Aplicación de Redes de Petri a la Evaluación de Desempeño de Sistemas de Comunicaciones

Mario E. Salazar P.[1] Néstor Peña[2]

Abstract

Performance evaluation of communications systems is usually done by analytical techniques or simulation. The analytical approach is normally based on a probabilistic description of the system operations in terms of Markovian processes or queuing systems, but only is directly applicable for simple systems, or simplistic models of complex systems. Simulation is a powerful tool but requires enormous computations to yield accurate performance estimators. In this paper we present an alternative technique: Petri Nets, a graphical and mathematical modeling tool applicable to many systems. We used Stochastic Petri Nets to model and evaluate queuing systems and some protocols. Simple and accurate models were developed. The main definitions relating Petri nets in general, and to the particular classes of Petri nets are summarized in Section 1. The following sections contains queuing models (Section2), Stop and Wait and Aloha protocols (Section 3), DSP protocol (Section 4). Section 5 concludes the paper.

Palabras claves

Redes de Petri, evaluación de desemepeño, modelos Markovianos, protocolos de comunicación, modelos de tráfico.

1. Redes De Petri

El modelaje de un sistema es un proceso que abstrae un sistema del mundo real en un conjunto de parámetros que lo caracteriza. La representación gráfica del sistema en dicho proceso puede facilitar su comprensión y análisis. Las Redes de Petri (RdP) son una herramienta gráfica y matemática de modelaje aplicable a muchos sistemas, especialmente en aquellos con características asincrónicas, paralelas, concurrentes o estocásticas. En su forma gráfica, una RdP es un *grafo bipartido dirigido*[1] cuyos nodos se dividen en dos conjuntos disjuntos denominados *puestos y transiciones*. Las transiciones (representadas como barras) modelan eventos y los puestos (círculos) representan condiciones para dichos eventos. Únicamente se permiten conexiones entre pares de nodos de diferente clase por medio de *arcos* dirigidos. Los arcos dirigidos desde puestos hacia transiciones se denominan arcos de entrada, en tanto que aquellos arcos que se dirigen desde transiciones hacia puestos se conocen como arcos de salida. Para una transición, los *puestos de entrada* (o salida) son el conjunto de puestos que están conectados a ella por medio de un arco de entrada (salida). Adicionalmente se tienen los "tokens" (marcas) los cuales están asociados con los puestos, y se representan como puntos negros (o números) dentro de los puestos. Cada puesto contiene un número (entero) no negativo de tokens.

1 Departamento de Ingeniería Eléctrica y Electrónica, Universidad de los Andes

2 Profesor Asociado, Departamento de Ingeniería Eléctrica y Electrónica, Universidad de los Andes

La dinámica de una RdP se rige por una regla básica aplicada a las transiciones: Habilitación y disparo. Una transición está habilitada siempre y cuando haya por lo menos un token en cada puesto de entrada a ella, de acuerdo a esto, la transición T1 de la red A de la figura 1 estaría habilitada. El disparo de una transición corresponde a la realización de la actividad asociada a ella (o evento), y se indica removiendo un token de cada puesto de entrada y depositando uno en cada puesto de salida de la transición ,de esta manera, el disparo de T1 de la red A de la figura 1 conllevaría a la red B de la misma figura.

M_0 es una función que asigna a cada puesto de la red un número inicial de tokens, que corresponde únicamente a la distribución de éstos en el estado inicial.

El estado de una RdP se define por medio del número de tokens en cada puesto, y es representado por un vector (llamado marca) m=[# P_1, #P_2,.....#P_n], donde #P_i es el número de tokens en el puesto i, y n es el número de puestos de la red.

1.1 Propiedades

Secuencia de Activación

Considérese la red de la figura 2 con marca inicial M_0=[10000]. En este estado únicamente está habilitada T1, cuyo disparo conlleva al nuevo estado M_1=[01100]; de manera similar, en M1 el disparo de T3 trae el nuevo estado de la red M_2=[01010]. Es decir que se llegó M_2 a partir de Mo después de una activación sucesiva de T1 y T2. Si se denomina como S1={T1,T2} se dice que S1 es una secuencia de activación.

