públicos. Dicho costo está asociado con el número de personas posibles dentro del sistema (servidores (S) y capacidad de clientes (K) y es sensible a la variación de ellos.

Luego, para ver los costos que implica cambiar la capacidad o el número de servidores expresamos dicho costo por persona así:

$$C_u = \frac{C_{utot}}{(s+k)}$$

8.1.3 Costos de espera

La espera de clientes en el sistema está asociada con el tiempo ocioso que éstos pierden mientras esperan por ser atendidos. Evaluar este costo es un factor clave al momento de cuantificar el efecto que éste tiene en la eficiencia de un negocio. Existen empresas como la Aerolínea Jet Blue que como política de calidad paga a sus clientes que esperen un

¹¹⁰ Disponible en internet: espanol.business-opportunities.biz/.../como-elegir-un-local-comercial/

bono en su próximo vuelo¹¹¹; sin embargo, existen otras compañías en donde es muy difícil llegar a calcular dicho valor o costo debido a su naturaleza intangible.

La caracterización de estos costos dependerá de cada empresa y está definida según la función lineal expuesta por Hamdy Taha¹¹², en donde el costo de espera es proporcional al tamaño de la espera.

$$C_w = C_e L_q$$

Donde C_e es el costo de espera por cada cliente.

Por lo tanto, para determinar los costos de espera se puede recurrir a diferentes criterios de evaluación, que dependiendo del contexto, permitiría hacer tangibles los costos de espera. En otras empresas se podría calcular con base en el costo de oportunidad que implica dejar de vender un bien cuando el cliente abandona el sistema o podría igualmente determinarse mediante el cálculo de la disminución en la demanda de los productos debido al deterioro de la imagen corporativa que implica poseer grandes filas¹¹³.

8.1.4 Costos de puestos de trabajo

Cada servidor que sea contratado para el correcto desempeño de sus funciones debe contar con un apropiado puesto de trabajo, cuyo aditamento tiene un costo y mantenimiento mensual, expresado en recursos físicos, preparación e inducción al puesto y elementos de oficina, entre otros. Ligado a este costo se clasifican los costos tecnológicos que hoy en día son altos debido los monopolios en la generación de software.¹¹⁴

$$C_{wt} = C_{pt} S$$

Donde C_{pt} representa el costo mensual de tener un servidor en un puesto de trabajo. Este costo incluye: valor mensual de los equipos, valor mensual de licencias, valor de papelería, y causación mensual de los costos de organizar un puesto con silla, escritorio y demás artículos de oficina.

¹¹¹ Disponible en internet: <http://www.nytimes.com/2007/02/20/business/worldbusiness/20iht-jet.4656377.html>

¹¹² TAHA, Hamdy A. Investigación de operaciones. México: Prentice Hall, c2004. Séptima edición. p. 626.

¹¹³ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 911 – 917.

¹¹⁴ Disponible en internet: <http://www.margen.org/articu/tics.html>

8.1.5 Costos de oportunidad

Uno de los factores más importantes en una cola son los abandonos. Debido a la impaciencia o limitaciones de tiempo por parte del cliente existe la posibilidad de que éste decida abandonar el sistema, lo que para la empresa representa un costo de imagen, calidad y oportunidad al no poder vender y satisfacer las necesidades de un cliente. Este costo por lo tanto está determinado por los abandonos del sistema y se define así:

$$C_{op}=C_oAb$$

Donde C_p representa el costo de oportunidad que tiene un cliente que abandona el sistema, y Ab los abandonos por nodo.

8.1.6 Otros Costos

Finalmente, existen otros costos particulares según la empresa que pueden ser especificados. Por lo general, estos costos dependen de los clientes y de los servidores.

$$C_{ot}= C_{it}(S+K)$$

Donde C_{it} representa los costos indirectos por persona y otros costos relacionados con la prestación del servicio, por ejemplo costos indirectos de saturación de un local comercial como multas, daños materiales, desgaste de propiedad planta y equipo, etc.

8.2 MODELO DE OPTIMIZACIÓN

A partir de los rubros definidos anteriormente es posible construir un modelo de optimización con el fin de minimizar los costos de servicio y minimizar el número de personas en colas, y por consiguiente, los tiempos de espera de los clientes. Con este modelo se estará desarrollando una herramienta de apoyo para la toma de decisiones en la dirección de empresas de servicios que buscan mejorar la eficiencia y la calidad en sus sistemas de atención.

8.2.1 Definición de variables

Para establecer el modelo de optimización es necesario definir todas las variables y parámetros que interactúan en el sistema y que afectan su desempeño. Dentro de la configuración de los modelos y los factores que los afectan están: las tasas de llegada y de atención de los servidores, sus distribuciones de probabilidad, la capacidad de cada uno de los nodos en el sistema y el número de servidores en el sistema.

Dentro de esta lista la cantidad de servidores disponibles en el sistema y la capacidad son las variables, ya que si bien los demás factores afectan el objetivo de minimizar el tiempo de espera, éstos son valores que están por fuera de control de una empresa. Por ejemplo, el número de nodos en la red, es un factor que si bien puede ser modificado por las organizaciones solo sería posible con un estudio largo y minucioso de los procesos que en dicha red se realizan, además su modificación implica la transformación completa de un lugar de atención. Estas dificultades técnicas y económicas muestran que la variación de estos factores no es viable en horizontes de tiempo cercanos.

8.2.2 Funciones objetivo

Las empresas dentro de su proceso de mejora continua están atentas y ávidas de propuestas de cambio que permitan optimizar su operación, y en un centro de atención la mejor optimización esta en brindar el mejor servicio al menor costo posible.

Para cuantificar este objetivo, se proceden a formar ecuaciones que sirvan de guía para llegar a las soluciones óptimas del sistema.

La primera de las funciones objetivo es la descrita por Hamdy Taha¹¹⁵ en su libro de Investigación de Operaciones. En ésta se busca minimizar el costo total de un sistema de colas a partir de todos los costos de espera y servicio asociados. A esta ecuación se le adicionan otros costos identificados de utilización de sistema, puestos de trabajo y otros.

El resultado es la multiplicación del costo salarial de cada servidor por el total de servidores en cada uno de los nodos, además la multiplicación de los costos de utilización por la capacidad de cada nodo y finalmente el costo de espera por el número de personas en cola en cada uno de los nodos.

$$\min Z = (C_s + C_{pt}) \sum_{n=1}^{N_o} S_n + (C_u + C_{tl})(\sum_{n=1}^{N_o} K_n + \sum_{n=1}^{N_o} S_n) + C_e \sum_{n=1}^{N_o} Lq_n + C_o \sum_{n=1}^{N_o} Ab_n$$

Costo de servidores y puesto de trabajo
Costo de Capacidad y otros costos
Costo de Espera y Oportunidad

No obstante, aunque esta función objetivo es funcional no contempla situaciones en donde los costos de espera no son tangibles; de ser este el caso el resultado descontaría por completo el valor del Lq , cuyo resultado es vital en la optimización de un sistema de colas.

Lo ideal y recomendable en las compañías es realizar un ejercicio de evaluación de factores que permitan volver tangibles costos que no lo sean. A pesar de esto, si el último factor no es posible, adicional a la función objetivo actual se suma una función exclusiva para la minimización del tiempo de espera en cola que se deriva del número de personas en el sistema.

¹¹⁵ TAHA, Hamdy A. Investigación de operaciones. Op. cit., p. 626

$$\min Z = \sum_{n=1}^{N_o} Lq_n$$

Esta función objetivo se utilizará independientemente de si el costo de espera es tangible o no, y busca por un lado evitar que se elimine el efecto de Lq dentro de la optimización; y en el caso de que el costo de espera sea conocido, busca incorporar un peso mayor al número de personas en cola.

8.2.3 Restricciones

Las restricciones delimitan un marco de acción lógico y permisible de acuerdo con las necesidades de un problema. Para el caso de estudio éstas deben estar asociadas a las variables y el resultado de los objetivos.

8.2.3.1 Restricciones asociadas con los servidores

La primera de estas restricciones ve encaminada a restringir la cantidad de servidores máximos dentro del sistema. Esta restricción está relacionada con el máximo costo permitido por las organizaciones para la contratación de empleados y/o por las instalaciones físicas de atención que sólo permitan un número determinado de servidores.

$$S_n < S_{tn} \quad \forall n, \quad (1)$$

$$\sum_{n=1}^{N_o} S_n < \sum_{n=1}^{N_o} S_{tn} \quad (2)$$

8.2.3.2 Restricciones asociadas con la capacidad

La restricción de capacidad delimita un marco de referencia para evaluar hasta qué punto es posible ampliar la capacidad de un nodo.

$$K_n < K_{tn} \quad \forall n, \quad (3)$$

$$\sum_{n=1}^{N_o} K_n < \sum_{n=1}^{N_o} K_{tn} \quad (4)$$

8.2.3.3 Restricciones asociadas con los costos

Cada negocio posee un punto de equilibrio en donde el resultado del ejercicio periódico de producir un bien o un servicio no representa pérdidas ni ganancias¹¹⁶. Mediante el cálculo de este indicador es posible identificar cuál es el costo máximo que pueden soportar sin incurrir en pérdidas. Dicho valor está compuesto por la suma total de la

¹¹⁶ YOUNG Philip K. Y. Economía de empresa. Cuarta edición. Pearson Educación, 2004. P. 440.

nómina de los servidores, las cargas prestacionales, los aportes parafiscales, aportes a seguridad, el costo total de espera, el costo de mantenimiento del sistema, los costos del puesto de trabajo, tecnología y los otros costos referenciados.

$$C_{mo} + C_{utot} + C_w + C_{ot} + C_{wt} + C_o < C \quad (5)$$

8.2.4 Modelo Final

Luego de definir las variables y parámetros, construir las funciones objetivo y las restricciones, el modelo de optimización en su versión compacta es:

<p><u>Conjuntos:</u> N = Número de nodos</p> <p><u>Parámetros:</u> λ = Tasa de llegada μ = Tasa de atención por servidor K_n=Capacidad en el nodo n. S_t=Número total de servidores N_o = Número de nodos del sistema C = Costo máximo del sistema</p> <p><u>Variables:</u> S_n=Número de servidores en el nodo</p>	<p><u>Funciones objetivo:</u> $\min Z = (C_s + C_{pt}) \sum_{n=1}^{N_o} S_n + (C_u + C_{tl})(\sum_{n=1}^{N_o} K_n + \sum_{n=1}^{N_o} S_n) + C_e \sum_{n=1}^{N_o} Lq_n + C_o \sum_{n=1}^{N_o} Ab_n$ $\min Y = \sum_{n=1}^{N_o} Lq_n$</p> <p><u>Restricciones:</u> $S_n < S_{tn} \quad \forall n,$ $\sum_{n=1}^{N_o} S_n < \sum_{n=1}^{N_o} S_{tn}$ $K_n < K_{tn} \quad \forall n,$ $\sum_{n=1}^{N_o} K_n < \sum_{n=1}^{N_o} K_{tn}$ $C_{mo} + C_{utot} + C_w + C_{ot} + C_{wt} + C_o < C$ $N, K, S_n, S_t, N_o \in Z$ $\lambda, \mu, C \in R$</p>
---	--

8.3 SOLUCIÓN DEL MODELO

Con el fin de dar respuesta a este tipo de problemas donde se desea elegir la mejor combinación del sistema, se plantea el modelo de optimización cuya formulación muestra funciones objetivo que no pueden ser evaluadas por métodos matemáticos tradicionales, ya que por ahora no existen fórmulas que acoten los modelos aquí desarrollados.

Este tipo de problemas son complejos de resolver debido a la evaluación de las ecuaciones y al carácter incierto de los posibles valores que pueden tomar las funciones objetivo. Solo con la simulación de cada escenario se pueden evaluar las ecuaciones y encontrar los valores óptimos de las variables. La dificultad nace en que los sistemas de colas en general son problemas combinatorios donde el conjunto de soluciones es bastante grande y su evaluación es computacionalmente costosa (NP_hard), por lo tanto es necesario recurrir a métodos heurísticos de optimización.

Una metaheurística es un procedimiento para resolver problemas de alta complejidad matemática que bajo los procedimientos clásicos no se llegaría a una solución precisa en un periodo de tiempo prudente.¹¹⁷ Si bien este método busca soluciones muy buenas no necesariamente son las óptimas, pero si el algoritmo está bien diseñado eventualmente se alcanzarán soluciones muy cercanas a la óptima.

Si bien existen técnicas muy buenas de este tipo, es tiempo de explorar otros métodos de optimización empleando heurísticas de búsqueda que hagan uso de la teoría de juegos¹¹⁸. Por lo tanto, para la solución del actual modelo se desarrollará una metodología de optimización que involucre conceptos de toma de decisiones junto con algoritmos de búsqueda local.

8.3.1 Evaluación y selección de la heurística vía teoría de juegos

Las redes de colas como fueron modeladas en este estudio habrán de ser optimizadas con base en las dos funciones objetivo definidas: minimización del costo total (función Z) y minimización de Lq (función Y), cada una de las cuales posee un conjunto de objetos que se les denomina estados, y todos ellos juntos forman el espacio de búsqueda¹¹⁹.

La complicación de este modelo radica en que dentro de los conjuntos no hay elementos que produzcan optimalidad en ambas funciones objetivo de forma simultánea. El conflicto entre las funciones objetivo hará que la mejora en una función empeore la otra, es decir,

¹¹⁷ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 572

¹¹⁸ COELLO Carlos A.. Introducción a la Optimización Multiobjetivo Usando Metaheurísticas. En: Sección de Computación. Méjico Col. San Pedro Zacatenco. 2005.

¹¹⁹ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 572

que la mejora en el indicador de desempeño Lq solo será posible con el aumento de los costos de funcionamiento del sistema.

Para explicar y abordar esta situación se plantea el modelo en términos de un juego. Supóngase una organización con dos áreas: Calidad y Finanzas. Ambas dentro de su rol buscarán acciones con el fin de ejercer una mejor gestión y que ello se refleje en la productividad de la organización a la que pertenecen. Si el Presidente de esta organización preguntase en una reunión de ambas áreas sobre qué hacer para que su negocio sea mejor se encontraría con dos respuestas diferentes.

El Financista seguramente haciendo una exposición de las restricciones de la economía y los efectos de poseer una nomina elevada dentro de los costos de operación del negocio de atención y en el valor de la compañía misma, sugeriría la reducción de personal o bien sea no incrementar dicho recurso con el fin de valorizar la compañía y reducir las obligaciones que ésta posee.

El encargado de Calidad por su parte descartaría la solución anterior y propondría incrementar el nivel de atención con más trabajadores con el ánimo de generar un servicio de calidad que permita al negocio expandir su mercado y así mejorar sus utilidades.

Para el Presidente es un hecho que la calidad que ofrece en ese momento es inferior a la esperada en una compañía de su clase, pero también conoce las limitaciones de presupuesto y la necesidad de los accionistas de aumentar sus utilidades. Por lo tanto, propone a sus subalternos pensar en lo que le conviene más a la empresa.

Al interior de cada área se tendrán dos estrategias diferentes, una de ellas es presentar un plan A de acción convencidos de que su propuesta permitirá a la empresa mayor utilidad que la propuesta de la otra área incrementando sus indicadores internos de productividad al máximo, o presentar un plan B que si bien mejora sus indicadores estima un margen de acción de la propuesta para la otra área encargada.

Planteada la negociación de esta forma se conformaría un juego, en el cual cada área de la empresa tendría dos estrategias con una matriz de pagos definida por la mejora porcentual aportada por el área en los indicadores de productividad de toda la compañía.

La tabla 23 muestra el juego en forma matricial con las estrategias y los pagos (a,b) según los supuestos del juego. Si ambas áreas optaran por presentar el plan A, no se tomará ninguna decisión lo que hará que la mejora aportada por cada área sea nula, ya que las dos estrategias serán completamente opuestas. Si por el contrario, una de las áreas decide tomar el Plan A y la otra el Plan B, el aporte del área con el plan A a los indicadores de eficiencia de la compañía será una mejora del 10%, mientras que la otra no aportará nada. Finalmente, si ambas deciden presentar un plan B el aporte que cada área tiene en la mejora de los indicadores rodeará el 5%.

Tabla 23. Juego entre los objetivos.

		Finanzas		
C a l i d a d			Plan A	Plan B
	Plan A	(0, 0)	(10, 0)	
	Plan B	(0, 10)	(5, 5)	

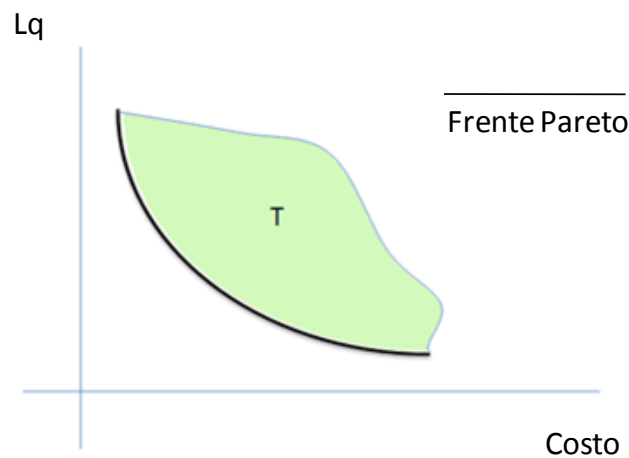
Fuente: Elaboración propia.

Si el problema de costos y calidad fuera estrictamente relacionado con el proceso de atención a clientes, este juego representaría el problema de optimización tal como se planteo en el numeral anterior. El vector (a,b) de mejoras estaría compuesto de elementos de la función Z y de la función Y respectivamente. El concepto de optimalidad es ahora relativo, ya que existen varias soluciones con resultados iguales pero con diferentes niveles de satisfacción. Por ejemplo, el resultado (10,0) y (0,10) son soluciones óptimas y con igual beneficio. Este concepto no es nuevo y se le conoce como eficiencia de Pareto. Si algo genera o produce provecho, comodidad, fruto o interés sin perjudicar a otro, provocará un proceso natural de optimización hasta alcanzar el punto óptimo¹²⁰.

Luego, una solución X formada por los elementos de las funciones objetivo Z y Y es un óptimo de Pareto si no existe otra solución que haga mejorar alguno de los objetivos respecto a los valores obtenidos para X sin que empeore de forma simultánea alguno de los otros. En general, la solución en el sentido de Pareto al problema de optimización multiobjetivo no será única y estará formada por todas las soluciones con mínimos en alguna función respecto a los valores de la otra.

¹²⁰ BUCHANAN Allen E. Ethics, efficiency and the market. Oxford University Press US, 1985. P. 4-8.

Figura 40. Espacio en R^2 de soluciones multiobjetivo.



Fuente: Elaboración propia

Por lo tanto, mientras que los costos puedan ser menores sin perjudicar el número de espera en colas se tienen situaciones no óptimas en el sentido Pareto. La línea que marca la frontera mínima de todas las posibles soluciones (f) en la figura 40 se le denomina frente Pareto, pues son todos los puntos óptimos Pareto dentro del problema.

Sin embargo, como se mostró anteriormente, dentro de dichos óptimos existe un desequilibrio en las funciones objetivo que castiga una de ellas con el fin de lograr mejores resultados en la otra. El objetivo entonces será llegar a un punto de satisfacción para ambos entes en el juego y en donde ambas funciones objetivo sean satisfechas.

Para cumplir con ello se debe buscar una solución en la cual no sea posible beneficiar más a un objetivo sin perjudicar al otro y que al mismo tiempo represente un equilibrio entre los objetivos propuestos inicialmente. Esta solución sería la combinación (5,5).

