

Utilización de Ciclos Ociosos de Servidores de Internet

Champredonde Raúl¹
Pasini Ariel²
La Battaglia Juan³

Laboratorio de Investigación y Desarrollo en Informática⁴
Facultad de Informática - Universidad Nacional de La Plata

1. Abstract

Pueden encontrarse numerosos trabajos de investigación sobre el uso de los recursos ociosos de una red de estaciones de trabajo.

Este trabajo intenta aplicar las experiencias obtenidas a partir de esos trabajos para aprovechar la potencia de cálculo no utilizada por muchos servidores de Web, servidores de correo electrónico, y otros que puedan ser alcanzados por medio de las tecnologías de Internet.

Keywords: Tiempo ocioso, Sistemas Distribuidos, Internet

¹ Profesor Adjunto semidedicación. Becario de Perfeccionamiento CONICET.
rchampre@lidi.info.unlp.edu.ar

² Ayudante Alumno. Becario LIDI. apasini@lidi.info.unlp.edu.ar

³ Becario LIDI. juanlb@lidi.info.unlp.edu.ar

⁴ Calle 50 y 115 Primer Piso. 1900 La Plata. Bs.As. Tel/Fax. 54 221 422 7707

2. Introducción

El aprovechamiento de recursos libres es un área de investigación que desde hace algunos años ha estado en crecimiento. Un crecimiento alimentado por diversos estudios que demostraron que un porcentaje importante de las estaciones de trabajo de una red (o de un cluster) permanecen ociosas durante una buena parte del tiempo [6][7][8].

En 1982, los programas “Worm” de Shoch y Hupp [1], utilizaron workstations Alto ociosas, aunque no tenían acceso a un sistema de archivos compartido.

Hagmann [2], Litzkow [3], Theimer [4] y Nichols [5] contribuyeron con trabajos que, con ciertas diferencias entre sí, todos utilizaron las estaciones de trabajo libres de otros usuarios.

Existen también proyectos que explotan un paralelismo súper masivo como el proyecto SETI [9].

Todos estos trabajos están basados en redes de estaciones de trabajo y han logrado resultados importantes. Sin embargo, si se tiene en cuenta que:

- habitualmente las estaciones de trabajo se encuentran funcionando una parte del día, mientras que los servidores de Internet se encuentran funcionando permanentemente,
- los servidores de Internet con objetivos locales a un país, o ciudad, tienen su mayor ocupación durante las horas del día, permaneciendo durante la noche con escasa o ninguna utilización,
- mientras es de día en una parte del mundo es de noche en otras,
- la potencia de cálculo de las máquinas en las que corren los servidores de Internet es generalmente muy superior a la de las estaciones de trabajo convencionales,
- los tiempos de respuesta requeridos por los usuarios de Internet no constituyen el mayor de los requerimientos de estos servicios,

se despierta un gran interés por investigar acerca de la posibilidad de explotar el tiempo en el que la CPU de los servidores de Internet permanecen ociosos, y de las posibilidades de aprovechamiento para el procesamiento paralelo.

En este trabajo se intenta aplicar la experiencia acumulada a través de las diversas investigaciones para analizar el aprovechamiento de los ciclos ociosos de CPU de un conjunto de servidores de Web, servidores de mail y otros que puedan ser alcanzados por medios de las tecnologías de Internet, para estimar la oportunidad de utilizar la potencia de cálculo desaprovechada en esas máquinas, estimar la frecuencia en que el cálculo debería detenerse para adaptarse a los cambios de la disponibilidad de procesadores y estimar el beneficio que un usuario puede esperar en cuanto a la potencia de cálculo alcanzable.

Luego con estas estimaciones puede definirse el perfil de los sistemas que mejor pueden aprovechar el tiempo ocioso de tales máquinas.

3. Nivel de Utilización de CPU

Con el objeto de determinar la potencia de cálculo desaprovechada por los servidores de Internet, se registró durante las 24 hs. de un día normal, la traza de porcentaje de utilización de CPU de un conjunto de servidores entre los cuales se encuentran servidores de web y servidores de mail o ambas cosas, algunos de ellos corriendo sobre Linux y otros corriendo sobre Windows NT.