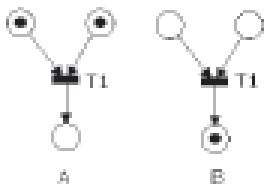

***Figura 1**: Activación y disparo.*

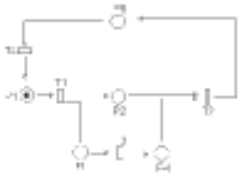

***Figura 2**: Red de Petri.*

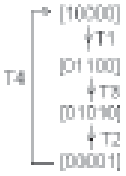

Figura 3: Árbol de cobertura.

Lo anterior se puede describir con la siguiente notación [2]:

Si $M_0<T1>\rightarrow M_1$ y $M_1<T2>\rightarrow M_2$

entonces $M_0<T1,T2>\rightarrow M_1$ ó $M_0<S>\rightarrow M_1$

Estados Alcanzables

Se denota como A(R,Mo) siendo R una RdP y Mo una marca inicial. Un estado Mi es alcanzable a partir de Mo si existe una secuencia de activación S tal que Mo resulta en el estado Mi después de ejecutar la activación de las transiciones de S. Es decir:

$A(R,Mo)=\{M_i, \exists S \; Mo<S>\rightarrow M_i\}$

La red de la figura 2 tiene 4 estados alcanzables. Es muy útil representar el conjunto de estados alcanzables de forma gráfica, para el ejemplo en cuestión dicha representación se presenta en la figura 3.

La gráfica de la figura 3 corresponde al denominado *Árbol de Cobertura de la red de Petri*.

En [1]- [4] se presentan propiedades adicionales y las técnicas de análisis de este tipo de redes clásicas.

1.2 Especificaciones temporales

La introducción de especificaciones temporales en este tipo de modelos se ha hecho (en la mayoría de casos) asociando a cada transición una temporización (tiempo que transcurre entre la habilitación y disparo). Cuando para una marca, más de una transición es factible (conflicto) aquella que tenga el menor tiempo asignado será quien se dispare; esta "política" de disparo se conoce como *modelo de carrera*."[5] El disparo de la transición que "gana la carrera" implica que la actividad asociada a dicha transición se ha realizado completamente, así, si para otra marca dicha transición se vuelve factible se procederá de igual manera a la descrita anteriormente. El comportamiento de aquellas transiciones que para una marca se habilitan pero que no se disparan puede ser especificado de diferentes maneras, por ejemplo es posible que para las futuras habilitaciones "recuerden" el tiempo que han permanecido habilitadas o simplemente no.

El uso de distribuciones exponenciales para la definición de especificaciones temporales es particularmente atractivo por dos razones. Primero, la propiedad sin memoria de la distribución exponencial hace innecesaria la distinción entre la distribución de la temporización por si misma, y la distribución del tiempo restante después de un cambio de estado, evitando así la necesidad de especificar el comportamiento de aquellas transiciones habilitadas para una marca que no fueron disparadas. Segundo, el proceso de Marca $\{M(t), t>0\}$ es un proceso estocástico que es isomorfo a una cadena continua de Markov [6], por lo tanto el análisis de la Red se puede hacer a través de técnicas numéricas aplicadas a los procesos de Markov. Se denominan como SPN (Stochastic Petri Nets) a RdP con transiciones exponenciales. A continuación se presentan otros tipos de RdP estocásticas.

1.2.1 Redes de Petri estocásticas generalizadas (GSPN):

GSPN [7] son SPN que incluyen las siguientes especificaciones:

Se permiten dos tipos de transiciones: Temporizadas exponencialmente (como en SPN) o sin temporización, denominas inmediatas y que tienen prioridad sobre cualquier otra transición.

Las tasas de disparo asociadas a cada transición pueden ser dependientes de la marca (Es decir que puede depender del número de tokens de alguno de los puestos de la red en dicha marca o estado).

Se introducen arcos inhibitorios. Cuando un puesto se conecta a una transición a través de un arco inhibitorio, ésta sólo se puede habilitar cuando el número de tokens en el mencionado puesto sea menor a la multiplicidad del arco de conexión entre los nodos en cuestión.