Debido a la gran cantidad de posibles combinaciones, la búsqueda de soluciones óptimas dentro de la frontera Pareto, se hará con un método de búsqueda local. Este algoritmo seleccionará dentro del universo de posibilidades las posibles soluciones cercanas a la óptimalidad de Pareto y luego de ello, seleccionará aquella que representen el mejor balance entre las funciones objetivo planteadas.

8.3.2 Diseño del Algoritmo

La búsqueda local es una técnica que proviene de la inteligencia artificial y está basada en una estructura de memoria adaptativa que le permite explorar un espacio de soluciones de manera eficiente a través de la estructura de vecindad.¹²¹ Debido a la complejidad de buscar dentro del conjunto de soluciones, es necesario hacer un recorrido que permita encontrar puntos cercanos a los óptimos Pareto y desde ahí evaluar la conveniencia o no de la solución.

8.3.2.1 Configuración inicial

Una configuración es el conjunto de variables enteras del problema dispuestas en un arreglo (vector compuesto por la configuración de los factores en una cola)¹²², y la configuración inicial es a partir de donde parte el algoritmo con el fin de mejorar su estado actual. La configuración inicial será un óptimo Pareto conocido. Para este caso, se tomará un servidor en cada uno de los nodos existentes, que si bien representa el ahorro máximo, no tiene en cuenta para nada los resultados del Lq.

8.3.2.2 Generación del vecindario

Un vecino de una configuración X es una configuración X' obtenida a partir de X. Para el problema de colas, un vecino se genera aumentando en una unidad el número de servidores en el nodo i y consecutivamente aumentando la capacidad.

8.3.2.3 Selección del mejor vecino

En la generación de la vecindad se introduce otro concepto usado en la investigación de operaciones y es la optimización dinámica. En ella se muestra que es posible descomponer un problema en etapas y a partir de los óptimos encontrados en una de ellas es posible buscar un óptimo en las siguientes etapas para conformar una solución óptima de todo el problema.¹²³

Si se buscara dentro del universo de posibilidades una configuración, existen muchas posibilidades, pero muchas de ellas no presentarían mejoría en el resultado buscado. Con la optimización por etapas se busca limitar el vecindario usando el conocimiento que se

¹²¹ HILLIER, Frederick S. y LIEBERMAN, Gerald J. Op. cit., p. 425

¹²² Ibid.

¹²³ Ibid.

posee del sistema para agregar inteligencia en la escogencia de soluciones probables dentro de la frontera Pareto.

De acuerdo a lo visto en el capítulo de análisis de resultados, el impacto que tiene el aumento de un servidor en las medidas de desempeño es más evidente si dicho aumento se da en los primeros nodos, esto quiere decir que a partir del óptimo de servidores en el primer nodo es posible encontrar óptimos en los nodos siguientes que dependan de la configuración óptima del nodo uno ya conocida.

En ese mismo capítulo también se evidenció cómo la variación de servidores y capacidades tienen un mayor impacto en L_q cuando se aumenta entre los primeros niveles; luego es de suponer que el impacto en las colas de la variación de uno a dos servidores es mucho más significativa que la variación de la cantidad cuando existen más servidores.

Esto permite concluir que en la medida que se avance en la búsqueda de soluciones óptimas en el sentido Pareto, se llegará a puntos donde la mejoraría comienza a hacerse menos notable; y a partir de ahí, sin importar el número de iteraciones adicionales, la frontera Pareto ya estará definida.

8.3.2.4 Evaluación de los vecinos

Un vecino que sea generado a partir de una posición X no puede ser seleccionado nuevamente a no ser que sea dominante sobre las configuraciones anteriores. El criterio de optimalidad Pareto dice que X es una solución compuesta por el vector de funciones objetivo (z, y) si y sólo si no existe otro vector que lo domine.

Un vector $x = (x_1; x_2)$ domina a otro $x' = (x'_1; x'_2)$ si y sólo si x es parcialmente menor a x' .

$$\forall i \in \{1,2\}, x_i \leq x'_i \wedge \exists i \in \{1,2\} | x_i < x'_i$$

Con este criterio se recorrerá el espacio de búsqueda para encontrar dominancias que conduzcan a los óptimos Pareto o, por lo menos, a resultados cercanos a éste donde sea posible seleccionar el equilibrio entre las funciones objetivo.

Luego de encontrar la probable frontera Pareto viene el dilema de cual de dichas soluciones representa un equilibrio entre las funciones. Es entonces donde se regresa al juego planteado: ¿qué solución haría simultáneamente que ambas áreas (fianzas y calidad) estuvieran satisfechas? Para dicho ejemplo seria el punto (5,5), pero ampliando mas el espectro de posibilidades y orientados hacia el comportamiento racional de un individuo se habrá de suponer que el equilibrio entre los intereses de las áreas solo será posible en la medida que una solución represente beneficios equivalentes. Para tal fin y debido a la incompatibilidad de las unidades entre las funciones objetivo se recurre a un

método, que si bien fue concebido en otro contexto, resulta útil para la solución de este juego y es la elasticidad.

En economía, la elasticidad es la razón formada entre el cambio proporcional de una variable con respecto al cambio proporcional de otra¹²⁴. Matemáticamente hablando, la elasticidad es el cambio proporcional de una variable dependiente (función de costo) relativamente a otra variable independiente (Lq):

$$E = \frac{\Delta y/Y}{\Delta z/Z}$$

Enfocando el concepto y la fórmula al modelo, si dicho indicador es mayor o menor que la unidad representa que la variación de una de las funciones objetivo de una solución óptima Pareto a otra tiende a beneficiar más a un área; y si por el contrario, la elasticidad es unitaria nos encontramos ante un equilibrio en los beneficios que representa el punto óptimo de satisfacción que se busca.

Adicional a este planteamiento, puede que las directivas del centro de servicio crean que el peso de alguna de las dos funciones es más importante que la otra, por lo cual se le agrega al modelo un parámetro α , el cual asignará un peso a las funciones objetivo para escoger el equilibrio entre ellas según el criterio seleccionado.

De acuerdo a esto, la fórmula de evaluación de un óptimo Pareto X (z, y) frente a otro será:

$$E = \frac{\Delta y Z \alpha}{\Delta z Y (1 - \alpha)}$$

El óptimo que generará el modelo será la solución cuyo cambio posea una elasticidad más cercana a uno.

8.3.2.5 Criterio de parada

El algoritmo puede buscar y seguir escogiendo vecinos indefinidamente, hasta recorrer todo el problema. Sin embargo, en problemas grandes y con muchas posibilidades de solución esta tarea puede tornarse lenta y arrojar resultados en tiempos bastante grandes. Con el objetivo de permitir la convergencia del resultado en un tiempo prudente y con una buena aproximación de lo que sería un óptimo se establece un criterio de parada.

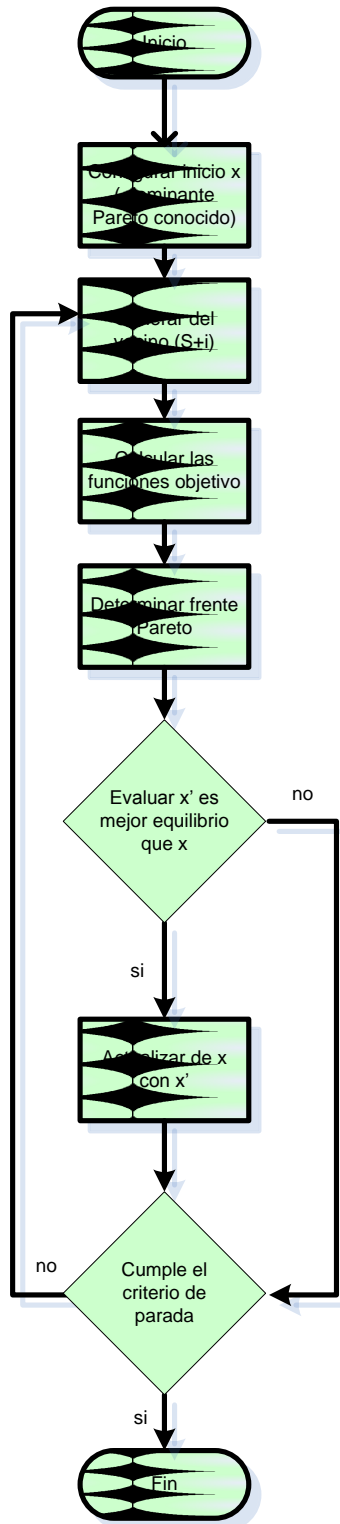
¹²⁴ PARKIN Esquivel, Michael. Microeconomía. Tercera Edición. P. 86.

Como se vio en el capítulo 8, la variación de los servidores y la capacidad tienen un impacto mayor si existen menos servidores en el sistema y la capacidad es más limitada; luego, al ir aumentando el número de valores o iterando entre los vecinos para encontrar soluciones, los resultados van a perder cada vez mayor impacto en los indicadores de desempeño.

Por lo tanto, el criterio de parada está determinado por sucesivos intentos de mejora sin éxito alguno. Si en las iteraciones siguientes, luego de encontrar un óptimo local, no ocurre mejoría y esto se prolonga por al menos $K/10$ (número de niveles posibles de capacidad dentro del problema) iteraciones, el algoritmo detiene su proceso de búsqueda, ya que está en un punto cercano al óptimo, o por lo menos, uno mucho mejor que sus predecesores.

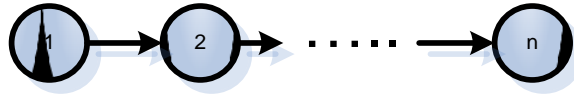
8.3.3 Funcionamiento

En resumen, el diseño del algoritmo aquí desarrollado consta de las siguientes etapas:



Dicho procedimiento se completará en cada una de las etapas que estarán representadas por los nodos.

Figura 41. Fraccionamiento de la red en nodos.



Fuente: Elaboración Propia.

Para la elaboración de este algoritmo se desarrolla un código en Visual Basic el cual se encuentra en el anexo G.

8.3.4 Evaluación

Dentro de este tipo de heurísticos que usan la frontera Pareto como criterio de optimización se han desarrollado medidas de dispersión u error con el fin de evaluar que tan aproximado es el algoritmo en la obtención de los puntos que cumplen con el criterio Pareto.

Una de dichas medidas fue desarrollada por Jason Schott¹²⁵ con el fin de medir la varianza de la distancia de cada miembro del conjunto óptimo Pareto con respecto a su vecino más cercano.

$$D = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{d} - d_i)^2}$$

Donde $d_i = \min (|z_i - z_j| + |y_i - y_j|)$, $i, j = 1, 2, \dots, n$; \bar{d} es el promedio de todos los d_i y n es el número de elementos del conjunto de Pareto obtenidos hasta el momento. Si $D=0$ significa que el algoritmo ha encontrado la distribución ideal de vectores no dominados.

8.4 EJEMPLO

Para hacer un detalle del cómo funciona el algoritmo junto con los resultados que arroja, se busca desarrollar un ejemplo ilustrativo.

¹²⁵ JASON Schott. Single multicriteria genetic algorithm. En: Departamento de tecnología de estado de Massachussets. 2005.

Los parámetros de entrada al modelo se determinan haciendo que el máximo nivel del experimento de este estudio represente las restricciones del modelo de optimización, con el fin de hacer equivalentes los resultados y vislumbrar la frontera Pareto calculada con relación al universo de posibilidades del problema. Los parámetros, costos y demás se especifican a continuación:







Parametros:

Lamda (λ) = 0.52 *Personas/ hora*
Miu (μ) = 0.16 *Personas/ hora*
Cantidad de Nodos: 3 Nodos
Distribucion t entre llegadas: Exponencial

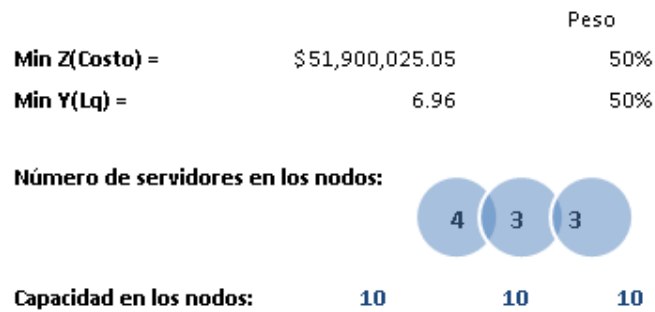
Costos:

C_s	\$	500,000
C_{utot}	\$	10,000,000
C_e	\$	9,000
C_{pt}	\$	100,000
C_o	\$	12,000
C_{tl}	\$	100,000

Restricciones:

C	\$	10,000,000,000.00
S_{t1}	10	
S_{t2}	10	
S_{t3}	10	
Capacidad en el	1	<u>30</u> 
nodos:	2	<u>30</u> 
	3	<u>30</u> 

El resultado del algoritmo es:



La varianza entre los puntos de la frontera Pareto según la fórmula anterior es:

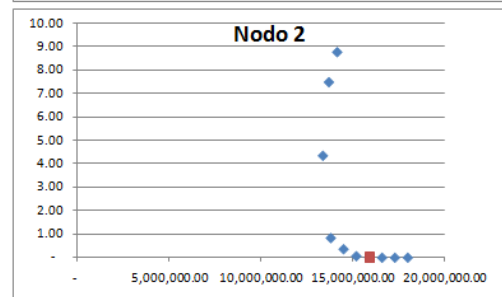
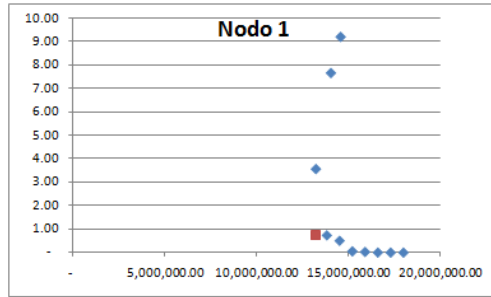
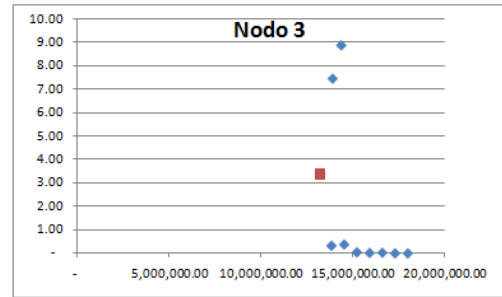
$$D= 0,16 \text{ (promedio para cada nodo)}$$

Este valor de D dice que es una buena aproximación del algoritmo a la optimización. Para conocer detalles del cálculo ver anexo F.

La figura 42. Presenta un cuadro con los resultados del algoritmo, la tabla ahí expuesta clasifica según el número de servidores los resultados de las funciones objetivo y además gráfica la frontera Pareto identificada junto con el óptimo (punto rojo) seleccionado.

Figura 42. Frontera Pareto por nodos en una red de colas

	Nodo 1		Nodo 2		Nodo 3	
1.00	14,563,160.94	9.24	14,158,972.59	8.77	14,339,780.14	8.86
2.00	14,029,246.66	7.69	13,707,423.04	7.49	13,867,024.88	7.45
3.00	13,212,214.62	3.58	13,379,171.16	4.35	13,190,426.46	3.38
4.00	13,806,654.76	0.74	13,807,534.76	0.84	13,802,930.34	0.33
5.00	14,504,552.87	0.51	14,503,250.25	0.36	14,503,397.42	0.38
6.00	15,200,479.78	0.05	15,200,578.40	0.06	15,200,411.88	0.05
7.00	15,900,258.71	0.03	15,900,009.77	0.00	15,900,162.64	0.02
8.00	16,600,009.92	0.00	16,600,000.00	-	16,600,262.60	0.03
9.00	17,300,000.00	-	17,300,000.00	-	17,300,025.05	0.00
10.00	18,000,000.00	-	18,000,000.00	-	18,000,000.00	-



Fuente: Elaboración Propia

9 HERRAMIENTA DE APLICACIÓN

El desarrollo y análisis que sustenta este estudio se basa en la teoría de los diversos conocimientos de la investigación de operaciones e ingeniería que pueden no ser totalmente conocidos o dominados por una persona a cargo de la administración de un sistema de colas. Con el fin de que tanto el análisis como la metodología de solución sean útiles para la toma de decisiones a nivel empresarial, se elabora una herramienta que simplifique y reúna los supuestos y procedimientos en un aplicativo sencillo, de fácil entendimiento y fácil uso. A través de este aplicativo el usuario podrá optimizar su sistema de colas según los parámetros, variables y características propias de su empresa, siempre y cuando éstas se puedan acoplar al alcance de este estudio.

9.1 GENERALIDADES

9.1.1 Construcción

La herramienta está construida en el lenguaje de programación de Visual Basic para poder conectar el software de simulación Rockwell Arena con la hoja de cálculo Microsoft Excel. Ver anexo G. Código de la herramienta.

9.1.2 Información Utilizada

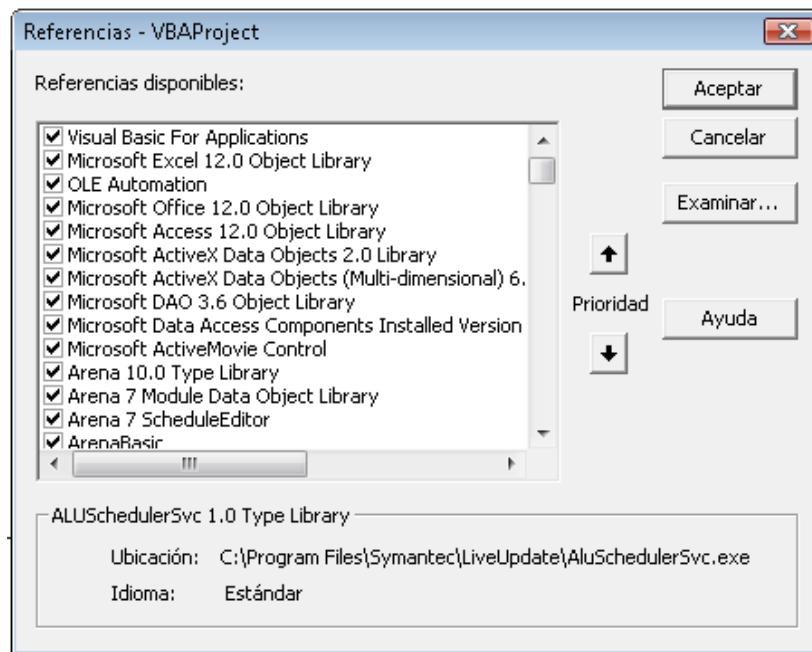
La herramienta involucra el uso de cuatro archivos en Arena y dos en Excel. El primer archivo en Arena es la interface del aplicativo como tal (Optimización.doe); y los tres restantes son las redes de uno, dos y tres nodos a simular con la información de entrada (nodo 1, 2 y 3 finito.doe). Por otra parte, los archivos de Microsoft Excel son para imprimir los resultados de la optimización (Macro Base.xslm) y para conservar los datos de la simulación (Resultados.xslm).

9.1.3 Requerimientos de Funcionamiento

Las características necesarias a tener en cuenta para ejecutar satisfactoriamente el aplicativo son:

- Tener Microsoft Office 2007
- Instalar el programa de simulación Rockwell Arena versión 10 o superior, preferiblemente con licencia comercial. No obstante, la versión académica puede ser útil si el número de entidades dentro del sistema es inferior a 150.
- Configurar el formato decimal del sistema con punto (ejemplo: 3.05)
En caso de tenerlo con coma, dirigirse al *Panel de Control/Configuración Regional y de Idioma/Personalizar Formato*.
- Chequear la activación de los complementos de bases de datos, hoja de cálculo y modelos de Arena, entre otros, en de las librerías y referencias de Arena y Excel (ver figura 43).
- Ubicar los seis archivos en la misma carpeta.

