

Simulación Versus Modelos de Redes de Colas para Estimar Performance de Redes Informáticas

Resumen: En el presente trabajo se busca determinar la aproximación de los resultados obtenidos mediante dos técnicas de modelización para evaluar la performance de un sistema de información, la primera basada en la simulación por eventos discretos y la segunda mediante modelos operacionales que responden a la teoría de las redes de cola. El grado de proximidad de resultados permitirá concluir cuando los modelos operacionales pueden servir para validar los resultados de la simulación ante la falta de datos del sistema real. El desarrollo de este trabajo considera un sistema de comercio electrónico, en el que se estiman métricas de desempeño por simulación y por modelo operacional.

Palabras Claves: Simulación; Modelo Operacional; Performance; Validación.

Abstract: The present work seeks to determine the approximation of the results obtained by two modeling techniques to evaluate the performance of an information system, the first based on discrete event simulation and the second through operational models that match the queue network theory. The closeness of results allows us to conclude when operational models can be used to validate the simulation results in the absence of data from the real system. The development of this paper considers a system of electronic commerce, in which performance metrics are estimated by simulation and operational model.

Keywords: Simulation; Operational Model; Performance; Validation.

Esteban J. Massuh, José R. Basualdo, Lisandro D. Gallego

CIDISI (Lavaise 610, Santa Fe, Argentina), Facultad Regional Santa Fe, UTN.

INTRODUCCIÓN

Cuando se busca analizar los comportamientos de sistemas informáticos se aplican técnicas de evaluación de performance experimentales o analíticas, para disponer de medidas o métricas cuantitativas tales como las utilizaciones, las velocidades de procesamiento y los tiempos de respuesta. Para ese fin se consideran los recursos de hardware y de software y la carga de trabajo de los usuarios. Ambos tipos de técnicas implican el desarrollo de estrategias de mediciones y de estimaciones. En el caso de técnicas experimentales existen monitores de los propios sistemas operativos como así también aplicaciones que emulan el trabajo de usuarios y analizan distintos escenarios. Entre las técnicas analíticas se aplican los modelos analíticos y de simulación. En el caso de los modelos analíticos se tienen en cuenta a los recursos como redes de colas y se emplea la teoría relacionada. Para los modelos desimulación, el enfoque a los sistemas de eventos discretos son los más adecuados para representar a los sistemas informáticos, dado que la carga de trabajo de los usuarios es aleatoria y se puede representar con distribuciones de probabilidad. Asimismo los tiempos de servicio de los recursos de hardware pueden ser variables, con un parámetro de distribución de probabilidad antes que un valor constante [Menascé y Almeida (1998)].

Tanto los modelos analíticos provenientes de la teoría de redes de cola como la simulación requieren de datos de entrada o parámetros que se obtienen de la monitorización de los recursos de hardware. De esta forma siempre hay una relación entre técnicas experimentales y de predicción. No obstante, predecir comportamientos en un sistema real con herramientas de monitorización resulta dificultoso ya que algunas de ellas si bien pueden saturar a los sistemas informáticos, no representa siempre la situación real. Y además cuando se desea analizar cambios en la

configuración de los recursos, como aumentar la velocidad de procesamiento, en términos prácticos significa disponer de recursos más rápidos, con la inversión correspondiente, sin saber si se justifica o no. En ese sentido, las técnicas analíticas tienen esa capacidad de predicción. Pero los resultados obtenidos mediante técnicas analíticas y/o simulación deben validarse con datos del sistema real. Sin embargo, no siempre se disponen de los mismos, razón por la cual una de las formas de comparar los resultados de una técnica analítica es hacerlo con los correspondientes a una simulación [Banks (2010)].

El interrogante es conocer si en todos los casos se pueden comparar los resultados obtenidos de dos técnicas de estimación en todos los escenarios que se implementen. Por esta razón, el objetivo del presente trabajo es el de comparar resultados de una simulación con enfoque de eventos discretos realizada con una herramienta de simulación DEVS JAVA [Ziegler (2003)] y un modelo analítico de la teoría de redes de cola para un sistema informático de comercio electrónico [Chezzi et. al (2007)]. DEVS JAVA proviene de implementar el formalismo DEVS (DiscretEventSystemSpecification) en una herramienta de simulación programada en JAVA. La ventaja principal de esta herramienta es representar problemas complejos con un modelado jerárquico, donde los elementos de un sistema pueden agruparse en un modelo acoplado y simplificar la representación o reutilizarse en otros modelos.