Las trazas están compuestas por alrededor de 86400 valores entre 0 y 100, cada uno de los cuales corresponde al porcentaje de utilización de la CPU medido en cada segundo.

Para el análisis posterior se tomaron las dos trazas de utilización de CPU más representativas de las obtenidas, de acuerdo a su forma general. Una de ellas corresponde a un servidor de Web corriendo sobre Windows NT y la otra a un servidor de Web y mail corriendo sobre Linux.

Los valores de tales trazas fueron agrupados de a 600 y luego promediados de forma tal de posibilitar su representación por medio de una curva que permita apreciar sus características básicas.

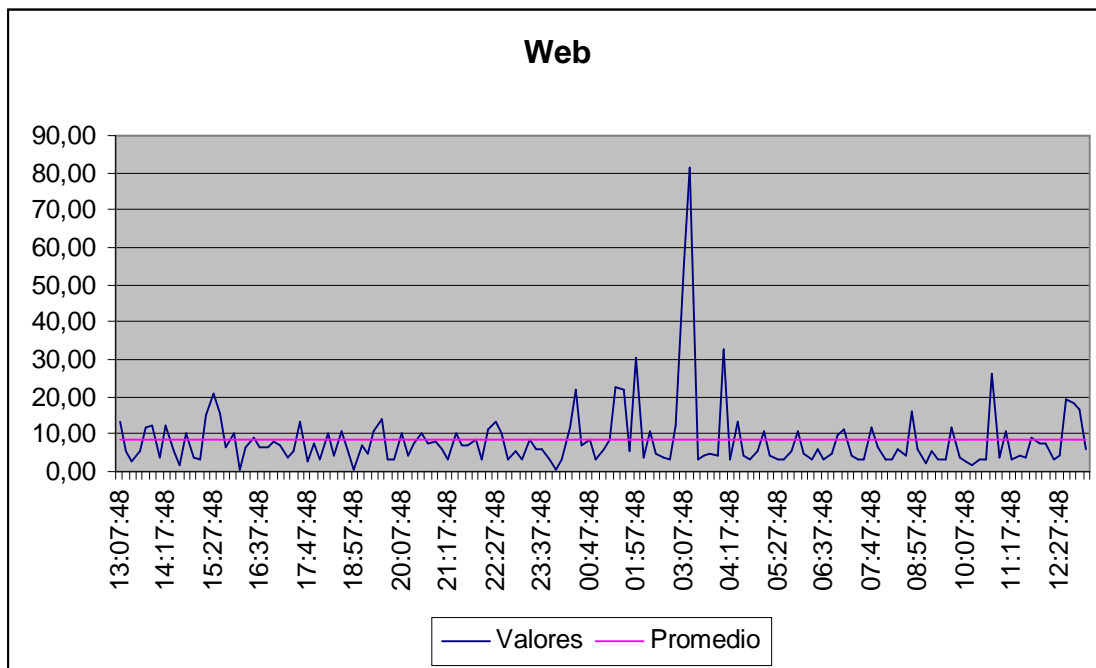


Figura 1: Servidor Web promediado

Como puede verse en las curvas representadas en las figuras 1 y 2, el porcentaje de utilización de CPU es, en términos generales, muy bajo. Pero además pueden obtenerse otros datos más significativos. Más allá de los picos de utilización más pronunciados ocasionados por los sistemas automáticos de backup, o aquellos esporádicos, no tan pronunciados, que no tienen una explicación más que el mismo comportamiento errático de los usuarios de Internet, o el ruido constante que introduce el registro mismo de la traza, se diferencia claramente el nivel de utilización en los horarios diurnos y en los horarios nocturnos.