Figura 43. Listado de referencias VBA Project

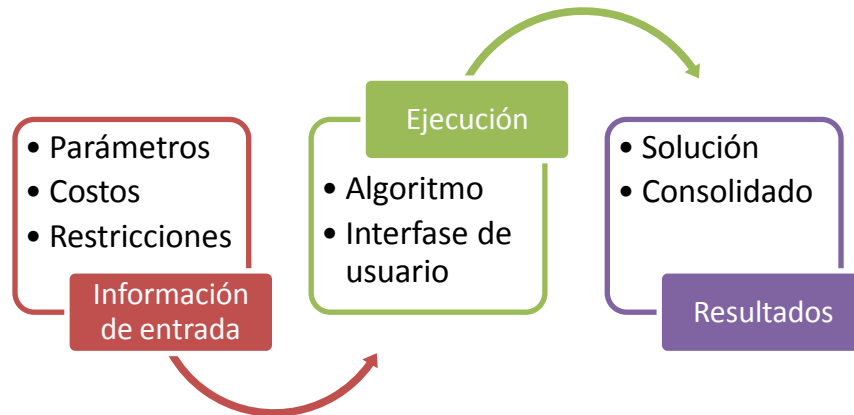


Fuente: Elaboración Propia.

9.1.4 Funcionamiento de la Herramienta

El aplicativo se basa en un proceso de alimentación de información, calculo interno de resultados y presentación de ellos al usuario. Este procedimiento se muestra en la figura 44.

Figura 44. Esquema de la herramienta



Fuente: Elaboración Propia.

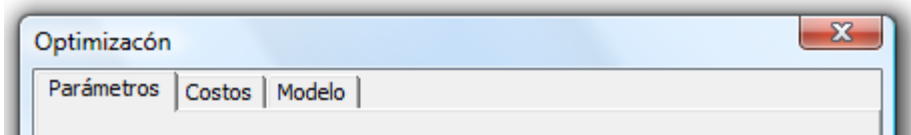
9.2 DISEÑO DE LA HERRAMIENTA

El aplicativo cuenta con una interfaz amigable y de fácil entendimiento que permite al usuario interactuar rápidamente para optimizar su sistema una vez conoce la información de entrada. Con la idea de simplificar y armonizar el aplicativo se diseñaron una serie de ventanas, formatos y formularios que sean útiles para el ingreso de las entradas y para el entendimiento de las salidas.

9.2.1 Información de entrada

El usuario de la herramienta debe suministrar ciertos datos de entrada que emulen su sistema de colas. Para el correcto funcionamiento del algoritmo y por cuestiones de diseño, dicha información se divide en tres pestañas independientes: Parámetros, Costos y Modelo. A continuación se describe cada una de ellas.

Figura 45. Pestañas del menú de entrada.



Fuente: Elaboración Propia.

9.2.1.1 Parámetros

En esta pestaña se deben ingresar los parámetros del modelo de optimización. Algunos de ellos están limitados por los niveles del diseño del experimento del presente estudio y otros son de ingreso libre al depender de cada empresa, como las tasas de llegadas y de servicio.

Figura 46. Formulario de Parámetros

A screenshot of the 'Optimización' window, showing the 'Parámetros' tab. The window has a title bar with a close button (X). The tabbed interface shows 'Parámetros', 'Costos', and 'Modelo' tabs, with 'Parámetros' selected. The main area contains several input fields and dropdown menus:

- 'Lamda (λ)': A text input field.
- 'Miu (μ)': A text input field.
- 'Unidades': A dropdown menu.
- 'Distribución tiempo entre llegadas:': A dropdown menu.
- 'Cantidad de Nodos:': A dropdown menu.

At the bottom right of the window, there is a note: '*Recuerde ingresar los decimales con punto'.

Fuente: Elaboración Propia

9.2.1.2 Costos

Los costos de un sistema deberán introducirse en un formulario con las mismas unidades y teniendo en cuenta una periodicidad mensual.

Figura 47. Formulario de Costos

Optimización

Parámetros | Costos | Modelo

Costo de un servidor (C_s) =

Costo de utilización Total (C_{utot}) =

Costo de espera por cliente (C_e) =

Costo de un puesto de trabajo (C_{pt}) =

Costo de oportunidad por cliente (C_o) =

Otros costos de servidores y clientes (C_{tI}) =

* No deje ningun espacio en blanco. Ingrese cero (0)
* No utilice separadores de miles

Fuente: Elaboración Propia

9.2.1.3 Modelo

Es esta pestaña se introducen los parámetros que definen las restricciones y la importancia de las funciones objetivo. Aquí se encuentra el boton de *Optimizar* que se debe pulsar una vez toda la información haya sido registrada en el sistema. Si no se ha registrado toda la información y se pulsa el boton de Optimizar. Es necesario para el correcto funcionamiento del modelo introducir los valores faltantes o desconocidos como ceros en el modelo, de no ser así pueden ocurrir errores de procesamiento impidiendo la solución del modelo.

Figura 48. Formulario del Modelo

The screenshot shows a software window titled 'Optimización' with three tabs: 'Parámetros', 'Costos', and 'Modelo'. The 'Modelo' tab is active. It contains two main sections: 'Funciones Objetivo' and 'Restricciones'. In the 'Funciones Objetivo' section, there are two input fields: 'Z = Costo' and 'Y = Personas en cola'. To the right of these fields is a 'Peso (Alpha)' input field and an 'Optimizar' button. The 'Restricciones' section includes a 'Costo Máximo Total' input field, and two rows of constraints: 'Máximo de Servidores en nodos' and 'Maxima capacidad en los nodos'. Each of these two rows has three input fields corresponding to nodes 1, 2, and 3.

Fuente: Elaboración Propia

Es importante tener en cuenta que la capacidad determina el número de personas que es posible atender en cada nodo, por lo tanto, dicho parámetro está definido por una variable entera positiva que debe ser múltiplo de 10 según los niveles del experimento. De introducir un número que no sea múltiplo de 10 el modelo pedirá re-ajuste en los valores introducidos ya que este supuesto es fundamental para el funcionamiento correcto del algoritmo.

9.2.1.4 Información de salida

La información de salida es la configuración óptima de servidores y capacidad de cada uno de los nodos, junto con información detallada de las funciones objetivo y soluciones dominantes según el criterio de optimización.

Figura 49. Solución del modelo

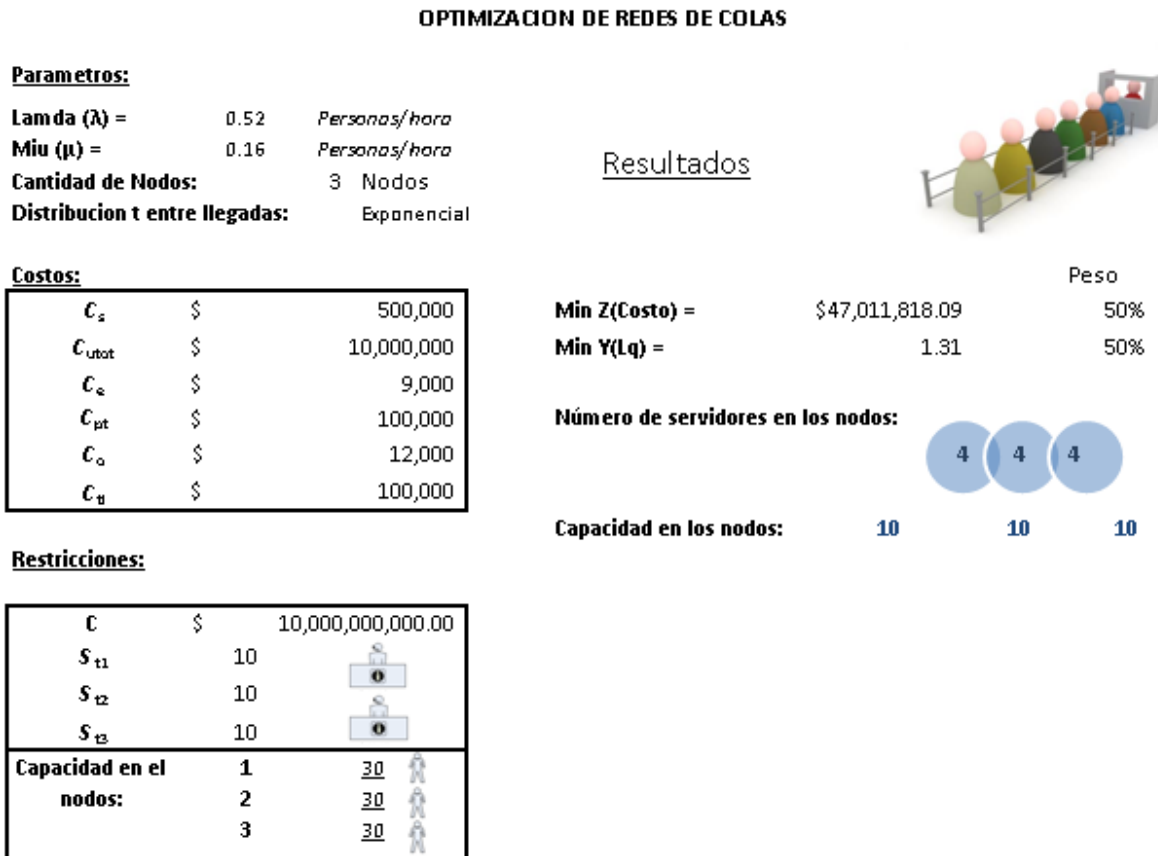
The screenshot shows a window titled 'Solución' displaying the results of the optimization. It lists the values for the objective functions Z and Y, and the optimal configuration for three nodes (1, 2, and 3) in terms of the number of servers and capacity.

Z:	1815.44865735	Nodos:	1	2	3
Y:	8.84828857847	Servidores	4	1	7
		Capacidad	10	20	10

Fuente: Elaboración Propia

Se guardarán los resultados y en el archivo Resultados.xlsm se encontrará información clave del modelo y solución del mismo. Este archivo se abrirá automáticamente una vez termine de correrse el modelo de simulación.

Figura 50. Información impresa generada en el archivo aplicativo.xlsx



Fuente: Elaboración Propia.

9.3 USO APROPIADO DE LA HERRAMIENTA

Después de verificar y asegurarse de la apropiada configuración del sistema, la instalación de los programas necesarios y la localización de los archivos de la herramienta, se podrá hacer uso del aplicativo siguiendo los siguientes pasos:

9.3.1 Verificar Compatibilidad

El modelo tiene características y metodologías que se usaron pensando en un análisis teórico y la posibilidad de una aplicación práctica no especificada, por lo tanto, el modelo puede o no ajustarse a un sistema de colas empresarial real. Para verificar la compatibilidad del modelo, es necesario asegurarse que la aplicación empresarial cuenta con las siguientes características:

- Red de colas en serie de uno, dos o tres nodos.
- Distribución de probabilidad entre llegadas Exponencial o Híper-Erlang.
- Distribución de probabilidad de tiempos de atención Exponencial.
- Capacidad restringida del sistema en unidades múltiplos de diez.
- Tasas iguales de atención de todos los servidores pertenecientes al mismo nodo de servicio.
- Disciplina de atención FIFO en todos los nodos.

9.3.2 Recolección de la información

Si el modelo es compatible, el usuario puede proceder a recolectar la información necesaria para simular el sistema, como lo es la tasa de llegadas, la tasa de atención, los costos asociados con la operación, las limitaciones de espacio y el dinero máximo disponibles a invertir para optimizar el sistema de colas.

9.3.3 Procedimiento de uso

Para usar la herramienta se deben seguir los siguientes pasos:

- **Instalar herramienta:** Se deben instalar los programas Arena 10 y Excel 2007, junto con los complementos que necesita el programa (ver figura 43). Adicionalmente, se deben guardar los archivos Optimización.doe, nodo1finito.doe, nodo2finito.doe, nodo3finito.doe, Resultados.xlsm y Macro Base.xlsm en una misma carpeta.

- **Ejecutar el aplicativo:** Abrir el archivo Optimización.doe y al iniciar el programa presionar *Ctrl+o*. Inmediatamente se desplegará la interface para introducir la información de entrada. Posterior a esto, oprimir el botón Optimizar
- **Revisar resultados:** Una vez termina de correr el algoritmo, se abre automáticamente el archivo Resultados.xlsm, en donde se presenta un resumen del modelo y los resultados.
- **Realizar nueva optimización:** Si se quiere correr una nueva optimización se debe acceder a los archivos donde se encuentran las redes y inicializar su configuración con sólo un servidor en cada nodo.

10 CONCLUSIONES

- La elección de la distribución de probabilidad de los tiempos entre llegadas debe ser bien estudiada, ya que ésta determina sustancialmente el comportamiento del sistema y la medición de su desempeño.
- En sistemas de colas gobernados por tiempos entre llegadas con alta variabilidad, la distribución Híper-Erlang permite realizar un modelamiento más acertado que lo que la exponencial permite.
- La distribución Híper-Erlang demostró tener un impacto notable en el modelamiento de las colas, su uso es recomendable cuando los sistemas presentan alta variación de los tiempos entre llegada.
- El comportamiento que tiene las variables respuesta (L_q y W_q) según las combinaciones entre los factores es el mismo; si bien las cantidades son distintas se conserva los mismos patrones y tendencias.
- Existen cambios en los valores de L_q y W_q superiores al 100% al realizar comparaciones según el nivel de utilización entre la distribución Híper-Erlang con respecto a la Exponencial, por lo que su elección se convierte en un factor fundamental para el modelamiento eficiente de un sistema.
- Los cambios en los factores que inciden en un sistema de colas generan un impacto cada vez menor conforme éstos aumenten dentro de sus niveles. Por ejemplo, el impacto de aumentar un servidor tiene más efecto entre menos servidores existan en el sistema; al cambiar de uno a dos servidores, el cambio de L_q es de 45%, mientras que al pasar de 7 a 10, solo cambia un 2% cuando los tiempos entre llegadas son Híper-Erlang.
- Los factores que más inciden en el desempeño de las redes de colas son las distribuciones de probabilidad entre los tiempo de llegada, el número de servidores en los nodos -especialmente en el nodo uno- y la capacidad del sistema con niveles de utilización cercanos a uno.

-

- Los sistemas con distribuciones Híper-Erlang presentan una desestabilización temprana a lo largo de los nodos aun cuando el nivel de utilización es inferior a 1.
- Para niveles de utilización inferiores a 0.7, la capacidad no afecta el desempeño del sistema.
- El aumento del número de servidores con niveles de utilización inferiores al 0.7 no tiene un efecto significativo en la variación de las medidas de desempeño del sistema.
- Las redes de colas en serie presentan cuellos de botella en el primer nodo. De ahí en adelante es visible una disminución de las colas cuya variación entre nodos oscilan entre el 30% para la distribución Híper Erlang y el 19% para la exponencial.
- Los resultados para ambas distribuciones difieren en cuanto a los valores de las medidas de desempeño, pero en general, el efecto de los factores de influencia a lo largo de todas las combinaciones muestra comportamientos y tendencias que resultan similares para ambas distribuciones. Esto lleva a pensar que dichos comportamientos pueden ser propios de todos los sistemas, independiente del tipo de distribución que se use para modelar los tiempos entre llegadas.
- La minimización de costos sólo es posible mediante su cuantificación, debido a la existencia de costos intangibles los modelos multi-objetivos parecen mediar entre los conflictos de costos y personas o tiempo en cola.
- La optimización mediante un juego permite descontar dificultades de análisis por falta de información y ejecutar un recorrido inteligente del sistema mediante los patrones observados. Incluir dicho método como medio de evaluación de optimización permite llegar a soluciones equilibradas y eficientes.
- Los modelos de optimización en redes de colas deben enfocarse a liberar la congestión del primer nodo y velar por un equilibrio en la capacidad de atención según las llegadas en cada punto de atención del sistema.
- Bajo las mismas condiciones, un sistema gobernado por llegadas Híper-Erlang necesita más recursos para su optimización que uno con tiempos exponenciales.
- Con el desarrollo de la simulación es posible conocer el comportamiento de las redes de colas según la combinación de los factores que influyen en ella y a partir de ese conocimiento alimentar un modelo de optimización con el ánimo de asignar eficientemente los recursos.

12 RECOMENDACIONES

- El análisis de las redes de colas puede extenderse hacia otro tipo de distribuciones de probabilidad para modelar el tiempo entre llegadas y de atención, ya que dichas variaciones pueden modificar el comportamiento aquí descrito.
- Para conocer la fuente de la variación de las tasas de nodo a nodo puede ser conveniente estudiar los abandonos que se presentan a lo largo del sistema y relacionarlos con el cambio en la tasa de entrada a los nodos dentro de la red.
- Se recomienda realizar el análisis de simulación con una distribución de probabilidad de los tiempos de atención Híper-Erlang para así ver la influencia que tiene esta variación en las medidas de desempeño y conclusiones de este estudio.
- Para la robustez de la herramienta de optimización es necesario estudiar las redes de colas bajo diferentes parámetros con el fin de continuar abasteciendo este algoritmo con la lógica necesaria para su funcionamiento en una mayor cantidad de casos.

13 BIBLIOGRAFÍA

ALFARO, Víctor M. Modelado y Análisis de los Sistemas Dinámicos utilizando la Red Generalizada. En: Escuela de ingeniería electrónica Costa Rica. (2003); p. 25.

ALVARADO Vera, Sandra. Diseño de una estrategia para el mejoramiento de la calidad del servicio de talleres por medio del entrenamiento técnico tomando como base de análisis y estudio a General Motors Colmotores. Trabajo de grado.(2004). p. 60-70.

Análisis Cuantitativo con WINQSB. Disponible en la web:
< <http://www.eumed.net/libros/2006c/216/1j.htm>>

ANDERSON David R. Métodos cuantitativos para los negocios. Soluciones empresariales. Cengage Learning Editores, 2004. p. 618.

BARRAS, J. S. y MAKOWSKI, D. J. K. Competing queues with geometric service requirements and linear costs: the μc -rule is always optimal. En: System and control letters. Vol. 6 (1985); p. 207.

BURKE, P. J. Proof of a Conjecture on the Interarrival-Time Distribution in an M/M/1 Queue with Feedback. En: IEEE Transactions On Communications, Vol. 29 (1976); p. 589.

_____. The Output of a queuing system, En: Operations research, Vol. 4 (1956); p. 704.

CAO Ricardo , ABAD Ricardo , Introducción a la Simulación y a la Teoría de Colas. Editor Net biblo, (2002). P.103 -110

COCHRANA, Jeffery K. y ROCHEB, Kevin T. A multi-class queuing network analysis methodology for improving hospital emergency department performance. En: Computers & Operations Research. Vol. 36 (2009) p. 138.

COHEN, William A. Plan de mercadotecnia. Primera edición. México: Compañía Editorial Continental: CECSA, 2004. p.400.

DANE. Disponible en Internet:
<http://www.dane.gov.co/index.php?option=com_content&task=category§ionid=15&id=54&Itemid=236>

DEMBO, Ron S.; MULVEY, John M. y ZENIOS, Stavros A. Large-Scale Nonlinear Network Models And Their Application. En: Operations Research. Vol. 3 (1989); p. 353.

DÍAZ REDONDO, Rebeca P; PAZOS ARIAS, José Juan y SUÁREZ GONZÁLEZ, Andrés. Teoría de colas y simulación de eventos discretos. Madrid: Pearson, 2003. p. 309.

DOWNB, Bonalda D. Stability of mixed generalized Jackson networks. En: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Vol. 7 (1998); p. 56.

ESCUADERO, Laureano F. Aplicaciones de la teoría de colas: análisis y solución de los problemas de cuellos de botellas en almacén, producción y ventas con referencias breves a Business Games y teoría de juegos. Primera edición. Bilbao: Deusto, 1972. p. 259.

EVANS, Michael J. Probabilidad y estadística Editor Reverte, (2005) .p.675.