1.2.2 MRSPN (Markov Regenerative Stochastic Petri Nets)

Son RdP en las que se pueden incluir especificaciones temporales por medio de variables aleatorias con distribución general. El proceso de marca M(t), bajo ciertas condiciones, es un proceso regenerativo de Markov y por lo tanto existe una solución analítica [8]

1.3 Herramientas computacionales

Aunque el análisis y solución de una RdP se puede hacer, en la mayoría de los casos, resolviendo el proceso estocástico subyacente a la red, dicho proceso puede resultar bastante difícil si no se cuenta con alguna ayuda computacional, especialmente en la evaluación de modelos con gran número de estados. Afortunadamente existe una buena cantidad de herramientas computacionales, algunas de dominio público, de manera que la evaluación de desempeño del sistema modelado se puede realizar a través de un simulador. En [1] se encuentra una descripción y evaluación de las herramientas Pesim [9], Winttpn [10] y Webspn [11].

2. Sistema De Espera

En la figura 4 se presenta un modelo con el que se puede evaluar el desempeño de la mayoría sistemas de espera. Es importante destacar, que además de su sencillez, su estructura es independiente de la distribución de los tiempos de llegada o servicio, por lo que se puede utilizar para el evaluar el desempeño de sistemas para el que no existe solución analítica. Descripción:

P_1: Modela el tamaño del sistema (K); P_2: Modela los usuarios esperando a ser atendidos.

P_3: Modela los usuarios siendo atendidos; P_4: Modela el número de servidores; T_1: Transición cuya temporización corresponde a la distribución entre el tiempo de llegadas; T_2: Transición inmediata; T_3: Transición cuya temporización corresponde a la distribución del tiempo de servicio; $M_0[K\,0\,0\,C]$.

Probabilidad de bloqueo

Se dice que el sistema está "bloqueado" cuando no permite la llegada de nuevos usuarios, esto quiere decir que en el sistema ya ha alcanzado el límite de su capacidad (K usuarios). Esta probabilidad se puede obtener del modelo propuesto, como la probabilidad de que no haya ningún token en P1.

Longitud media de la cola (Lq)

En P2 se modela la cola del sistema, n tokens en P2 indican n usuarios en la cola. La longitud media de la cola es entonces el número medio de tokens en P2.

Número medio de usuarios en el sistema (L)

Se puede obtener teniendo en cuenta el número medio de usuarios en la fila y el número medio de usuarios en servicio.(Suma de fila y servicio). El número medio de tokens en P3 corresponde al número medio de usuario en servicio.

Retardo en la fila y en el sistema (Wq,W)

Se pueden obtener directamente con la ley de Little [12] para sistemas cuya llegada sigue un proceso de *Poisson*.

Con el modelo propuesto de evaluó el desempeño de los sistemas M/M/1/K; M/M/1; M/M/C/K obteniendo los resultados de la solución analítica. En la Figura 5 se presenta el número medio de usuarios en la fila y el sistema (para un sistema M/M/C/K), en función de diferentes valores de tráfico ofrecido. Se utilizó K=10, C=5. (ro=λ/cμ).

Es muy importante destacar que la estructura de los modelos propuestos es independiente del tipo de distribuciones del tiempo de servicio y de llegada, así por ejemplo, es posible utilizar el modelo de la figura 4 para evaluar el desempeño de un sistema M/G/1/K. La figura 6 presenta el desempeño para este sistema, asumiendo tiempo de atención determinístico 1[unidad de tiempo] y una tasa λ =0.5 [usuarios/unidad de tiempo]. (Sistema M/D1/K). Para K suficientemente grande se puede aproximar al modelo M/G/1, que para el caso propuesto, de acuerdo a las ecuaciones de Pollaczek-Khinchin, la fila media es 0.25 usuarios, resultado que coincide con el presentado en la figura 6. Un ejemplo del sistema **M/G/C/K** se puede apreciar en la figura 7, donde se presenta la longitud de la fila para este sistema, tomando el tiempo de servicio como una variable uniforme distribuida entre [0.5 y 1], y el tiempo entre llegadas exponencial con tasa 2 [usuarios/unidad de tiempo].