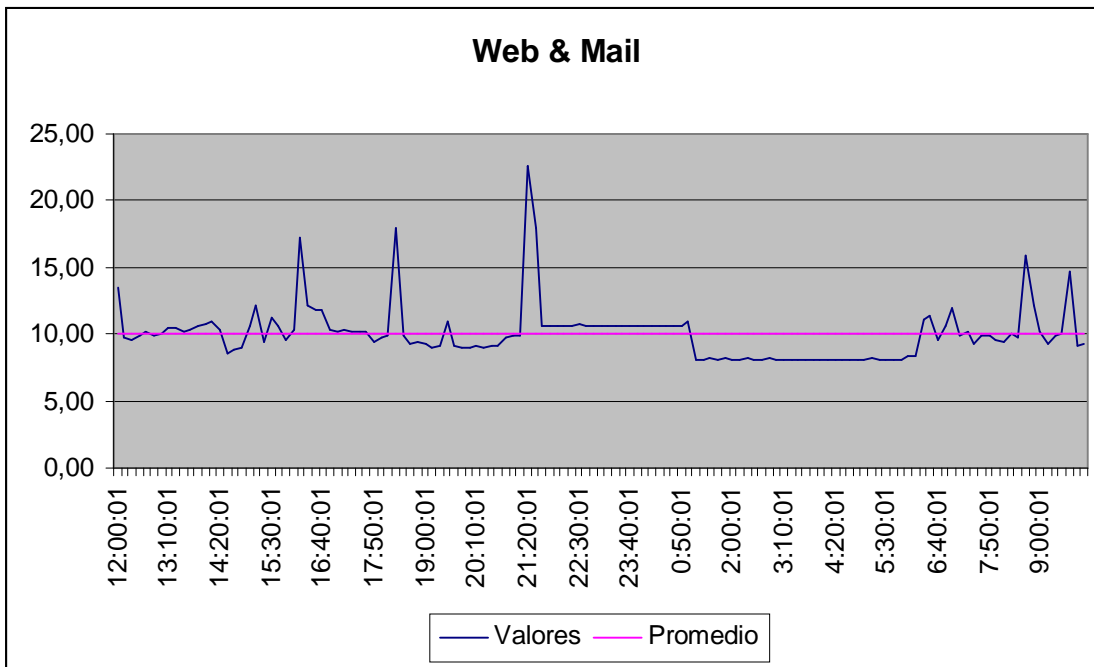


Figura 2: Servidor Web & Mail promediado

En las siguientes figuras se puede ver la suma de los tiempos en los que la utilización de la CPU se encuentra en los rangos de 0 a 5%, de 5 a 10%, y así sucesivamente.

En el caso del servidor de Web el 78% del tiempo total permanece por debajo del 5% de utilización de CPU, un 89% del tiempo por debajo del 10% de utilización y un 97% del tiempo por debajo del 20% de utilización.

El servidor de Web y mails permanece por debajo del 20% de utilización durante el 98% del tiempo total, por debajo del 10% durante el 74% del tiempo, pero por encima del 5% casi todo el tiempo medido.