FELDMAN, Zohar; MANDELBAUM, Avishai; MASSEY, William A. y WHITT, Ward. Staffing of Time-Varying Queues to Achieve Time-Stable Performance. En: Management Science. Vol. 54 (2008); p. 23.

Fondo Monetario Internacional. Disponible en Internet:

<<http://www.docstoc.com/docs/1014168/Vuelve-Japn---Daniel-Citrin-y-Alexander-Wolfson--Finanzas-y-Desarrollo---Junio-de-2006>>

GÓMEZ, Emilio y SARABIA, José María. La distribución binomial-exponencial truncada con aplicaciones en el sector del seguro de automóviles. En: IAE, Vol. 14, (2008); p. 3-20.

GONGA, Qiguo y SHOUYANG, Laib. Supply chain networks: Closed Jackson network models and properties. En: Int. J. Production Economics, Vol. 113 (2008); p. 574.

GREEN, Linda y KOLESAR, Peter. The Pointwise Stationary Approximation for Queues with Non stationary Arrivals. En: Management Science; Vol.37 (1991); p. 84.

GROSS, Donald y HARRIS, Carl M. Fundamentals of queueing theory. Tercera edición. New York ; Chichester : John Wiley, 1998. p. 441.

HAMDY A. Taha. Investigación de operaciones. Méjico: Prentice Hall. ed. 7. (2004). P. 649 - 652.

HILLIER, Frederick S. y LIEBERMAN, Gerald J. Investigación de operaciones. Séptima edición. México: McGraw-Hill, 2002. p. 1123.

_____. Métodos cuantitativos para administración: un enfoque de modelos y casos de estudio, con hoja de cálculo. México; Buenos Aires : McGraw-Hill/Irwin, 2002. p. 855.

HONGXIA, Zhou. WANG, Junfeng. XU, Fanjiang y LI, Lei. Hyper-Erlang Based Model for Network Traffic Approximation. En: China Postdoctoral Science Foundation, (2005).

HOOKER, John Norman. Integrated methods for optimization. New York: Springer, 2007. P. 486.

INTERNATIONAL BUSINESS REPORT (IBR). Disponible en internet:
http://www.internationalbusinessreport.com/Reports/Focus-reports/2008_global_overview.asp

JACKSON, J. R. Jobshop-like queuing systems. En: management science, Vol. 10 (1963); p. 142.

_____. Networks of waiting lines. En: Operation Research. Vol. 5 (1957); p. 581.

JAMHOURA, Edgard; PENNA, Manoel; NABHEN, Ricardo y PUJOLLE, Guy. Modeling a multi-queue network node with a fuzzy predictor. En: Fuzzy sets and systems. (2007); p. 58.

J. M. Juran. Juran y la calidad por el diseño. (1996). P. 75.

KAMLESH, Mathur y SOLOW, Daniel. Investigación de operaciones: el arte de la toma de decisiones. Primera edición. México; New York: Prentice-Hall Hispanoamericana, 1996. p. 977.

KELTON, W. David. Simulación con software Arena. 4a ed. México; Bogotá: McGraw-Hill Interamericana Editores, 2008. p. 631.

KERBACHE, Laoucine y MACGREGOR, James. Queueing networks and the topological design of supply chain systems. En: J. Production Economics. Vol. 91 (2004); p. 278.

KECECIOGLU, Dimitri. Reliability and Life Testing Handbook. Prentice Hall, Inc. Volume I, New Jersey, (1993).

KIM, Jeongsim y KIM, Bara. Regularly varying tails in a queue with discrete autoregressive arrivals of order p. En: Queueing Syst. Vol. 56 (2007); p. 253.

KO, Hyun Jeung; KO, Chang Seong y TAIOUN Kim. A hybrid optimization/simulation approach for a distribution network design of 3PLS. En: Computers & Industrial Engineering. Vol. 50 (2006); p. 586.

KOTLER Philip, ARMSTRONG Gary. Fundamentos de marketing. Pearson Educación, (2003). P.312.

KUEHL, Robert. Principios estadísticos para el diseño y análisis de investigaciones. Segunda Edición Thomson, (2000). p. 20-225.

KURODA M. y KAWADA A. Adaptive input control for job-shop type production systems with varying demands using inverse queueing network analysis. En: Int. J. Production Economics. Vol. 41 (1995); 356.

LOVELOCK, Christopher H. Mercadotecnia de servicios. Primera edición. México: Prentice-Hall Hispanoamericana, 1997. p. 661.

MALHOTRA, Manish. y REIBMAN, Andrew. Selecting and implementing phase approximations for semi-

markov models. En: Stochastic Models, Vol. 9, No. 3 (1993); p. 473-506.

MAZUMDAR, Ravi; MASON, Lome G. y DOULIGERIS, Christos. Fairness in Network Optimal Control: Optimality of Product Flow Forms. En: IEEE Transactions On Communications, Vol. 39 (1991); p. 776.

MERCHANT, D. K. y NEMHAUSER, G. L. A model and an algorithm for the dynamic traffic assignment problem. En: trans. Sci. (1978) p. 193.

NAHMIAS, Steven. Análisis de la producción y las operaciones. Primera edición. México: CECSA, 1999. p. 425.

NOGUEIRA COBRA, Marcos Enrique. Marketing de servicios: estrategias para turismo, finanzas, salud y comunicación. Segunda edición. Bogotá; Buenos Aires: McGraw Hill, 2000. p. 255.

PORTAFOLIO. Disponible en Internet:

<http://www.portafolio.com.co/economia/economiahoy/2008-11-24/ARTICULO-WEB-NOTA_INTERIOR_PORTA-468314>

POTOMAC. New Study May Eliminate Queuing Costs. En: Airline Financial News. (Oct 21, 1996). P.88.

PRAWDA WITENBERG, Juan. Métodos y modelos de investigación de operaciones. México: Limusa, 1976. p. 2v.

R. COSS, Bu. Simulación Enfoque Práctico. Méjico: LIMUSA, (1982) p. 60 – 65.

RENDER Barry, HEIZER Jay H . Principios de administración de operaciones. Pearson Educación, (2004). p. 575

THOMPSON, G. L. y ZAWACK D. J. A problem expanding parametric programming method for solving the job shop scheduling problem. En: Operation Research. (1986); p. 342.

VAN MIEGHEM, Ann. Price and Service Discrimination in Queuing Systems: Incentive Compatibility of Gc Scheduling. En: Management Science, Vol. 46 (2000); p. 1267.

VANDA, Angelis; FELICI, Giovanni y IMPELLUSO, Paolo. Integrating simulation and optimization in health care centre management. En: European Journal of Operational Research. Vol. 150 (2003); p.114.

VAUGHN, Richard C. Introducción a la ingeniería industrial. Segunda edición. Barcelona: reverte, 1993. p. 474.

VELASCO, Enio Edgardo. Queueing theoretic approaches for the analysis and synthesis of air traffic control problems. En: Operation Research. Vol. 23 (1989); p. 456.

WANG, Junfeng. LIU, Jin y CHUNDONG, She. Segment-based adaptive hyper-Erlang model for long-tailed network traffic approximation. En: J Supercomput, Vol. 45 (2008); p. 296 – 312.

ANEXOS

Anexo A. Macro de configuración y ejecución de la simulación

Figura 51. Esquema de la información en Excel.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Landa	Miu	Factor de Utilización	Servidores Nodo 1	Capacidad Nodo 1	Servidores Nodo 2	Capacidad Nodo 2	Servidores Nodo 3	Capacidad Nodo 3	Wq Nodo 1	Desv. Wq1	Lq N
2	0.672	0.16	0.7	1	10	1	10	1	10			
3	0.672	0.16	0.7	1	10	1	10	2	10			
4	0.672	0.16	0.7	1	10	1	10	3	10			
5	0.672	0.16	0.7	1	10	1	10	5	10			
6	0.672	0.16	0.7	1	10	1	10	7	10			
7	0.672	0.16	0.7	1	10	1	10	10	10			
8	0.672	0.16	0.7	1	10	2	10	1	10			
9	0.672	0.16	0.7	1	10	2	10	2	10			
10	0.672	0.16	0.7	1	10	2	10	3	10			
11	0.672	0.16	0.7	1	10	2	10	5	10			
12	0.672	0.16	0.7	1	10	2	10	7	10			
13	0.672	0.16	0.7	1	10	2	10	10	10			
14	0.672	0.16	0.7	1	10	3	10	1	10			
15	0.672	0.16	0.7	1	10	3	10	2	10			
16	0.672	0.16	0.7	1	10	3	10	3	10			
17	0.672	0.16	0.7	1	10	3	10	5	10			
18	0.672	0.16	0.7	1	10	3	10	7	10			
19	0.672	0.16	0.7	1	10	3	10	10	10			
20	0.672	0.16	0.7	1	10	5	10	1	10			
21	0.672	0.16	0.7	1	10	5	10	2	10			
22	0.672	0.16	0.7	1	10	5	10	3	10			
23	0.672	0.16	0.7	1	10	5	10	5	10			
24	0.672	0.16	0.7	1	10	5	10	7	10			

Fuente: Elaboración del Autor

Macro 1: Modelo con un nodo.

```
Dim i, j As Integer
Sub macromodelo1()
  'Abrir Excel
  camino = Model.Path
  Excel.Workbooks.Open camino & "Prueba"
  Excel.ActiveWorkbook.Sheets(1).Activate

  For i = 3 To 86
    Models.Open camino & "nodofinito.doe"
    Models.Item(2).SaveAs "Nodos1Modelo " & i - 2
    lamda = cells(i, 1)
    C = 3
    miu = cells(i, 2)
    expr3 = "EXPO(" & 1 / miu & ")"

    'Nodo 1
    capacidad1 = cells(i, 5)
    archivo1 = "Archivo1Nodo1Modelo" & i - 2
    expr1 = "(-1/ lamda )*(ln(UNIF(0 , 1 )"
    *UNIF(0 , 1 ))+ln(UNIF(0 , 1 )*UNIF(0 , 1 )))"
    expr2 = "NQ(Servicio.Queue)<" & capacidad1
    servidores = cells(i, 4)
    indicador = 1
    Call servidor(servidores, indicador)
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.11")).Data("Interarrival Type")="Expression"
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.11")).Data("Expression") = expr1
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.31")).Data("Value") = expr2
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.48")).Data("Expression") = expr3
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.247")).Data("OSFile") = camino & archivo1

    Models.Item(2).BatchMode = True
    Models.Item(2).Go
    Models.Item(2).End
    Models.Item(2).Save
    Models.Item(2).Close

  Next
End Sub
```

Macro 2: Modelo con dos nodos.

```
Dim i, j As Integer
Sub macromodelo2 ()
  'Abrir Excel
  camino = Model.Path
  Excel.Workbooks.Open camino & "Prueba"
  Excel.ActiveWorkbook.Sheets(2).Activate

  For i = 3 To 1766
    Models.Open camino & "nodo2finito.doe"
    Models.Item(2).SaveAs "Nodos2Modelo " & i - 2
    lamda = cells(i, 1)
    C = 3
    miu = cells(i, 2)
    expr3 = "EXPO(" & 1 / miu & ")"

    'Nodo 1
    capacidad1 = cells(i, 5)
    archivo1 = "Archivo2Nodo1Modelo" & i - 2
    expr1 = "(-1/ " & lamda & " )*(ln(UNIF(0 , 1 )"
    *UNIF(0 , 1 ))+ln(UNIF(0 , 1 )*UNIF(0 , 1 )))"
    expr2 = "NQ(Servicio.Queue)<" & capacidad1
    servidores = cells(i, 4)
    indicador = 1
    Call servidor(servidores, indicador)
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.11")).Data("Interarrival Type") = "Expression"
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.11")).Data("Expression") = expr1
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.31")).Data("Value") = expr2
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.48")).Data("Expression") = expr3
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, 'object.247')).Data("OSFile") = camino & archivo1

    'Nodo 2
    capacidad2 = cells(i, 7)
    archivo2 = "Archivo2Nodo2Modelo" & i - 2
    expr5 = "NQ(Servicio2.Queue)<" & capacidad2
    servidores = cells(i, 6)
    indicador = 2
    Call servidor(servidores, indicador)
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.99")).Data("Value") = expr5
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.116")).Data("Expression") = expr3
    Models.Item(2).Modules(Models.Item(2).Modules.
    Find(smFindTag, "object.249")).Data("OSFile") = camino & archivo2
```

```
Models.Item(2).BatchMode = True
Models.Item(2).Go
Models.Item(2).End
Models.Item(2).Save
Models.Item(2).Close
Next
End Sub
```

Macro 3: Modelo con tres nodos

```
Dim i, j As Integer
Sub macromodelo3()
    'Abrir Excel
    camino = Model.Path
    Excel.Workbooks.Open camino & "Prueba"
    Excel.ActiveWorkbook.Sheets(3).Activate

    For i = 37500 To 37999
        Models.Open camino & "nodo3finito.doe"
        Models.Item(2).SaveAs "Nodos3Modelo " & i - 2
        lamda = cells(i, 1)
        C = 3
        miu = cells(i, 2)
        expr3 = "EXPO(" & 1 / miu & ")"

        'Nodo 1
        capacidad1 = cells(i, 5)
        archivo1 = "Archivo2Nodo1Modelo" & i - 2
        expr1 = "(-1/ " & lamda & ")*(ln(UNIF(0 , 1 )"
        *UNIF(0 , 1 ))+ln(UNIF(0 , 1 )*UNIF(0 , 1 )))"
        expr2 = "NQ(Servicio.Queue)<" & capacidad1
        servidores = cells(i, 4)
        indicador = 1
        Call servidor(servidores, indicador)
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, "object.11")).Data("Interarrival Type") = "Expression"
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, "object.11")).Data("Expression") = expr1
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, "object.31")).Data("Value") = expr2
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, "object.48")).Data("Expression") = expr3
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, 'object.247')).Data("OSFile") = camino & archivo1

        'Nodo 2
        capacidad2 = cells(i, 7)
        archivo2 = "Archivo2Nodo2Modelo" & i - 2
        expr5 = "NQ(Servicio2.Queue)<" & capacidad2
        servidores = cells(i, 6)
        indicador = 2
        Call servidor(servidores, indicador)
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, "object.99")).Data("Value") = expr5
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, "object.116")).Data("Expression") = expr3
        Models.Item(2).Modules(Models.Item(2).Modules.
        Find(smFindTag, "object.249")).Data("OSFile") = camino & archivo2
    
```

```

'Nodo 3
capacidad3 = cells(i, 9)
archivo3 = "Archivo3Nodo3Modelo" & i - 2
expr7 = "NQ(Servicio3.Queue)<" & capacidad3
servidores = cells(i, 8)
indicador = 3
Call servidor(servidores, indicador)
Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.165")).Data("Value") = expr7
Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.182")).Data("Expression") = expr3
Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.251")).Data("OSFile") = camino & archivo3

Models.Item(2).BatchMode = True
Models.Item(2).Go
Models.Item(2).End
Models.Item(2).Save
Models.Item(2).Close

Next
End Sub

```

Funciones de las macros

```
Function servidor(j, indicador)
servidor = "Resource " & j + 15 * (indicador - 1)
If indicador = 1 Then
  If j = 1 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(1)") = "Resource 1"
  If j >= 2 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(2)") = "Resource 2"
  If j >= 3 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(3)") = "Resource 3"
  If j >= 4 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(4)") = "Resource 4"
  If j >= 5 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(5)") = "Resource 5"
  If j >= 6 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(6)") = "Resource 6"
  If j >= 7 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(7)") = "Resource 7"
  If j >= 8 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(8)") = "Resource 8"
  If j >= 9 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(9)") = "Resource 9"
  If j >= 10 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(10)") = "Resource 10"
  If j >= 11 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(11)") = "Resource 11"
  If j >= 12 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(12)") = "Resource 12"
  If j >= 13 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(13)") = "Resource 13"
  If j >= 14 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(14)") = "Resource 14"
  If j >= 15 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.50").Data("Resource Name(15)") = "Resource 15"
End If
If indicador = 2 Then
  If j = 1 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(1)") = "Resource 16"
  If j >= 2 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(2)") = "Resource 17"
  If j >= 3 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(3)") = "Resource 18"
  If j >= 4 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(4)") = "Resource 19"
  If j >= 5 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(5)") = "Resource 20"
  If j >= 6 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(6)") = "Resource 21"
  If j >= 7 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(7)") = "Resource 22"
  If j >= 8 Then Models.Item(2).Modules(models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(8)") = "Resource 23"
```



```

If j >= 9 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(9)") = "Resource 24"
If j >= 10 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(10)") = "Resource 25"
If j >= 11 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(11)") = "Resource 26"
If j >= 12 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(12)") = "Resource 27"
If j >= 13 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(13)") = "Resource 28"
If j >= 14 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(14)") = "Resource 29"
If j >= 15 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.118").Data("Resource Name(15)") = "Resource 30"
End If

```

```

If indicador = 3 Then
If j >= 1 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(1)") = "Resource 31"
If j >= 2 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(2)") = "Resource 32"
If j >= 3 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(3)") = "Resource 33"
If j >= 4 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(4)") = "Resource 34"
If j >= 5 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(5)") = "Resource 35"
If j >= 6 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(6)") = "Resource 36"
If j >= 7 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(7)") = "Resource 37"
If j >= 8 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(8)") = "Resource 38"
If j >= 9 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(9)") = "Resource 39"
If j >= 10 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(10)") = "Resource 40"
If j >= 11 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(11)") = "Resource 41"
If j >= 12 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(12)") = "Resource 42"
If j >= 13 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(13)") = "Resource 43"
If j >= 13 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(13)") = "Resource 43"
If j >= 14 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(14)") = "Resource 44"
If j >= 15 Then Models.Item(2).Modules(Models.Item(2).Modules.
Find(smFindTag, "object.184").Data("Resource Name(15)") = "Resource 45"
End If

```

Anexo B. Macro para la administración y unificación de la información desde Excel.

```
Sub Principal()  
  
Workbooks.Open (Range("ruta_destino") & "\" _  
& "Base con la simulacion.xlsx")  
  
For nodo = 1 To 3  
  
    Select Case nodo  
  
        Case 1  
            Call nodo1  
        Case 2  
            Call nodo2  
        Case 3  
            Call nodo3  
  
    End Select  
  
Next  
  
End Sub
```

Funciones:

```
Function Access_Excel(modelo As String)  
'Dimensionar objetos  
  
Dim ruta_destino As String  
Dim Hoja As String  
Dim ruta_origen As String  
Dim reporte As String  
Dim objDB As Database  
Dim faltantes(1 To 100) As String  
  
'datos de busqueda  
    ruta_destino = Range("ruta_destino") & _  
    "\Excel_Base_Transitorio.xls"  
    Hoja = "hoja1"  
    ruta_origen = Range("ruta_origen") & "\" & modelo  
    reporte = "Statistic"  
  
Set objDB = OpenDatabase(ruta_origen)
```

```

'Si el fichero excel existe, puedes borrarlo aqui
If Dir(ruta_destino) <> "" Then Kill ruta_destino

objDB.Execute _
"SELECT * INTO [Excel 8.0;DATABASE=" & ruta_destino & "]"
[" & Hoja & "] FROM " & "[" & reporte & "]"
objDB.Close
Set objDB = Nothing

End Function

Function sacar_dato(definicion As Integer)

    Dim datos(1 To 1, 1 To 2)

    Sheets("hoja1").Select
    ActiveSheet.Range("$A$1:$J$30489").AutoFilter Field:=3, _
    Criterial:=definicion
    Range("f2").Select
    Range(Selection, Selection.End(xlDown)).Select
    Selection.Copy
    Range("M38000").Select
    ActiveSheet.Paste
    Application.CutCopyMode = False
    Range("M38000").Select
    vector = Range("m38000", Selection.End(xlDown))
    promedio = Application.WorksheetFunction.Average(vector)
    desvesta = Application.WorksheetFunction.StDev(vector)
    datos(1, 1) = promedio
    datos(1, 2) = desvesta
    sacar_dato = datos
    Columns("M:M").Select
    Selection.Delete Shift:=xlToLeft

End Function