El sistema en estudio es el de comercio electrónico del tipo Business toConsumer (B2C), ya que es uno de las formas actuales de comercialización de creciente popularidad y cuyo desempeño de negocios depende de los recursos tecnológicos del sitio web donde está implementado [Menascé y Almeida (2000) - Chezzi (2013)]. La estimación de la performance a nivel de recursos tecnológicos y de negocios mediante la simu-

lación o los modelos analíticos de teoría de redes de cola provee información de importancia a la hora de tomar decisiones tanto en relación a los beneficios económicos como a la calidad de servicio de los recursos informáticos.

Teoría de redes de colas

En las computadoras es frecuente encontrar a diversos trabajos compartiendo un número limitado de recursos, tales como la CPU, los discos u otros dispositivos [Dominguez et. al (2015)]. Generalmente, sólo uno de los trabajos puede utilizar el recurso mientras los demás esperan en cola.

La teoría de colas es una herramienta matemática que permite cuantificar el fenómeno de formación de colas. Se basa en el modelo general de colas el cual forma un modelo abstracto que se denomina estación de servicio, conformada por un centro de servicios y una cola de espera de aquellos clientes que no pueden ser atendidos.

En la figura 1 se muestra el esquema de un modelo de cola con centro simple de servicio.



Fig. 1 - Modelo de cola.

El centro de servicio tiene asignado un tiempo de servicio para atender los requerimientos que arriban según un tiempo entre llegadas. Estos pueden ser atendidos o esperar en cola en caso que el centro esté ocupado.

En los sistemas informáticos, cada uno de los recursos que lo componen es análogo al funcionamiento del centro de servicio con su cola y los requerimientos de usuarios. El tiempo de respuesta de un requerimiento es la suma del tiempo que estuvo en cola y del tiempo de servicio en el centro para su

atención. Según la teoría de colas así se determinan los tiempos de residencia del requerimiento en los centros de servicio.

El modelo de cola de la Figura 1 tiene dos parámetros: la tasa de arribo de los clientes (intensidad de la carga) y el requerimiento de servicio promedio de un cliente (demanda de servicio). Con ecuaciones de las leyes fundamentales se pueden obtener medidas de performance tales como utilización del recurso, tiempo promedio del requerimiento en atención (tiempo de residencia), número promedio de requerimientos en cola y velocidad de procesamiento. Es importante destacar que el número promedio de clientes en cola incluye los requerimientos en cola y en servicio.

La teoría de colas representa al sistema completo como un solo centro. La teoría de redes de colas permite representar a cada uno de los recursos de los sistemas informáticos, conectados entre sí como una red de cola.

La Figura 2 presenta una red de colas compuesta de una CPU y un Disco. Los requerimientos arriban a la CPU y son atendidos si esta se encuentra desocupada, en otro caso quedan en cola. Una vez atendidos por la CPU pasan al Disco, el cual los procesa. Luego, los requerimientos vuelven a la CPU, para finalmente salir del sistema como requerimientos atendidos.

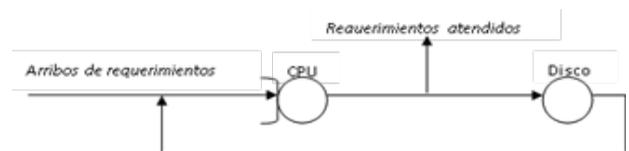


Fig. 2 - Modelo de red de colas.

Los parámetros de este modelo son análogos al de la teoría de colas con la diferencia que existen demandas de servicio para cada recurso de la red. Para la obtención de las medidas de performance de los recursos y del sistema se plantean ecuaciones basadas en Leyes

Fundamentales: Ley de Little, Ley de Utilización, Ley del Tiempo de Respuesta y Ley de Flujo Forzado. Con estas leyes y con los distintos enfoques para los modelos se generan algoritmos asociados para estimar las medidas de performance.

La arquitectura del sistema en estudio presenta un estilo basado en capas como la indicada en la Figura 3, la primera capa consiste del cluster de servidores Webs, la segunda, del cluster de servidores de Aplicación y la tercera, del cluster de Base de Datos. La conexión entre capas se efectúa a través de tres redes LANs. Cada LAN se considera una conexión de tipo full duplex. El arribo de los requerimientos de Internet se hace a través del Router, quien los deriva a los clusters del Sitio de Negocios Electrónicos. Cada grupo de servidores posee un coordinador que toma los requerimientos y los asigna al servidor desocupado.