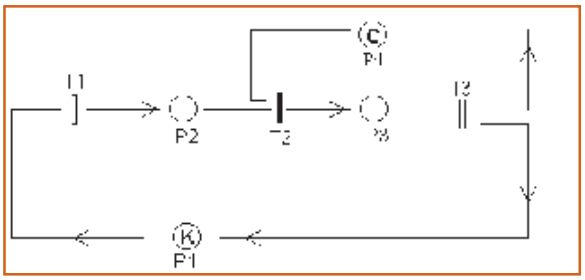

***Figura 4**: Modelo general de sistema de espera.*

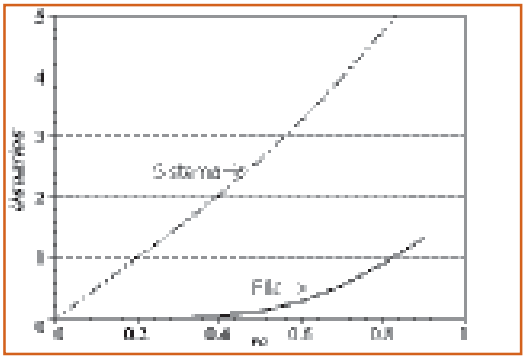

***Figura 5**: Longitud fila y sistema(C=5 K=10)*

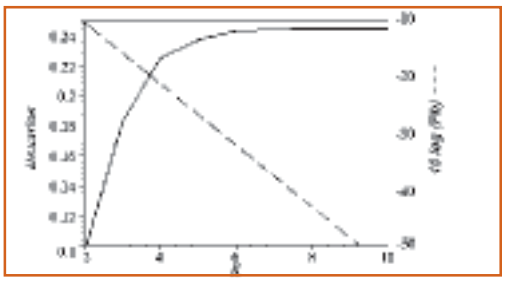

***Figura 6**: Probabilidad de bloqueo y usuario en fila (M/D/1/K)*

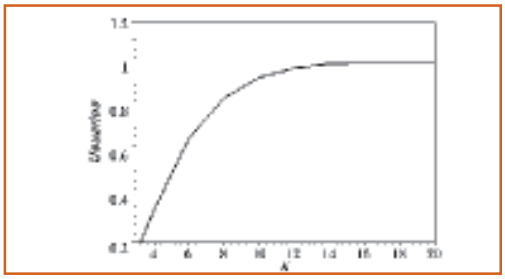

Figura 7: Longitud fila (M/G/2/K. G ~ uniforme [0.5; 1]).

3. Protocolos de Comunicaciones

3.1 Protocolo Parar y Esperar

En este protocolo, cada vez que se envía un paquete, el transmisor debe esperar una señal de confirmación de envió (*ack*) por parte del receptor, antes de enviar el siguiente [13]. Si finaliza la temporización del emisor (*timeout*) antes de recibir el respectivo ack entonces se retransmite el paquete. En la figura 8 se presenta el modelo desarrollado para este protocolo.

La lista de puestos se presenta a continuación: P1:disponible, P2:Listo, P3:Rec1, P5:falla, P4 Esperar, P6: Recibido, P7:Rec2, P8:Listo2. El estado inicial es el estado M0=[100000010], en el cual la única transición habilitada es *NMsg*, cuya temporización corresponde a una tasa λ de llegada de mensajes. Cuando llega un nuevo mensaje para ser transmitido se llega al estado M1=[010000010], e indica la condición previa a la transmisión del mismo. Con la transición *enviar* se especifica la velocidad de transmisión del emisor, la transición inmediata p modela la pérdida del mensaje enviado. El valor de la transición q es por lo tanto 1-p. Es importante destacar que las transiciones *Rec* y *Ack Rec* deben modelar tanto el retardo en cada transmisión, como el tiempo de procesamiento respectivo. Resolviendo para estado estable se obtienen las probabilidades de estar en cada estado posible de la RdP, y a partir de dichos resultados se pueden calcular las siguientes medidas de desempeño:

Caudal

Si se denota como p, la probabilidad de que haya un tóken en el puesto Disponible, entonces el Caudal (S) se puede calcular como $S = \lambda.p$

Retardo

Se puede hallar con la ley de Little.