```

Nodos:

```
Sub nodo3 ()

j = 2
Windows("Macro Base.xlsm").Activate
ruta = Range("ruta_origen") & "\"
i = 1

h = 8
Do Until Range("h" & h) = ""

' nodos 3
  |   For i = Range("h" & h) To Range("i" & h)
  |   modelo = "Nodos3Modelo " & i
  |   Windows("macro base.xlsm").Activate
' el archivo existe?
d:
  If Dir(ruta & modelo & ".mdb") = "" Then
    Windows("macro base.xlsm").Activate
    Sheets(1).Select
    Range("a" & j).FormulaR1C1 = modelo
    j = j + 1
    GoTo g:
  End If
If i = 2 Then
d:

  Application.DisplayAlerts = False
  Application.ScreenUpdating = False

  Windows("Excel_Base_Transitorio.xls").Close

  Application.DisplayAlerts = True
  Application.ScreenUpdating = True

  Windows("macro base.xlsm").Activate
  Sheets(1).Select
  Range("a" & j).FormulaR1C1 = modelo
  j = j + 1
  GoTo g:
End If

' Abrir el modelo y pasarlo a excel

  Access_Excel (modelo)
```

```

'abrir archivo temporal excel

    Workbooks.Open (Range("ruta_destino") & "\" & _
|"Excel_Base_Transitorio.xls")
    On Error GoTo d:
    w1 = sacar_dato(9)
    w2 = sacar_dato(7)
    w3 = sacar_dato(8)

    l1 = sacar_dato(148)
    l2 = sacar_dato(146)
    l3 = sacar_dato(147)

    Application.DisplayAlerts = False
    Application.ScreenUpdating = False

    Windows("Excel_Base_Transitorio.xls").Close

    Application.DisplayAlerts = True
    Application.ScreenUpdating = True

'escribir los resultados en la base

    Windows("Base con la simulacion.xlsx").Activate
    Sheets(3).Activate

    Range("k" & h + 2).FormulaR1C1 = w1(1, 1)
    Range("l" & h + 2).FormulaR1C1 = w1(1, 2)

    Range("m" & h + 2).FormulaR1C1 = l1(1, 1)
    Range("n" & h + 2).FormulaR1C1 = l1(1, 2)

    Range("o" & h + 2).FormulaR1C1 = w2(1, 1)
    Range("p" & h + 2).FormulaR1C1 = w2(1, 2)

    Range("q" & h + 2).FormulaR1C1 = l2(1, 1)
    Range("r" & h + 2).FormulaR1C1 = l2(1, 2)

    Range("s" & h + 2).FormulaR1C1 = w3(1, 1)
    Range("t" & h + 2).FormulaR1C1 = w3(1, 2)

    Range("u" & h + 2).FormulaR1C1 = l3(1, 1)
    Range("v" & h + 2).FormulaR1C1 = l3(1, 2)

g:
Next i
Loop

End Sub

```

```

Sub nodo2 ()

Application.DisplayAlerts = False
Application.ScreenUpdating = False

j = 2
Windows("Macro Base.xlsm").Activate
ruta = Range("ruta_origen") & "\"
i = 1

' nodos 2

    For i = 1582 To 17064
        modelo = "Nodos2Modelo " & i
        Windows("macro base.xlsm").Activate
'el archivo existe?

        If Dir(ruta & modelo & ".mdb") = "" Then
d:
            Windows("macro base.xlsm").Activate
            Sheets(1).Select
            Range("a" & j).FormulaR1C1 = modelo
            j = j + 1
            GoTo g:
        End If

' Abrir el modelo y pasarlo a excel

        Access_Excel (modelo)

' abrir archivo temporal excel

        Workbooks.Open (Range("ruta_destino") _
            & "\" & "Excel_Base_Transitorio.xls")

        On Error GoTo d:|
        w1 = sacar_dato(8)
        w2 = sacar_dato(7)

        l1 = sacar_dato(101)
        l2 = sacar_dato(100)

        Application.DisplayAlerts = False
        Application.ScreenUpdating = False

        Windows("Excel_Base_Transitorio.xls").Close

        Application.DisplayAlerts = True
        Application.ScreenUpdating = True

```

```

'escribir los resultados en la base

    Windows("Base con la simulacion.xlsx").Activate
    Sheets(2).Activate

    Range("i" & i + 2).FormulaR1C1 = w1(1, 1)
    Range("j" & i + 2).FormulaR1C1 = w1(1, 2)

    Range("k" & i + 2).FormulaR1C1 = l1(1, 1)
    Range("l" & i + 2).FormulaR1C1 = l1(1, 2)

    Range("m" & i + 2).FormulaR1C1 = w2(1, 1)
    Range("n" & i + 2).FormulaR1C1 = w2(1, 2)

    Range("o" & i + 2).FormulaR1C1 = l2(1, 1)
    Range("p" & i + 2).FormulaR1C1 = l2(1, 2)

g:
Next i

Application.DisplayAlerts = True
Application.ScreenUpdating = True

End Sub

Sub nodos1()

j = 2
Windows("Macro Base.xlsm").Activate
ruta = Range("ruta_origen") & "\"
i = 1

'nodos 1

    For i = 1 To 84
        modelo = "Nodos1Modelo " & i
        Windows("macro base.xlsm").Activate
'es el archivo existe?

        If Dir(ruta & modelo & ".mdb") = "" Then
            Windows("macro base.xlsm").Activate
            Sheets(1).Select
            Range("a" & j).FormulaR1C1 = modelo
            j = j + 1
            GoTo g:
        End If

'Abrir el modelo y pasarlo a excel

        Access_Excel (modelo)

```

```

'abrir archivo temporal excel

    Workbooks.Open (Range("ruta_destino") & "\" _
        & "Excel_Base_Transitorio.xls")

    w1 = sacar_dato(7)

    l1 = sacar_dato(54)

    Application.DisplayAlerts = False
    Application.ScreenUpdating = False

    Windows("Excel_Base_Transitorio.xls").Close

    Application.DisplayAlerts = True
    Application.ScreenUpdating = True

'escribir los resultados en la base

    Windows("Base con la simulacion.xlsx").Activate
    Sheets(1).Activate

    Range("g" & i + 2).FormulaR1C1 = w1(1, 1)
    Range("h" & i + 2).FormulaR1C1 = w1(1, 2)

    Range("i" & i + 2).FormulaR1C1 = l1(1, 1)
    Range("j" & i + 2).FormulaR1C1 = l1(1, 2)

g:
Next i
End Sub

```


Anexo C. Resultado En SPSS

Figura 52. Base de datos con resultados dentro del programa SPSS

Resultados1.sav [Conjunto_de_datos1] - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

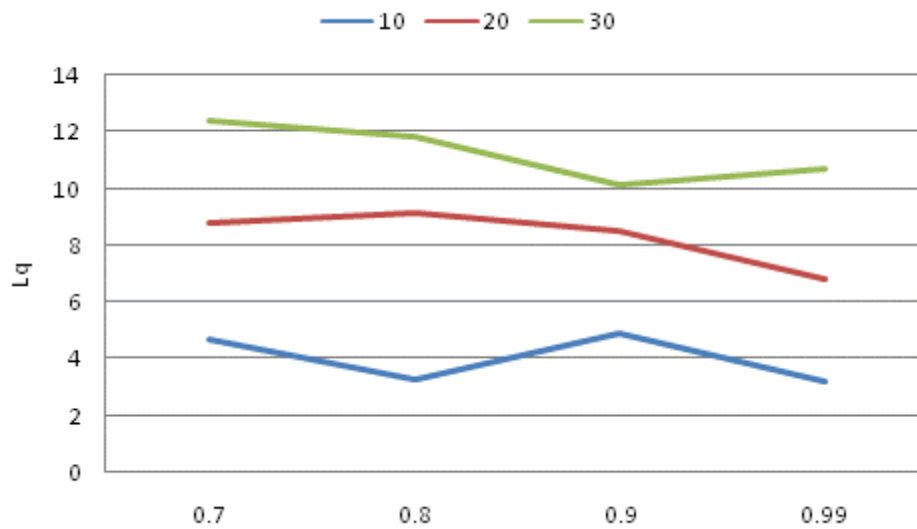
36 : FactordeUtilización 0.7 Visible: 27 d

	Landa	Miu	FactordeUtiliza	ServidoresNodo	CapacidadNodo	ServidoresNodo	CapacidadNodo	ServidoresNodo	CapacidadNodo	WqNodo1	Desv.Wq1
1	.672	.16	.7	1	10	1	10	1	10	42785317937	207191258683
2	.672	.16	.7	1	10	1	10	2	10	41641084503	194399452472
3	.672	.16	.7	1	10	1	10	3	10	43248573772	201935694428
4	.672	.16	.7	1	10	1	10	5	10	42503266502	201613015954
5	.672	.16	.7	1	10	1	10	7	10	42546710977	201190018991
6	.672	.16	.7	1	10	1	10	10	10	42546710977	201190018991
7	.672	.16	.7	1	10	2	10	1	10	57348421435	180806249890
8	.672	.16	.7	1	10	2	10	2	10	56815247975	173501322975
9	.672	.16	.7	1	10	2	10	3	10	57459213618	176857583680
10	.672	.16	.7	1	10	2	10	5	10	56790135016	170380795669
11	.672	.16	.7	1	10	2	10	7	10	56675478814	170907316123
12	.672	.16	.7	1	10	2	10	10	10	56675478814	170907316123
13	.672	.16	.7	1	10	3	10	1	10	49581915437	210083315102
14	.672	.16	.7	1	10	3	10	2	10	49849214526	205784526434
15	.672	.16	.7	1	10	3	10	3	10	50556224818	210075678750
16	.672	.16	.7	1	10	3	10	5	10	50787552822	210817019515
17	.672	.16	.7	1	10	3	10	7	10	50633302699	209515913225
18	.672	.16	.7	1	10	3	10	10	10	50675290381	209621045142
19	.672	.16	.7	1	10	5	10	1	10	50977416370	212875706994
20	.672	.16	.7	1	10	5	10	2	10	47850382468	202476126462
21	.672	.16	.7	1	10	5	10	3	10	48299347200	197593890517
22	.672	.16	.7	1	10	5	10	5	10	47458215448	191426008911
23	.672	.16	.7	1	10	5	10	7	10	47509638908	190219340885
24	.672	.16	.7	1	10	5	10	10	10	47509638908	190219340885

Fuente: Elaboración del Autor

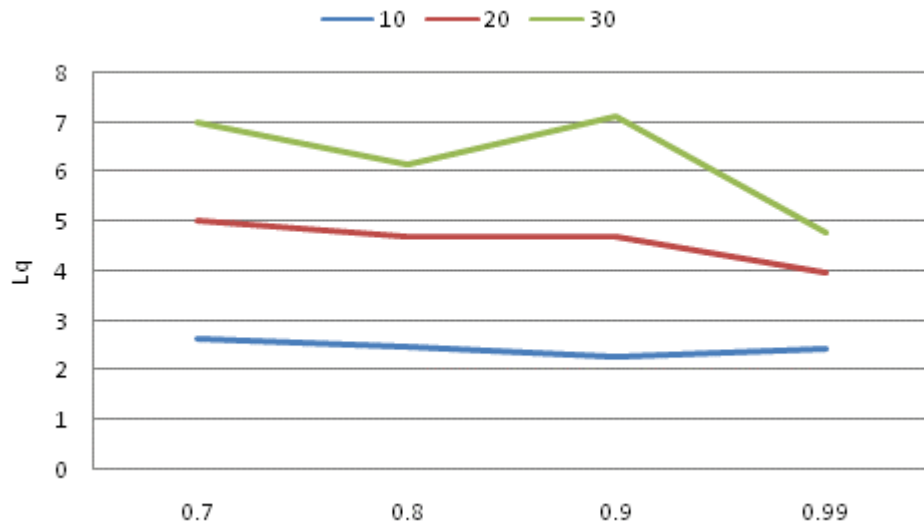
Anexo D. Resultado Lq en Redes de colas

Figura 53. Resultados de Lq según la capacidad en el segundo nodo bajo la distribución Híper-Erlang.



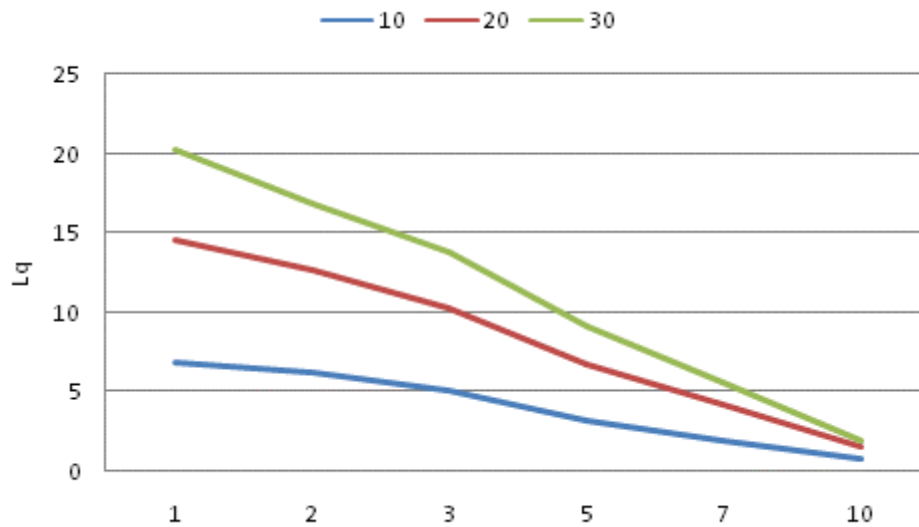
Fuente: Elaboración del Autor

Figura 54. Resultados de Lq según la capacidad en el tercer nodo bajo la distribución Híper-Erlang.



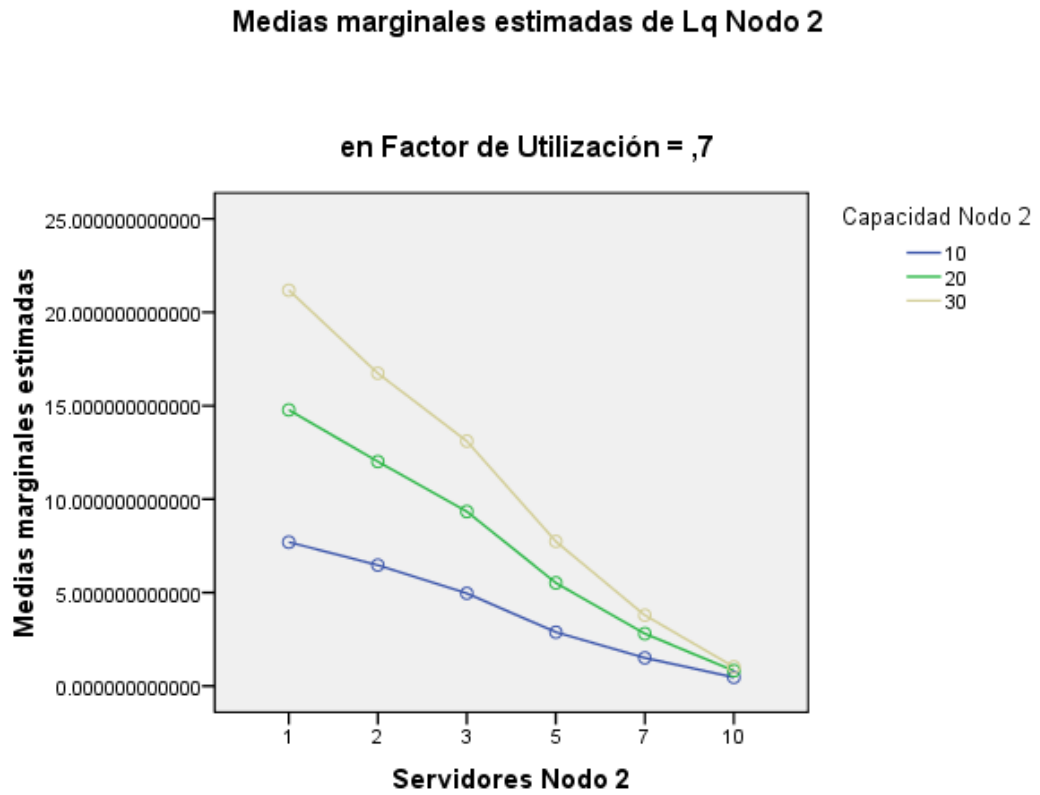
Fuente: Elaboración del Autor

Figura 55. Resultados de Lq según el número de servidores y la capacidad en el segundo nodo bajo la distribución Híper-Erlang