Se puede representar al sistema de comercio electrónico mencionado anteriormente de 3 capas de varios servidores cada una como un modelo de redes de cola de 3 capas con sólo un servidor en cada una de ellas, tal como se muestra en la Figura 4.

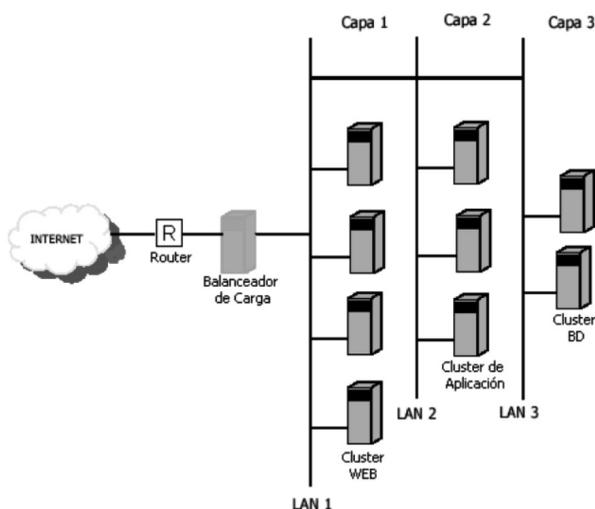


Fig. 3 - Sistema de 3 capas de servidores para comercio electrónico.

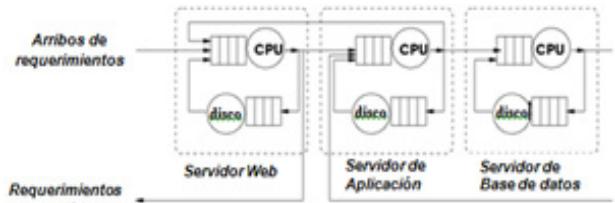


Fig 4 - Modelo de redes de cola del sistema de 3 capas para comercio electrónico.

Estos modelos de redes de colas se pueden resolver mediante las ecuaciones de las leyes antes mencionadas en una planilla de cálculo [Menascé y Almeida (1998)].

Simulación de sistemas por eventos discretos

Los sistemas reales se representan por lo general mediante el enfoque de sistemas discretos o continuos. En el primer caso, en los sistemas se producen cambios de estado instantáneos y en el cual permanecen hasta que un evento provoca un cambio de estado. En los sistemas continuos los cambios son instantáneos y continuos [Banks (2010)].

Los sistemas informáticos, y entre ellos los de comercio electrónico, se representan como sistemas discretos y la simulación de los mismos se realiza teniendo en cuenta este enfoque. De las numerosas herramientas, el formalismo DEVS y la herramienta de simulación DEVS JAVA permite una abstracción del problema en los Modelos de Recursos, Clientes y Funcional.

DESARROLLO

En el sistema de comercio electrónico se identificaron 3 tipos de sesiones denominadas clases 1, 2 y 3. La clase 1 visita la página principal (Home), la clase 2 busca un producto y la clase 3 hace la adquisición del producto.

En la Tabla I se encuentran los tiempos medios entre arribos de cada una de las clases. Es importante destacar que en los modelos de redes de colas los

tiempos medios entre arribos son exactamente los que aparecen en la tabla, mientras que en la simulación se calculan a partir de una distribución de probabilidad con media en esos valores.

En la Tabla II se hallan las demandas de servicio para cada una de las clases en los diferentes recursos. En la misma se puede observar que la clase 1 no hace uso de los servidores de aplicación ni de los servidores de base de datos porque sólo visita la página principal. La clase 2 no requiere de los servidores de base de datos. La clase 3, a diferencia de las anteriores, presenta una demanda de todos los recursos.

Clase 1	Clase 2	Clase 3
1	3	15

Tabla 1 -Tiempos medio entre arribos (segundos).

Recurso	Clase 1	Clase 2	Clase 3
Router	0,0004	0,0004	0,0004
LAN 1	0,0155	0,0155	0,0155
LAN 2	0,0155	0,0155	0,0155
LAN 3	0,0155	0,0155	0,0155
SERV WEB	0,712	0,699	0,724
SERV APP	0	0,891	0,909
SERV BD	0	0	1,333

Tabla 2 - Demandas de servicio para los recursos (segundos).