En la Figura 9 se presentan los resultados para p=0.01 y p=0.001. Los valores de las transiciones corresponden a una velocidad de transmisión de 9600 b/s, y paquetes de 1026 bits.

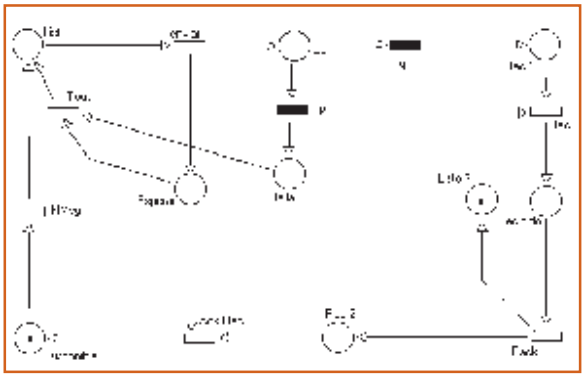

***Figura 8**: Modelo Petri Parar y Esperar.*

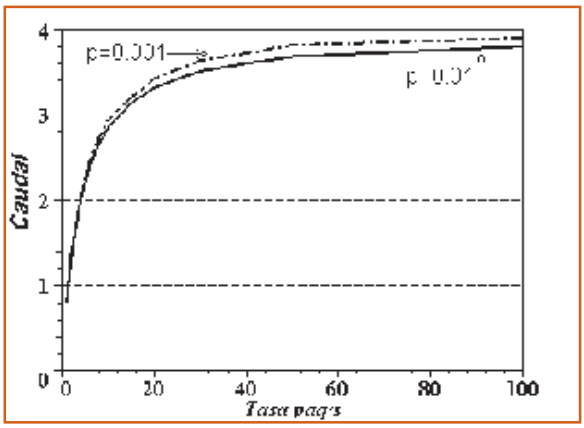

***Figura 9**: Caudal protocolo parar y esperar p=0.01; p=0.001.*

El máximo rendimiento se presenta cuando se transmite en un medio libre de pérdidas (p=0), y es aproximadamente 4.14 Paquetes/seg [1] resultado que coincide con la cota analítica propuesta por Schwartz en [13].

3.2 Protocolo Aloha Puro

El protocolo *Aloha puro* es una técnica de acceso aleatorio a un medio común [14]. Suponiendo que el tiempo de transmisión de un paquete es 1 (unidad de tiempo normalizada), entonces no se producirá colisión si ninguna otra estación intenta transmitir en un tiempo antes de 2 unidades de tiempo [11]. En la figura 10 se presenta el modelo desarrollado para este protocolo.

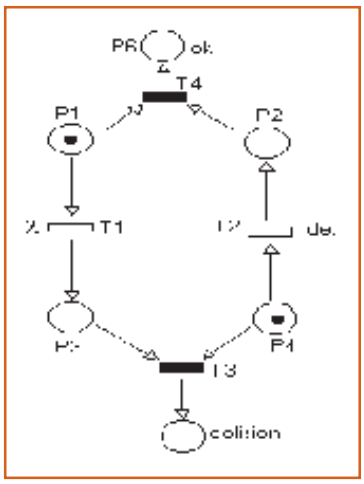

Figura 10:Modelo Aloha.

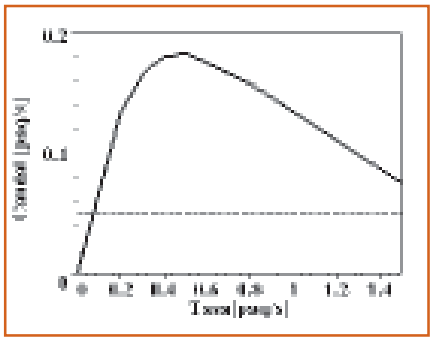

Figura 11: Caudal normalizado Aloha Puro

Si la transición T2 (determinística) se dispara antes de T1, entonces se tendrá que el mensaje fue enviado y ningún otro usuario intentó transmitir, si por el contrario, T1(exponencial) se dispara antes de T2 entonces se produce una colisión. La figura 11 presenta el desempeño de este protocolo. La solución analítica conlleva al mismo resultado[15]. En[1] se presentan los modelos de los protocolos Retorno a N, CSMA/CD.