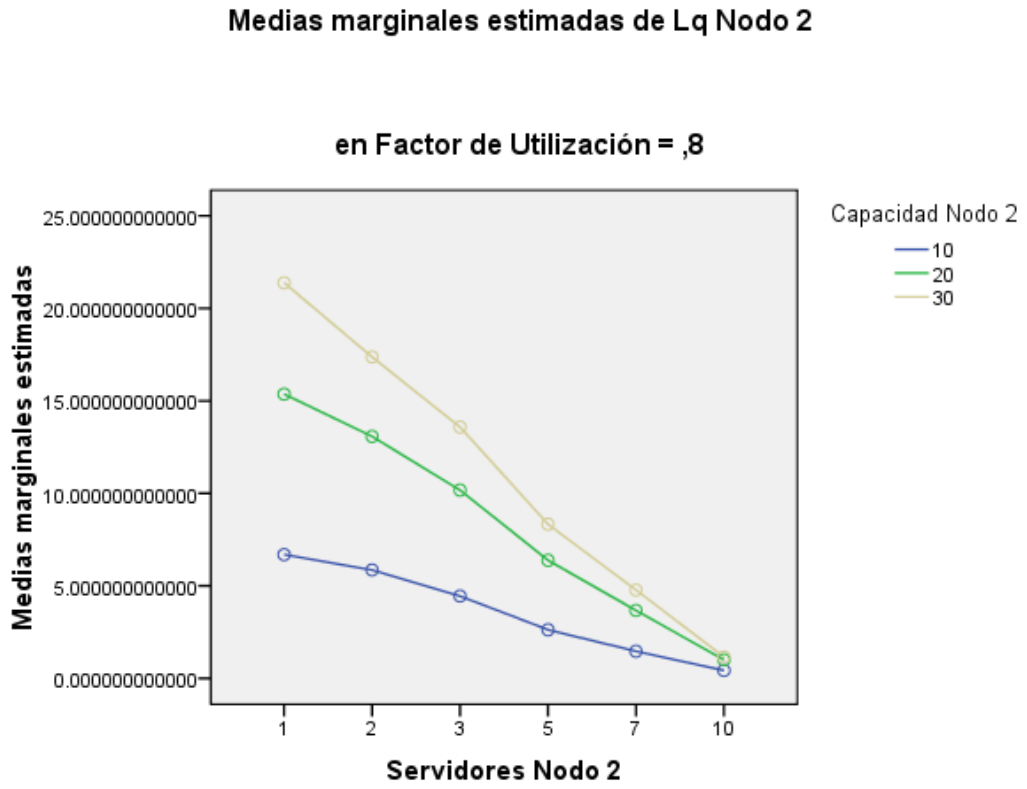
Fuente: Elaboración del Autor

Figura 56. Medias marginales estimadas del Lq Nodo 2, Rho 0,7



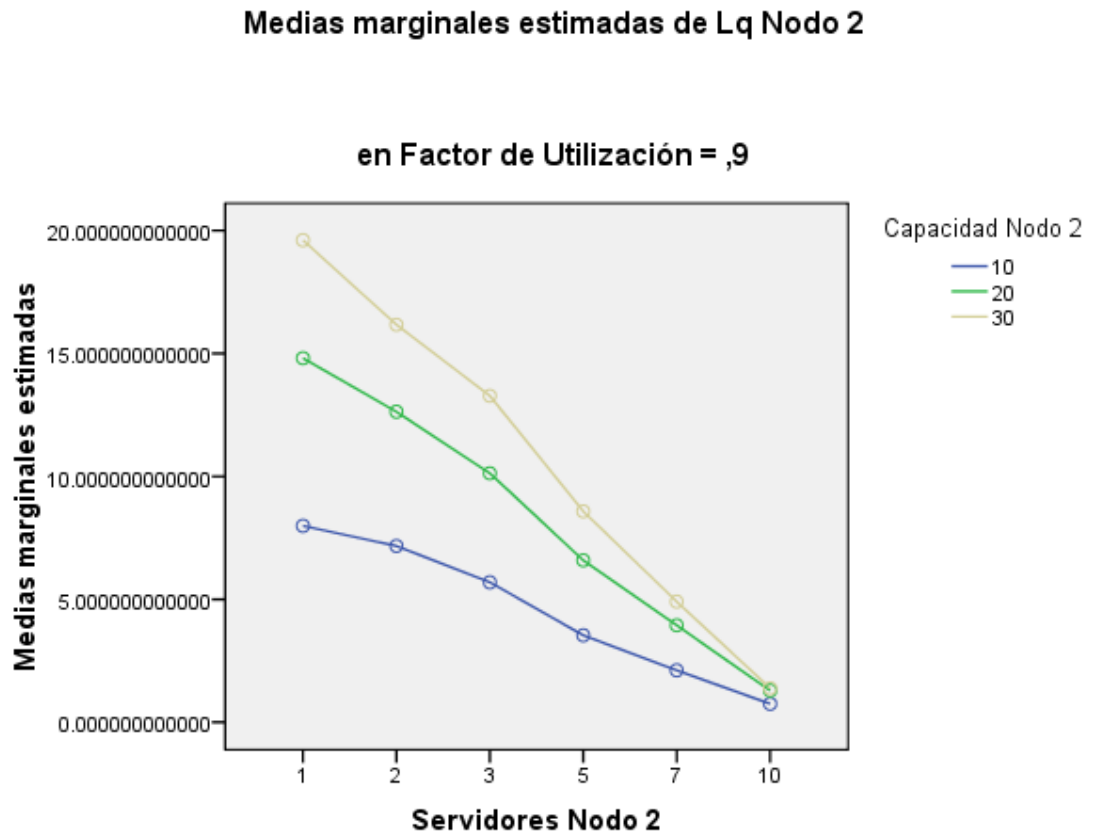
Fuente: Elaboración del Autor

Figura 57. Medias marginales estimadas del Lq Nodo 2, Rho 0,8



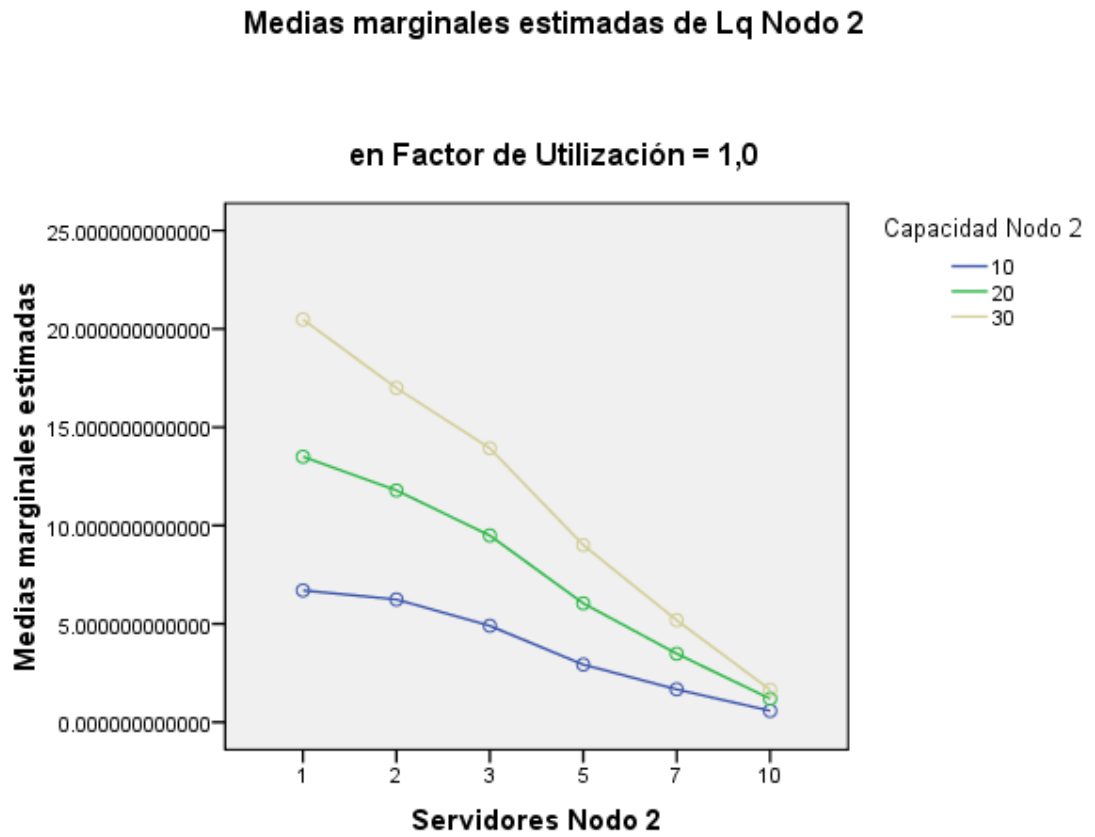
Fuente: Elaboración del Autor

Figura 58. Medias marginales estimadas del Lq Nodo 2, Rho 0,9



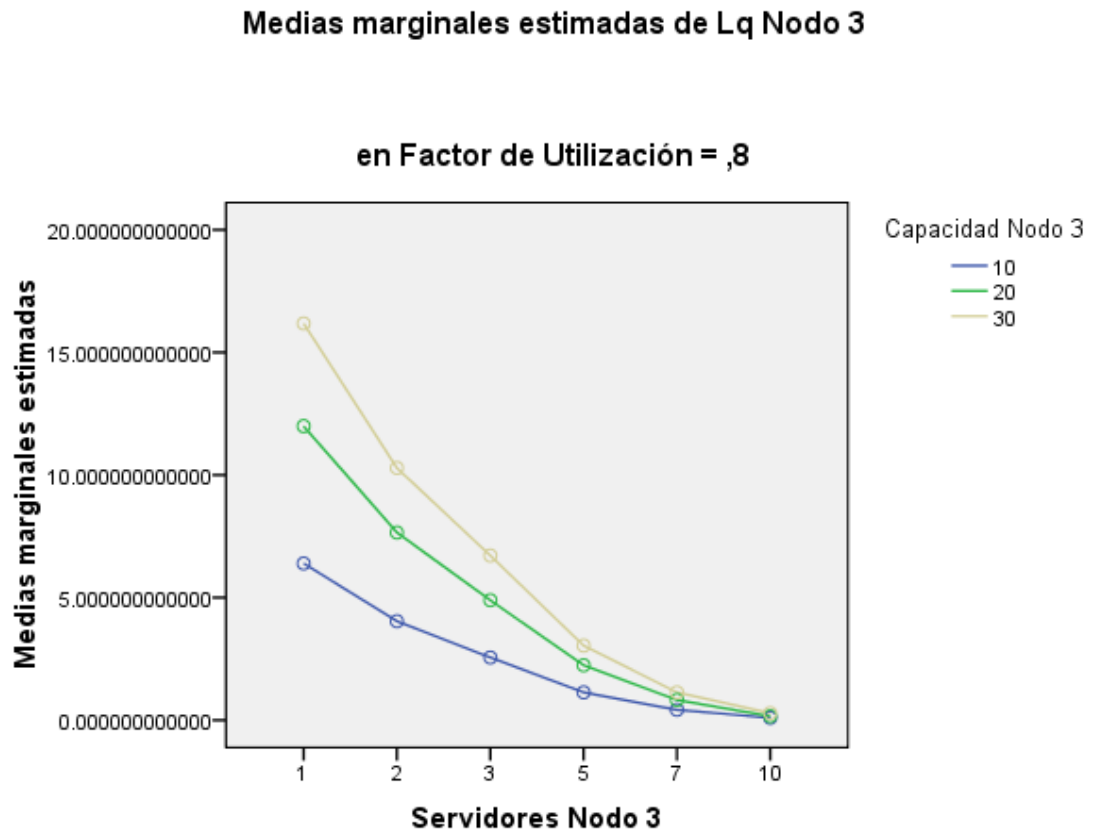
Fuente: Elaboración del Autor

Figura 59. Medias marginales estimadas del Lq Nodo 2. Rho 0,99



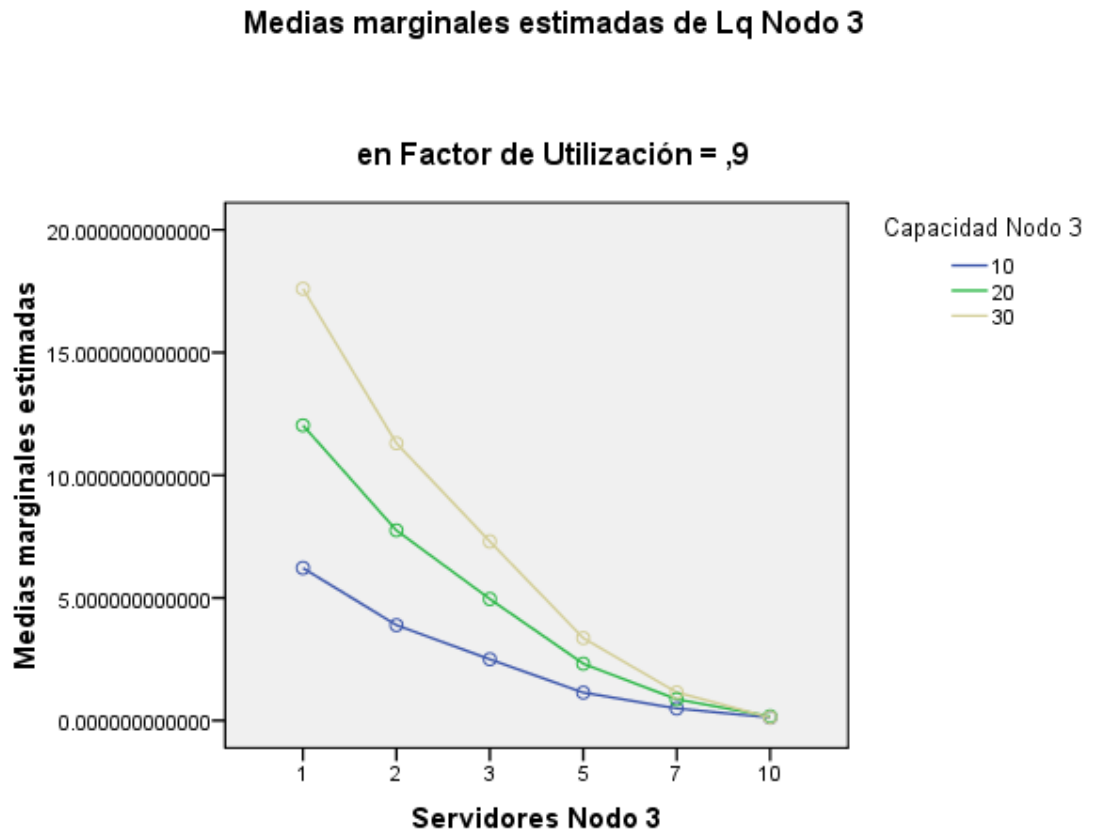
Fuente: Elaboración del Autor

Figura 60. Medias marginales estimadas del Lq Nodo 3. Rho 0,8.



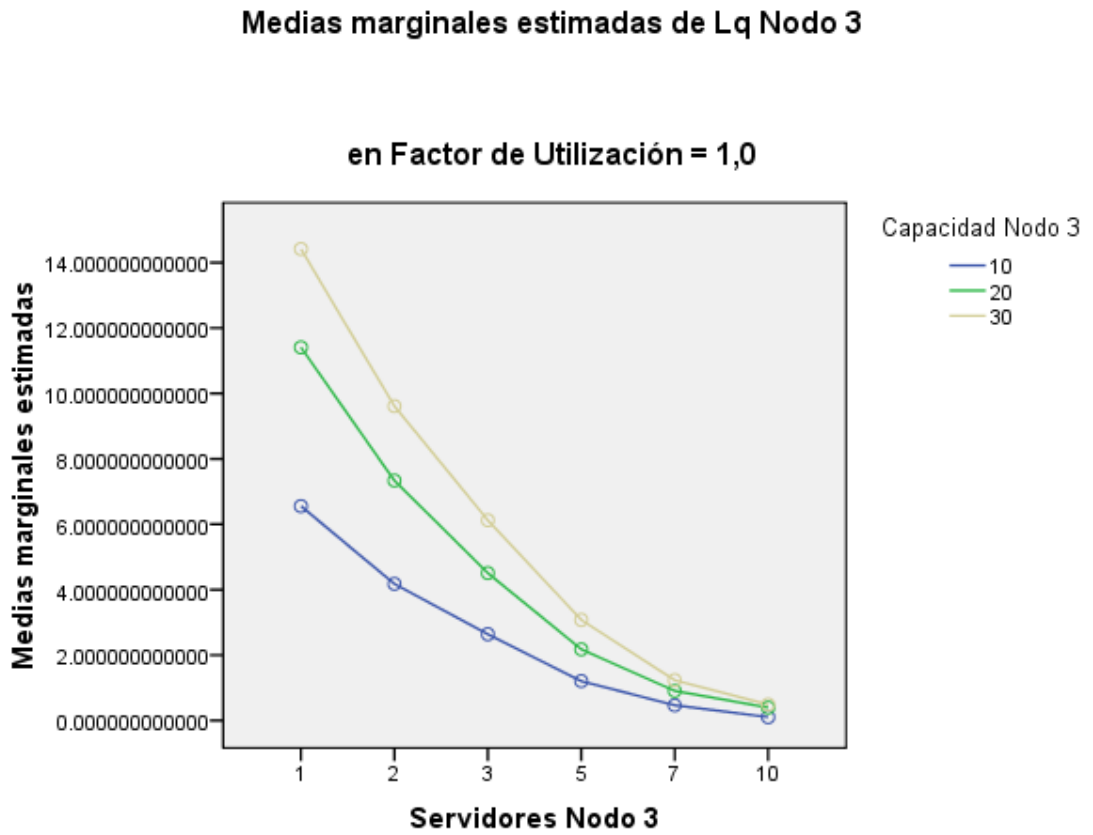
Fuente: Elaboración del Autor

Figura 61. Medias marginales estimadas del Lq Nodo 3. Rho 0,9.



Fuente: Elaboración del Autor

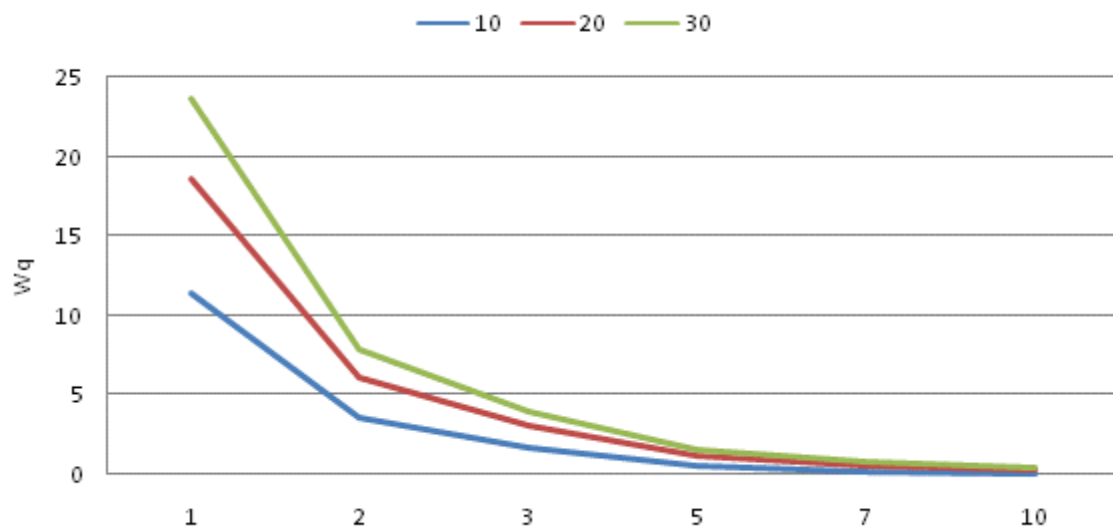
Figura 62. Medias marginales estimadas del Lq Nodo 3. Rho 0,99.



Fuente: Elaboración del Autor

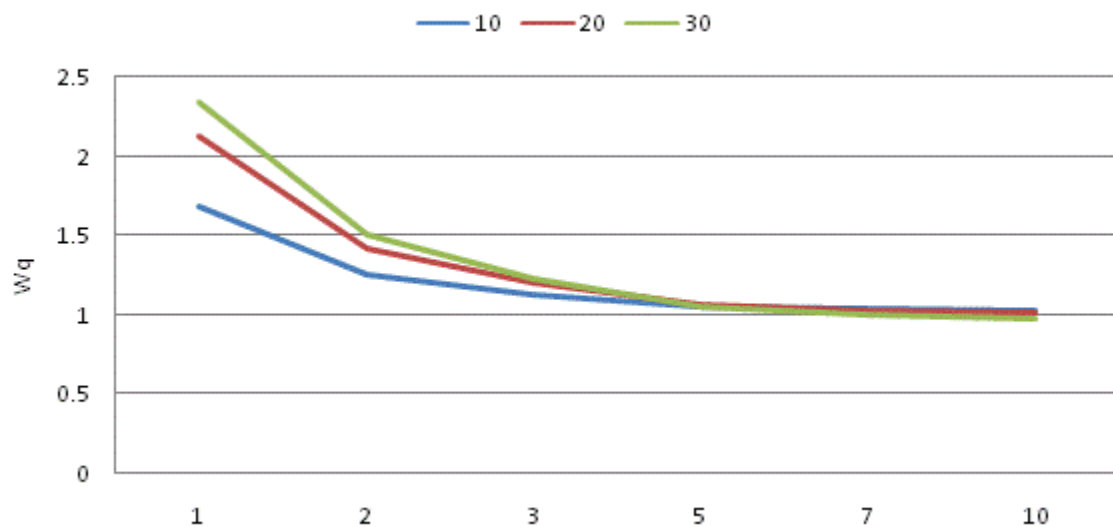
Anexo E. Resultados Wq redes de colas.