Los modelos de redes de colas se pueden resolver mediante una planilla de cálculo en Excel, denominada OpenQN (Open Queue Network Solver), cuya autoría es de [Menascé y Almeida (1998)]. La herramienta requiere definir el número de clases de trabajo, las velocidades y las demandas de servicio o tiempos de procesamiento de los recursos del sistema en estudio. En las Tablas 1 y 2 figuran los datos del sistema de referencia.

Para la simulación y sobre la base del Modelo de Recursos se formaliza la representación con DEVS, como Framework Orientado a Objetos. Cada Framework describe una componente del sistema en estudio como modelo atómico, definido por ciertas características, que se puede interconectar con otros modelos atómicos de forma modular y jerárquica. Un conjunto de modelos atómicos se constituye en un modelo acoplado o sistema más complejo, que a su vez termina cumpliendo la función de un nuevo modelo atómico, que se interconectará con otros.

El Formalismo es independiente de todo mecanismo de simulación. Como Modelo Funcional se construye el Diagrama de Interacciones que establece las interconexiones entre los distintos elementos del sistema. Estas interacciones generan los eventos que completan el Framework. Se plantean las métricas para el diseño de las salidas. Una vez obtenido el Framework DEVS se lo codifica en la herramienta DEVS JAVA para la ejecución de la simulación.

Esta herramienta se organiza en dos módulos fundamentales: un Marco Experimental y el Sistema en estudio. El Marco Experimental representa la carga de trabajo en el sitio web del comercio electrónico, la estructura navegacional de los usuarios en el sitio y las funciones del modelo de negocios, como así también el modelo de recursos de procesamiento.

Los parámetros utilizados para la simulación son los mismos que los del modelo de redes de colas.

Finalmente para analizar si es posible usar modelos de redes de cola como elemento para validar los resultados de la simulación se comparan las medidas de performance y se determinan las situaciones en que ambas técnicas dan resultados aproximados o muy diferentes. Por esta razón se hacen cambios en las demandas de servicio como escenarios para el análisis.

RESULTADOS

Los cambios introducidos en los escenarios que se presentan a continuación son realizados sobre estos datos.

Escenario 1: Demandas originales (Tablas 2) y arribos de Tabla 1.

	X modelo	X simulación	Diferencia Porcentual	R modelo	R simulación	Diferencia Porcentual
Clase 1	1	0,808231	23,73%	107,9254	32,6528	230,52%
Clase 2	0,3333	0,209224	59,32%	109,1246	63,90503	70,76%
Clase 3	0,0667	0,040417	64,94%	115,7659	69,55271	66,44%

Tabla 3 - Resultados escenario 1

(*) En los modelos abiertos, la velocidad de procesamiento (X) coincide con el la velocidad de arribo.

Escenario 2: Reducción en un 50 % de las demandas en los servidores.

	X modelo	X simulación	Diferencia Porcentual	R modelo	R simulación	Diferencia Porcentual
Clase 1	1	0,99923	0,08%	0,41540	0,38934	6,69%
Clase 2	0,333	0,33174	0,48%	0,86358	1,20867	28,55%
Clase 3	0,0667	0,06594	1,09%	1,55978	2,23160	30,10%

Tabla 4 - Resultados escenario 2

Escenario 3: Incremento en 4 veces las demandas de los servidores.

	X modelo (*)	X simulación	Diferencia Porcentual	R modelo	R simulación	Diferencia Porcentual
Clase 1	1	0,998	0,15%	0,81758	0,74650	9,52%
Clase 2	0,3333	0,3315	0,55%	1,73470	2,35426	26,32%
Clase 3	0,0666	0,0659	1,15%	3,14457	4,63288	32,12%

Tabla 5 - Resultados escenario 3

Escenario 4: Las demandas en los servidores Web y BD son 100 % respecto a las originales, mientras que en el de aplicación son 1,5 veces más grandes.

	X modelo	X simulación	Diferencia Porcentual	R modelo	R simulación	Diferencia Porcentual
Clase 1	1	1,02520	2,46%	1,82202	1,59382	14,32%
Clase 2	0,3333	0,35217	5,35%	3,22310	4,47958	28,05%
Clase 3	0,0667	0,06208	7,38%	4,67525	7,55793	38,14%

Tabla 6 - Resultados escenario 4

DISCUSIÓN

Como se puede observar en la Tabla III, del escenario 1, los resultados obtenidos con la simulación difieren en bajos porcentajes en relación a la velocidad de procesamiento X, pero esa diferencia es mayor para los tiempos de respuesta R, siendo el caso más notable la diferencia para la clase 3.