4. Optimización Temporizador Protocolo DSP

En [16] se presentan dos protocolos (TSP,DSP) distribuidos para la configuración y manejo de fallas de un conjunto de dispositivos autónomos inteligentes. DSP se encarga de tareas y aplicaciones, mientras TSP de comunicaciones y dispositivos. Adicionalmente a detectar fallas, DSP se encarga de localizar y seleccionar un dispositivo que pueda realizar la tarea que ha falla-

do. La asignación de tareas del protocolo DSP se fundamenta en una rutina 4 *hand shake* que contiene los siguientes mensajes:**Who can xx** (wc): mensaje broadcast donde xx es la identificación de la tarea que ha fallado. **I can** (ic): respuesta de un dispositivo afirmando que puede realizar xx. **Ok run xx** (run x): respuesta a mensaje I can. **I am running xx** (i am): Confirmación positiva mensaje runxx. Se desarrollaron dos modelos para evaluar el comportamiento para diferentes valores de temporización. En la figura 12 se presenta uno de los modelos propuestos.

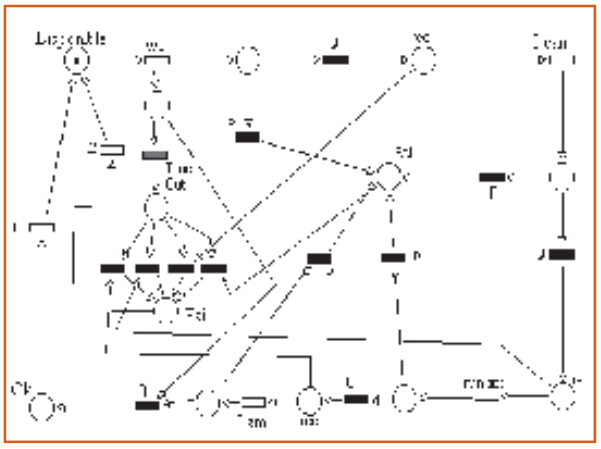

***Figura 12**: Modelo GSPN Asignación.*

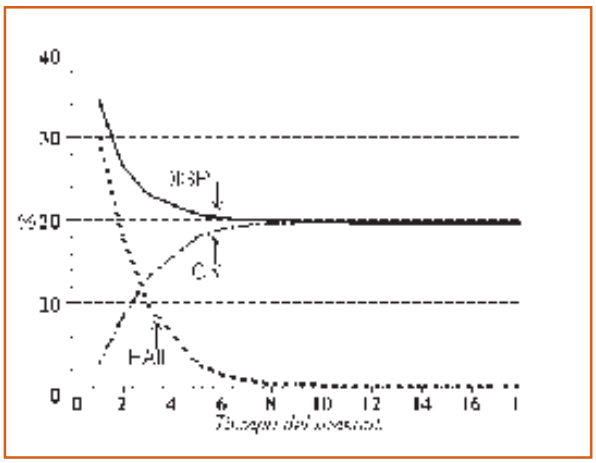

***Figura 13**: Probabilidades de estado.*

Un adecuado valor de *timeout* debe permitir la asignación efectiva de la tarea.

Una vez que se envía el medio alguno de los 4 posibles mensajes existe una probabilidad P de que no llegue a su destino. Cuando ocurre una pérdida del mensaje, se llega a un estado denominado Fail , en el que la única transición factible es *TimeOut*. La optimización busca minimizar la probabilidad de que el sistema llegue al estado Fail, maximizar la probabilidad de estar en el estado *OK*(adecuada asignación), maximizar la probabilidad del estado disponible (*Disp*), y minimizar el retardo de la rutina.