Figura 63. Resultados de Wq según el número de servidores y la capacidad en el segundo nodo bajo la distribución Exponencial



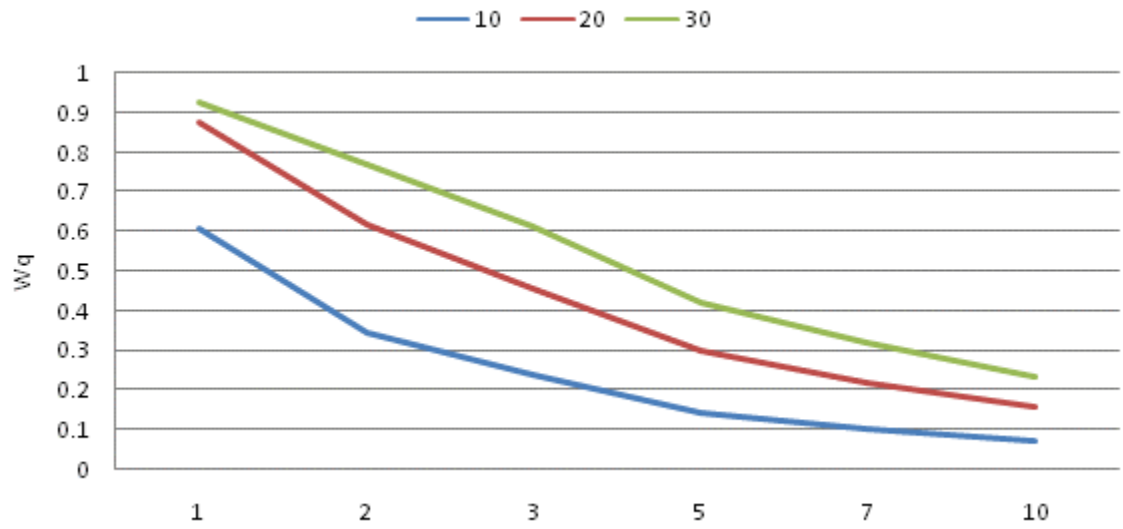
Fuente: Elaboración del Autor.

Figura 64. Resultados de Wq según el número de servidores y la capacidad en el tercer nodo bajo la distribución Híper- Erlang



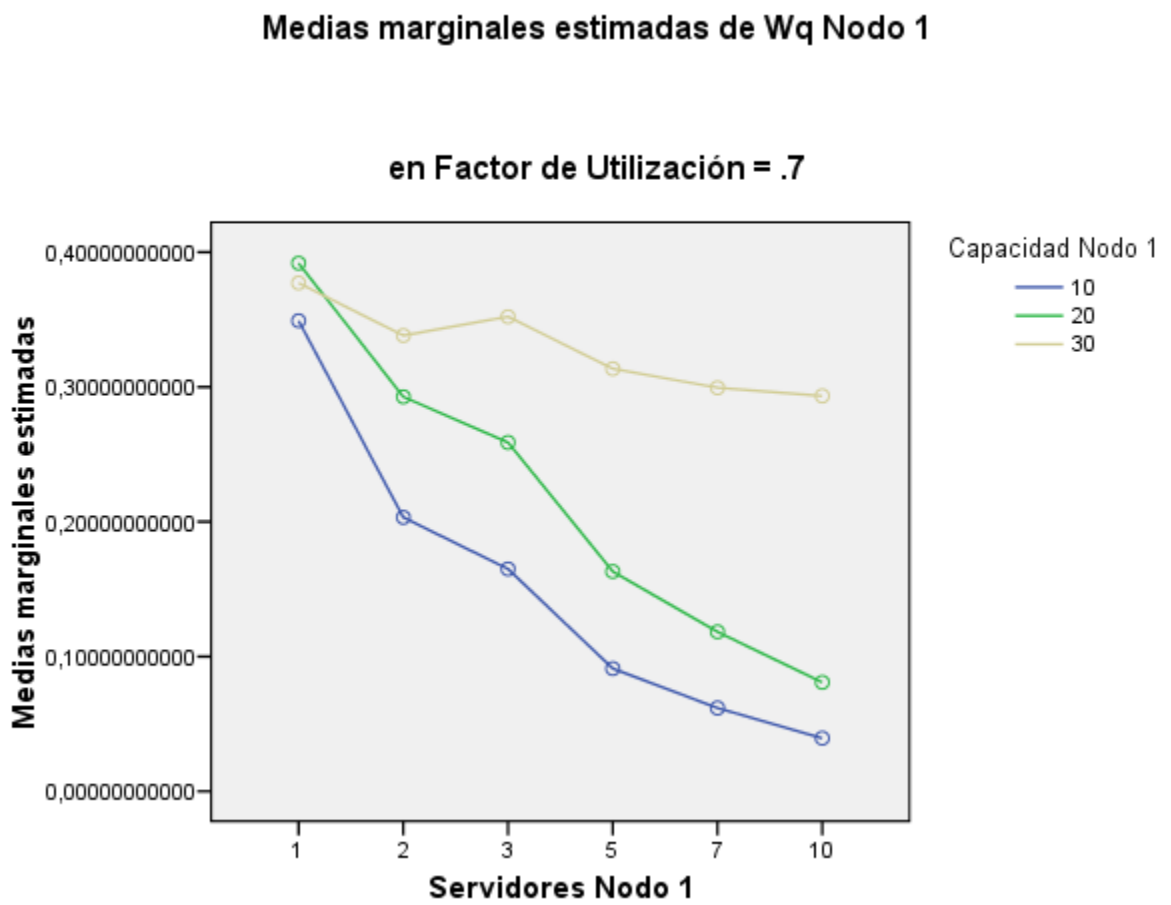
Fuente: Elaboración del Autor.

Figura 65. Resultados de L_q según el número de servidores y la capacidad en el tercer nodo bajo la distribución Híper-Erlang



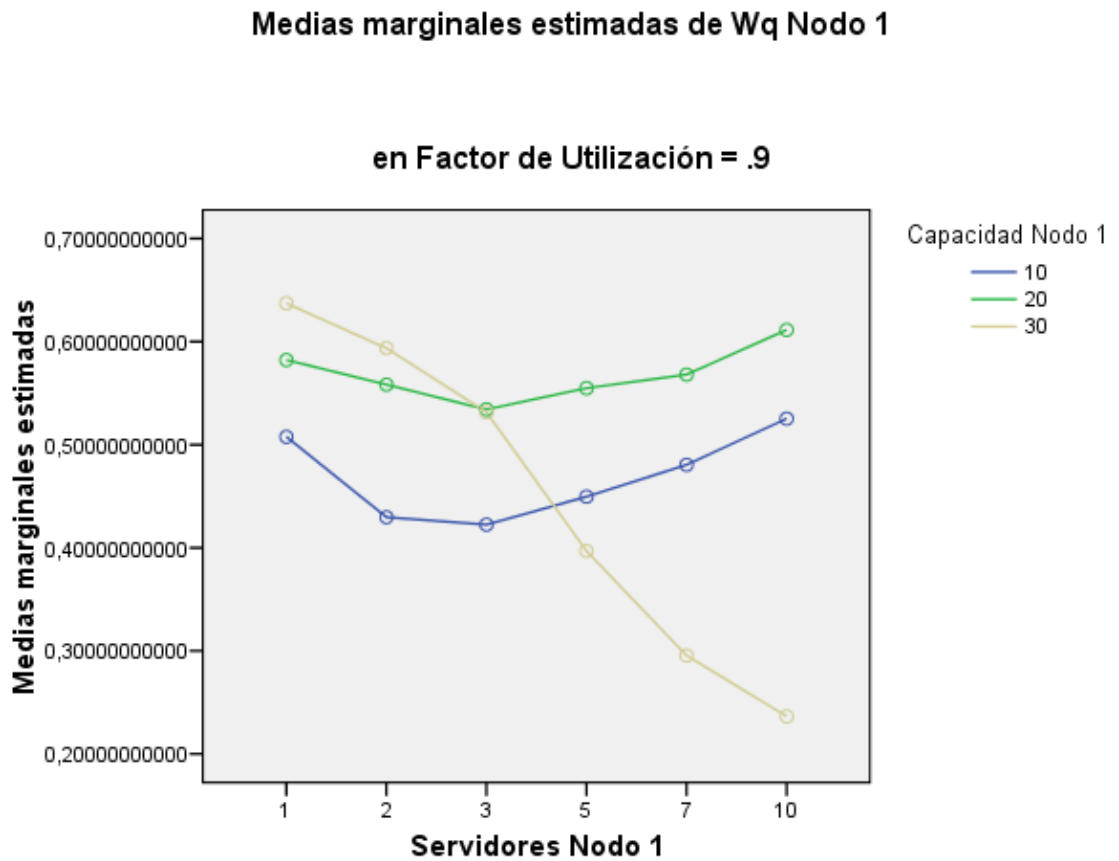
Fuente: Elaboración del Autor.

Figura 66. Medias marginales estimadas del Wq Nodo 1, Rho 0,7



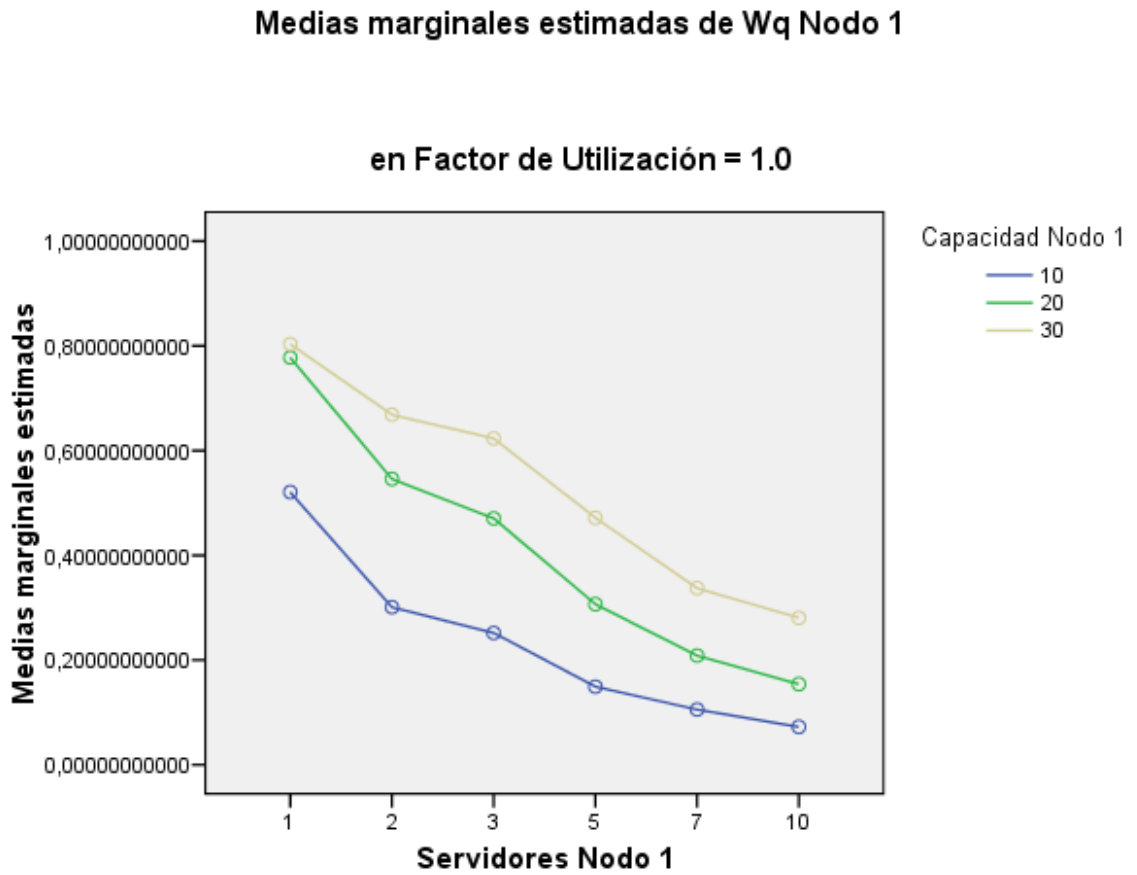
Fuente: Elaboración del Autor.

Figura 67. Medias marginales estimadas del Wq Nodo 1, Rho 0,9.



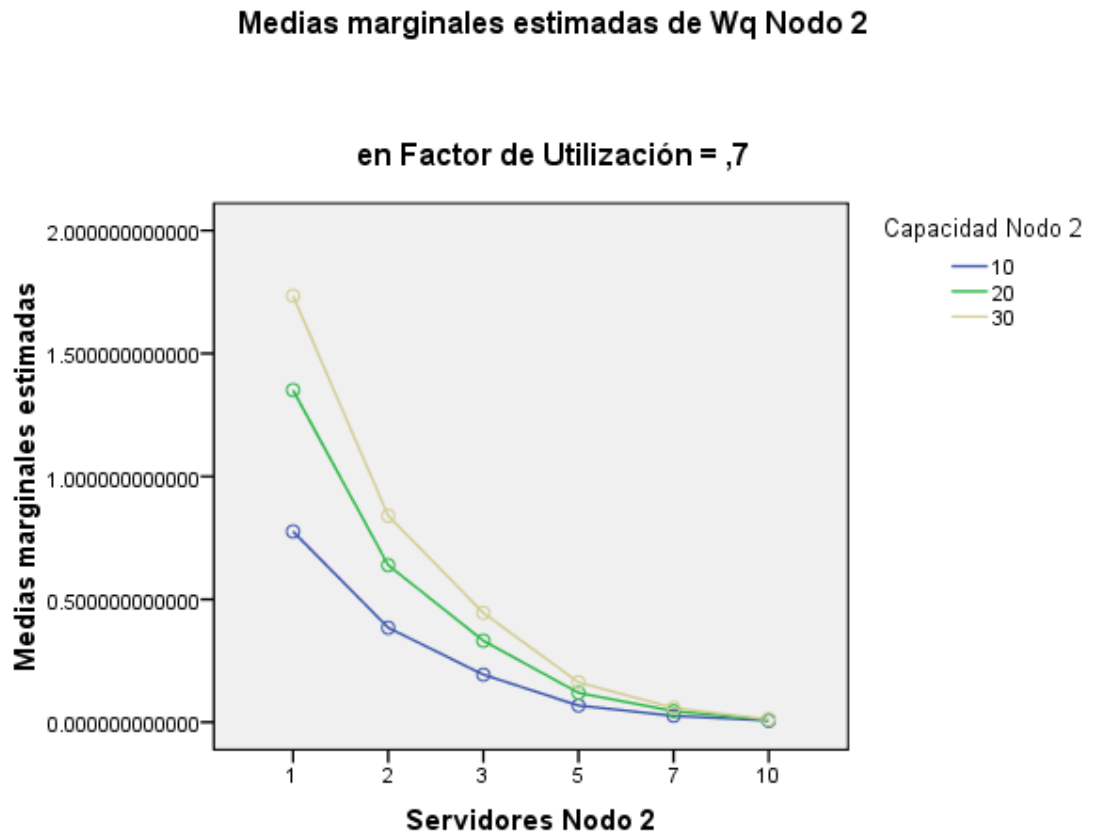
Fuente: Elaboración del Autor.

Figura 68. Medias marginales estimadas del Wq Nodo 1. Rho 0,99



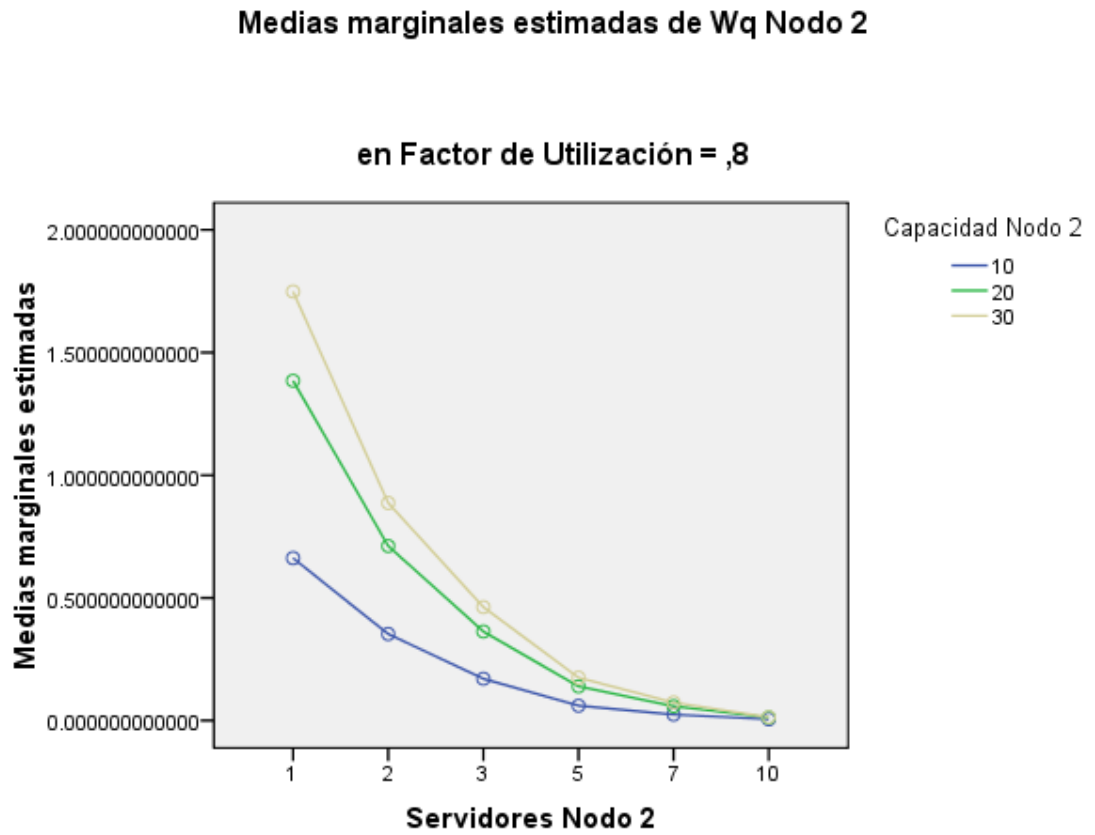
Fuente: Elaboración del Autor.

Figura 69. Medias marginales estimadas del wq Nodo 2. Rho 0,7.



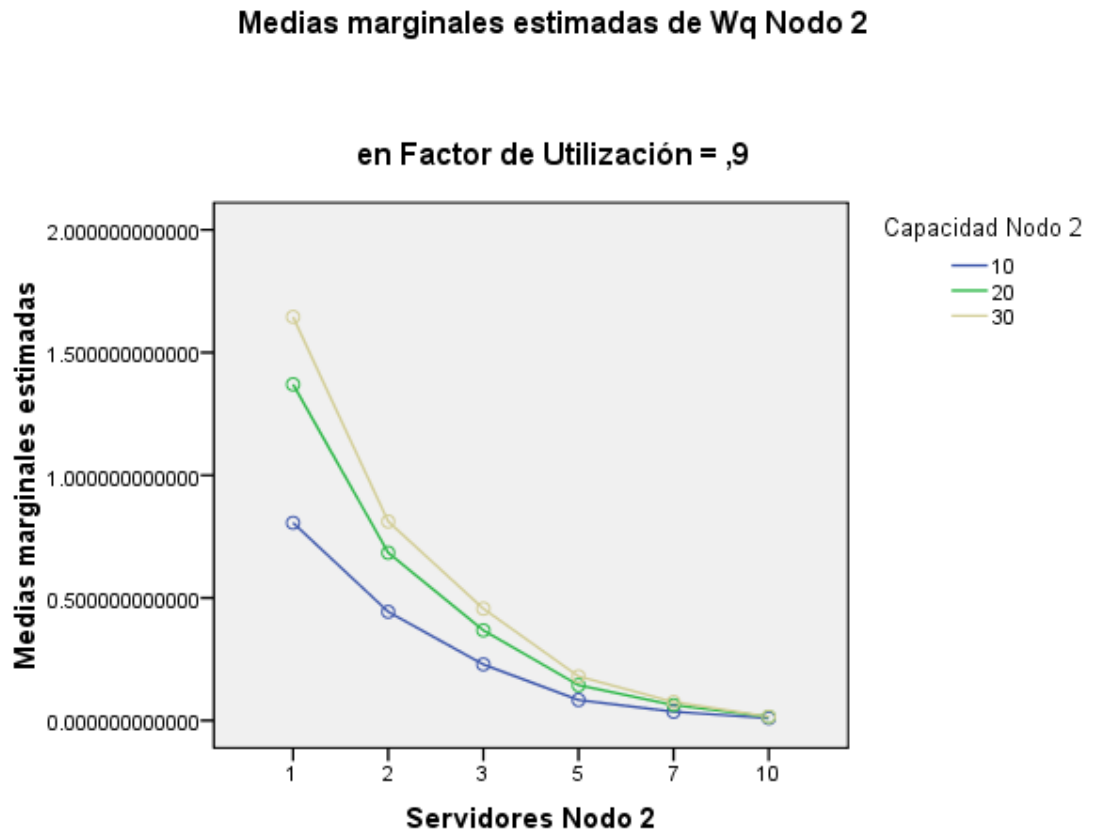
Fuente: Elaboración del Autor.