Respecto al escenario 2, en donde se reducen en todos los servidores las demandas de servicio a la mitad, las diferencias porcentuales se mantienen en tendencia dentro del mismo sistema pero disminuyen los valores en todos los casos respecto al escenario 1 o sistema original. Esto ocurre porque el sistema es más rápido para procesar los requerimientos.

La Tabla V muestra un escenario degradado dado que las demandas de servicio o tiempos de procesamiento de los recursos son mayores, y las diferencias entre la simulación y los modelos operacionales de la teoría de redes de colas son notables. Esto ocurre porque estos parámetros se acercan al estado de saturación del sistema, en donde las utilizaciones de los recursos están cercanas al 100 % y se generan colas de espera, como se muestra en el aumento de los tiempos de respuesta R de todas las clases en todos los recursos. Por otra parte los modelos de redes de cola tienen una serie de hipótesis

que dejan de cumplirse cuando el sistema se acerca a la saturación, por lo cual deja de ser una herramienta de predicción válida. En ese sentido, la simulación puede seguir representando lo que ocurre en el sistema aún en situación límite o de recursos críticos.

La Tabla VI es una situación intermedia entre el sistema original (Tabla III) y el sistema cercano a la saturación (Tabla V). Las diferencias en las medidas de performance del escenario 4 son mayores que las del escenario 1 pero menores que las del escenario 3, por lo que se puede decir que la tendencia del incremento de las diferencias de las medidas con el aumento de las demandas (y con ello, las utilizaciones) se mantiene también en este caso.

Por lo cual, en escenarios alejados de la saturación o del 100% de utilización de los recursos, las técnicas analítica y de simulación dan resultados aproximados entre 0 y menos de un 10% en velocidades de procesamiento y entre 7% y hasta menos de un 30% en tiempos de respuesta.

CONCLUSIONES

Dada la problemática inicial y a partir de los resultados obtenidos se puede concluir que los datos obte-

nidos del modelo operacional pueden usarse para la validación de los resultados de la simulación ante la falta de datos provenientes del sistema real, siempre y cuando las demandas no sean tan grandes tal que provoquen que las utilizaciones de los recursos se aproximen al 100% o al estado de saturación. Por lo tanto, la validación de la simulación puede hacerse con el modelo operacional de las redes de colas. Pero cuando las utilizaciones aumentan y tienden al 100%

las dos técnicas difieren notablemente, siendo los resultados de los modelos mayores porque dejan de cumplirse las hipótesis del modelo de redes de cola.

RECONOCIMIENTOS

Los autores agradecen a la Universidad Tecnológica Nacional por el financiamiento del proyecto de investigación en el que se enmarca el trabajo.

REFERENCIAS

Chezzi, C. M., Villamonte, A., Tymoschuk, A. R. (2007). "Modelado y Simulación de Performance de Transacciones Electrónicas Comerciales". *Proceeding IX Workshop de Ciencias de la Computación (WICC 2007)* (págs. 443-447).

Chezzi, C. M. (2013) "Modelado y Simulación de Desempeño de Procesos de Comercio Electrónico". *Tesis Doctorado en Ingeniería Mención Ingeniería de Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Santa Fe. Santa Fe, Argentina.*

Dominguez, M., Tymoschuk, A. R. (2015). "Modelos de performance: fundamentos basados en la teoría de cola y de redes de colas. Cátedra: Análisis y Evaluación de Performance en Arquitecturas Distribuidas". *UTN Facultad Regional Santa Fe.*

Zeigler, B., Sarjoughian, H. (2003). "Introduction to DEVS Modeling & Simulation with JAVA: Developing Compo-

nent-Based Simulation Models". Arizona. Web: http://www.acims.arizona.edu/EDUCATION/ECE575Fallo3/notes/Manuscript_MSDJ_090903.pdf

Menascé, D., Almeida, V. (2000). *Scaling for E-Business, Technologies, Models, Performance and Capacity Planning*. Prentice Hall Professional.

Menascé, D., Fonseca, R., Almeida, V., Mendes, M. (2000). "Resource Management Policies for E-commerce Servers". *Performance Evaluation* 27 (págs. 27-35).

Menascé, D., Almeida, V., Fonseca, R. and Mendes, M. (2000). "Business-oriented resource management policies for e-commerce servers". *Performance Evaluation* 42 (págs. 223-239).

Banks, J., Carson II, J.S., Nelson, B. L., Nicol, D.M. (2010). "Discrete-Event System Simulation", 5ta Edición, Prentice Hall Ed. New Jersey.