En las figuras 13 y 14 se presentan las probabilidades de los estados que se quieren optimizar y el retardo (el tiempo está normalizado respecto al tiempo del envío de un mensaje). Se asumió P=0.1%

Con la información de las gráficas 12 y 13 se tienen criterios cuantitativos para escoger el valor de *timeout* óptimo, de acuerdo a la importancia que considere en cada uno de los 4 aspectos considerados. En [1] se presentan otros resultados obtenidos de esta optimización.

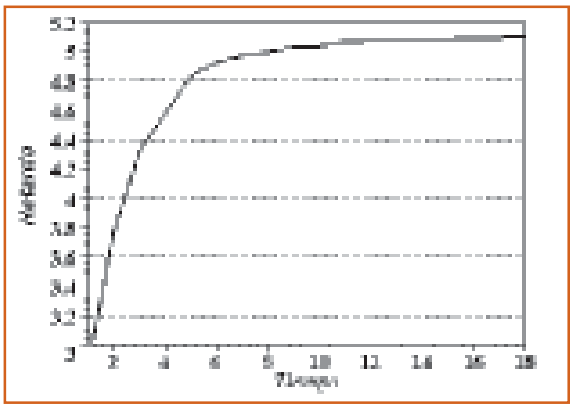

***Figura 14**: Retardo.*

5. Conclusiones

Se presentaron las Redes de Petri Estocásticas como una alternativa a las técnicas usuales empleadas en la evaluación de desempeño de sistemas de comunicaciones.

La exactitud de los resultados obtenidos permiten concluir que son una herramienta adecuada y eficiente en la evaluación de desempeño.

La exactitud de las medidas de desempeño son resultado de la solución analítica numérica subyacente a una RdP.

El modelaje de un sistema por medio de una RdP Estocástica se hace de una manera más fácil y efectiva que usando un modelo Markoviano directamente, o un modelo probabilístico, además es considerablemente más fácil de entender.

Las RdP permiten evaluar modelos para los que no existe solución analítica.

Referencias

[1] Mario E. Salazar. "Aplicación de Redes de Petri Estocásticas a la evaluación de desempeño de protocolos de comunicaciones". Tesis Maestria en Ing Eléctrica. Universidad de los Andes 2001.

[2] Hassane Alla "Du Grafcet aux réseaux de Petri". Hermes 1992.

[3] Tadao Murata. "Petri Nets: Properties, Analysis and Applications". Proceedings of the IEEE, Vol 77 No 4 Abril 1989.

[4] Robert Valette. " Les Résaux de Petri". http://www.laas.fr/~robert.

[5] Andrea Bobbio, Miklós Telek. "Non exponential stochastic Petri nets: an overview of methods and techniques". Computer Systems 339-351 1998.

[6] Cristoph Lindemann "A characterization of the Stochastic Proces underlaying a Stochastic Petri Net". IEEE Transactions on Software Engineering. Vol 20 No 7 Julio 1994.

[7] Ajmone Marsan. "An introduction to Generalized Stochastic Petri Nets".Microelectron.Reliab. Vol 31 No 4 1991.

[8] Hoon Choi. "Performance and Reliability modeling using Markov Regenerative Stochastic Petri Ners" Phd Dissertation. Duke University 1993.

[9] PeSim: www.aut.utt.ro/~mappy/petri/tools/pesim.htm

[10] Winttpn: www.diit.unict.it/users/scava/TPTPN/WinTTPNDownload.html.

[11] WebSpn: http://sun195.iit.unict.it/~webspn/webspn2.

[12] J.D Little. " A Proof of the Queueing Formula L=lW". Operarions Res. Vol 9 No 3 1961.

[13] Mischa Schwartz "Redes de Telecomunicaciones".Addison Wesley 1994.

[14] Robert Gallager, Dimitri Bertsekas. " Data Networks" Prentice Hall . 1992.

[15] Andrew Tanenbaum. "Redes de computadoras". Pearson 1997

[16] J.Pimentel. "Distribuited Protocols for Automatic, Reliable Configuration and Fault Management of Automatic Intelligent Devices".