Figura 70. Medias marginales estimadas del Wq Nodo 2. Rho 0,8.



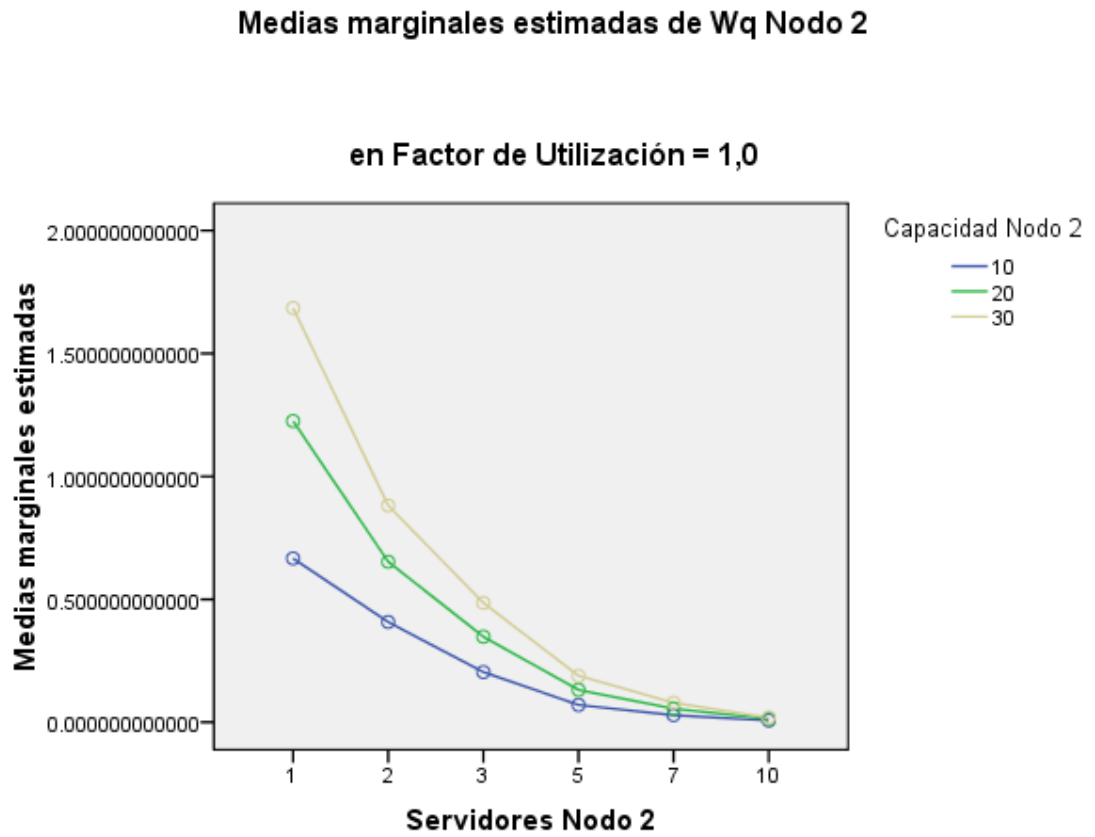
Fuente: Elaboración del Autor.

Figura 71. Medias marginales estimadas del Wq Nodo 2. Rho 0,9.



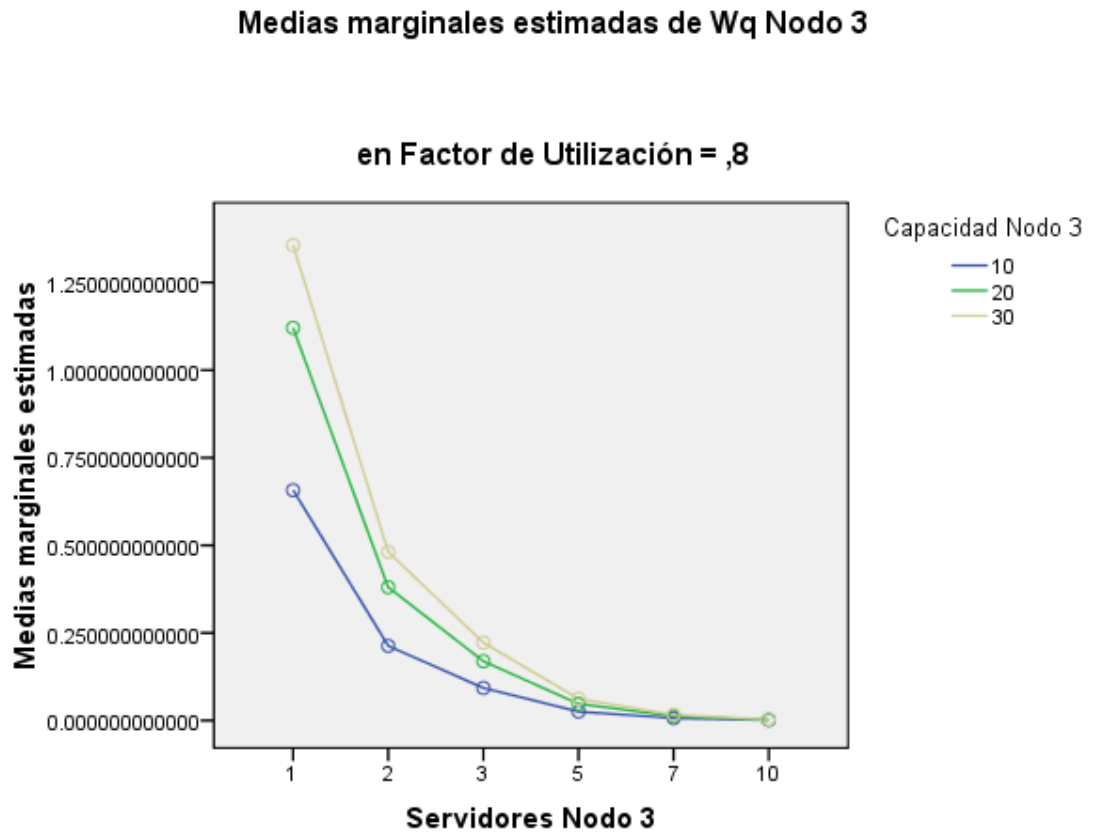
Fuente: Elaboración del Autor.

Figura 72. Medias marginales estimadas del Wq Nodo 2. Rho 0,99.



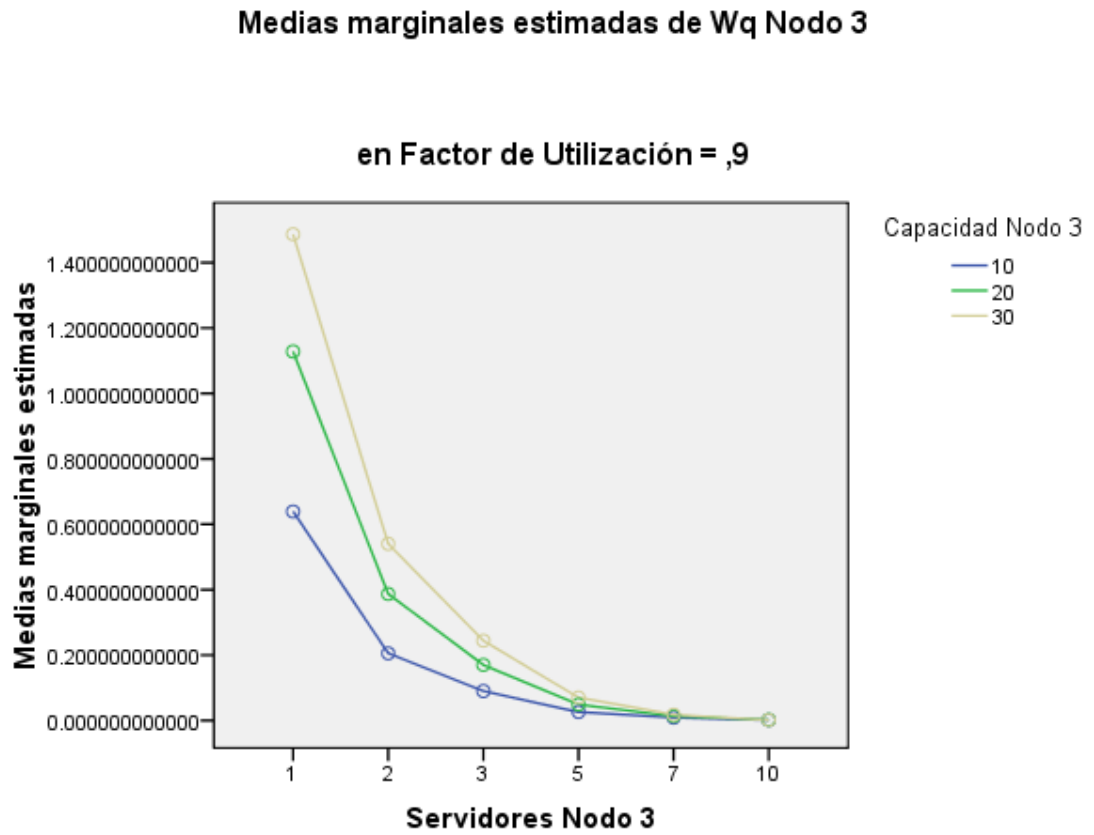
Fuente: Elaboración del Autor.

Figura 73. Medias marginales estimadas del Wq Nodo 3. Rho 0,8.



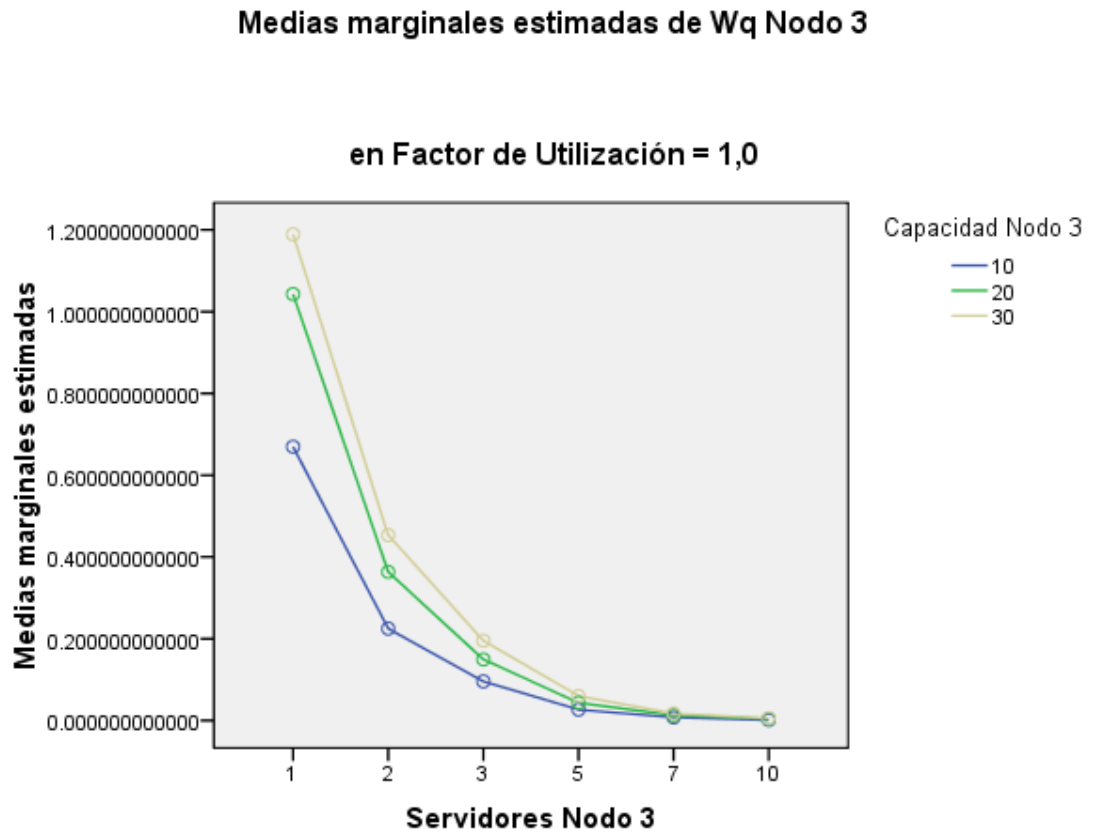
Fuente: Elaboración del Autor.

Figura 74. Medias marginales estimadas del Wq Nodo 3. Rho 0,9.



Fuente: Elaboración del Autor.

Figura 75. Medias marginales estimadas del Wq Nodo 3. Rho 0,99.



Fuente: Elaboración del Autor.

Anexo F. Medias de dispersión de la frontera Pareto

Figura 76. Cuadro de dispersiones por nodo según el algoritmo

Nodo 1	Nodo 2	Nodo 3
Di	Di	Di
0.22	0.22	0.21
0.11	0.06	0.09
0.06	0.04	0.05
0.16	0.09	0.32
0.05	0.06	0.05
0.10	0.05	0.08
0.05	0.04	0.07
0.04		0.05
		0.04

Promedio:		
0.10	0.08	0.11

Sumatoria de las diferencias:		
0.12	0.14	0.10
0.01	0.02	0.02
0.04	0.04	0.05
0.06	0.01	0.21
0.04	0.02	0.05
0.00	0.03	0.03
0.05	0.03	0.03
0.06	0.08	0.06
0.10	0.08	0.07

D =	0.16
-----	------

* Diferencias luego de la aplicación de la fórmula de elasticidad.

Anexo G. Herramienta de optimización

```
Sub optimizacion(lamda, miu, M, dtr, cpt, ctl, cutot, cs, ce, co, pesol, _  
ccc, stuno, stdos, sttres, capuno, capdos, captres, mp)  
  
camino = InputBox("Ruta de acceso, (ej.C:\Users\)")  
  
Excel.Workbooks.Open camino & "Macro Base opt.xlsm"  
  
For n = 1 To M  
  
sm = sam(n, stuno, stdos, sttres)  
ck = Kam(n, capuno, capdos, captres) / 10  
Km = Kam(n, capuno, capdos, captres)  
ReDim s(1 To sm * ck)  
ReDim Z(1 To sm * ck)  
ReDim y(1 To sm * ck)  
ReDim kp(1 To sm * ck)  
ReDim Z1(1 To sm * ck)  
ReDim y1(1 To sm * ck)  
ReDim kp1(1 To sm * ck)  
ReDim vz(1 To sm * ck)  
ReDim vy(1 To sm * ck)  
ReDim se(1 To sm * ck)  
Dim x(1 To 4, 1 To 3)  
distancia = 1E+48  
zoptimo = 1E+48  
yoptimo = 1E+20  
i = 0  
  
    For h = 1 To sm  
  
        Z(h) = 1E+24  
  
        k = 0  
  
        Do Until k = Km
```



```

k = k + 10

cu = cutot / (k + h)
i = 1 + i
'variables

s(i) = h

'funciones objetivo
If n = 1 Then
nodo1 camino, lamda, miu, dtr, k, h
Excel.Run ("Hoja2.nodo1")
Kill camino & "nodo1finito.mdb"
Excel.Sheets(2).Select
Lq = Excel.Range("a1").value
Salidas = Excel.Range("a2").value
End If

If n = 2 Then
nodo2 camino, lamda, miu, dtr, x(2, 1), x(1, 1), k, h
Excel.Run ("Hoja3.nodo2")
Kill camino & "nodo2finito.mdb"
Excel.Sheets(3).Select
Lq = Excel.Range("a1").value
Salidas = Excel.Range("a2").value
End If

If n = 3 Then
nodo3 camino, lamda, miu, dtr, x(2, 1), x(1, 1), x(2, 2), x(1, 2), k, h
Excel.Run ("Hoja4.nodo3")
Kill camino & "nodo3finito.mdb"
Excel.Sheets(4).Select
Lq = Excel.Range("a1").value
Salidas = Excel.Range("a2").value
End If

Z(i) = (cpt + cs) * s(i) + (cu + ctl) * (s(i) + k) + ce * Lq + Salidas * co
y(i) = Lq

```

```

'es pareto equilibrado?

'variacion

For h = 2 To li

If y1(h) - y1(h - 1) = 0 Or Z1
vz(h - 1) = 10
Else
vz(h - 1) = ((y1(h) - y1(h - 1)
/ ((Z1(h) - Z1(h - 1)) * y1(h)
End If

Next

'Mejora si o no?
|
For i = 1 To li - 1

d = Abs(1 - Abs(vz(i)))

'evalaucion de la distanci

If d < distancia Then
distancia = d
sub_optimo = s1(i)
End If

Next

If sub_optimo > 1 Then
sub_optimo = sub_optimo - 1
li = li - 1
End If

x(1, n) = sub_optimo
x(2, n) = kp1(li)
x(3, n) = Z1(li)
x(4, n) = y1(li)

```

```
Range("f" & 4 + i).FormulaR1C1 = n1p(i, 1)
Range("g" & 4 + i).FormulaR1C1 = n1p(i, 2)
Next
```

```
If M > 1 Then
For i = 1 To stdos
Range("h" & 4 + i).FormulaR1C1 = n2p(i, 1)
Range("i" & 4 + i).FormulaR1C1 = n2p(i, 2)
Next
End If
```

```
If M > 2 Then
For i = 1 To sttres
Range("j" & 4 + i).FormulaR1C1 = n3p(i, 1)
Range("k" & 4 + i).FormulaR1C1 = n3p(i, 2)
Next
End If
```

```
Sheets(2).Select
Range("lamda") = lamda
Range("miu") = miu
Range("Nodos") = M
Range("dtr") = dtr
Range("capuno") = capuno
Range("capdos") = capdos
Range("captres") = captres
```

```
Range("cpt") = cpt
Range("ctl") = ctl
Range("cutot") = cutot
Range("cs") = cs
Range("ce") = ce
```

```
Range("pesol") = pesol
```

```
Range("ccc") = ccc
Range("stuno") = stuno
Range("stdos") = stdos
Range("sttres") = sttres
```

```
Range("k12") = x(3, 1) + x(3, 2) + x(3, 3)
Range("k13") = x(4, 1) + x(4, 2) + x(4, 3)
```

```
Range("k16") = x(1, 1)
Range("l16") = x(1, 2)
Range("m16") = x(1, 3)
```

```
Range("k18") = x(2, 1)
Range("l18") = x(2, 2)
Range("m18") = x(2, 3)
```

```
Excel.Workbooks.Close
End If
```

```
UserForm2.Label7.Caption = x(3, 1) + x(3, 2) + x(3, 3)
UserForm2.Label8.Caption = x(4, 1) + x(4, 2) + x(4, 3)
UserForm2.Label9.Caption = x(1, 1)
UserForm2.Label10.Caption = x(2, 1)
UserForm2.Label11.Caption = x(1, 2)
UserForm2.Label12.Caption = x(2, 2)
UserForm2.Label13.Caption = x(1, 3)
UserForm2.Label14.Caption = x(2, 3)
UserForm2.Show
```

```
End Sub
```