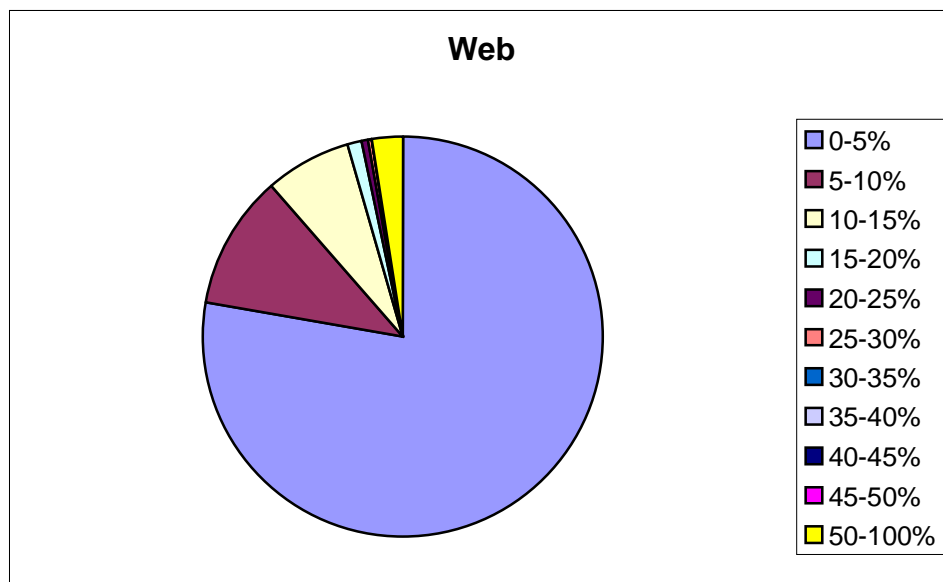


Figura 3: Web - Utilización por rangos

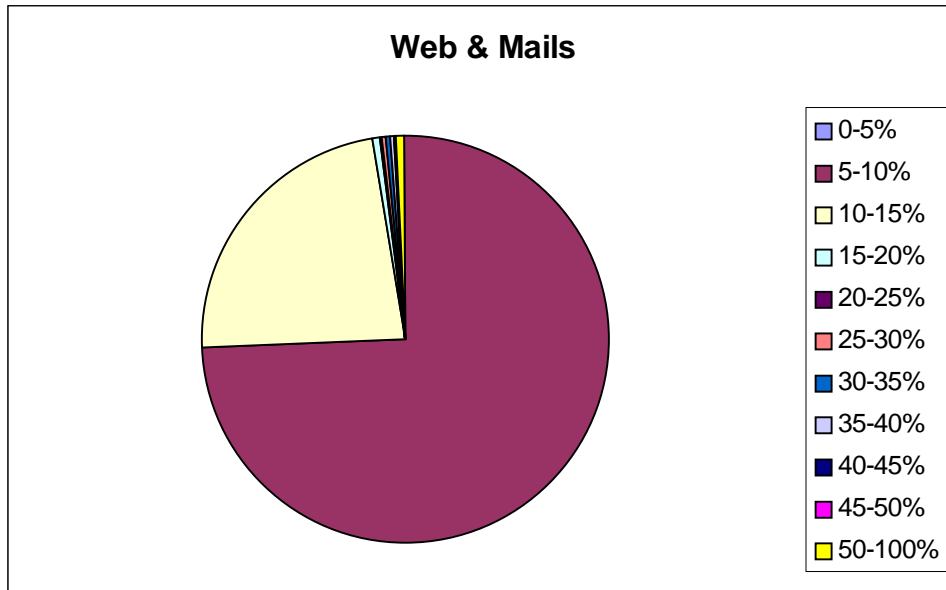


Figura 4: Web & Mail - Utilización por rangos

En promedio, los porcentajes de utilización de los dos servidores tomados como muestras representativas son 8,35% para el servidor de Web y 10,06% para el servidor de Web y mails. Considerando que se trata de computadoras que pueden lograr una potencia de cálculo que oscila alrededor de los 650 MFlops, se concluye que queda una potencia de cálculo de entre 584 y 595 MFlops sin utilizar.

4. Distribución del tiempo ocioso

Para estimar la frecuencia en la que el cálculo debería detenerse para adaptarse a los cambios de la disponibilidad de procesadores se calcula la distribución del tiempo en que la utilización de la CPU de una máquina permanece dentro de un rango.

El conocimiento de esta distribución de disponibilidad permite determinar el tipo de cálculos que pueden hacer uso de la potencia de cálculo no utilizada por la función normal de los servidores de Internet, y por ende determinar el tipo de aplicaciones que se desarrollen a tal fin.

Se pueden encontrar varios de trabajos en los que se han realizado estudios similares pero sobre las estaciones de trabajo de una red [10][6]. En general los resultados obtenidos en ellos muestran que las máquinas que han estado ociosas por un breve período de tiempo son más comúnmente requeridas que aquellas que han permanecido ociosas por un período relativamente largo. Se podría encontrar en esto un criterio análogo al del principio de localidad.

En este trabajo se utilizaron como indicadores lo que se llamó ráfagas máximas y ráfagas promedio. Se llama ráfaga a cada uno de los períodos de tiempo en los cuales la utilización de la CPU permanece por debajo de un cierto porcentaje. La ráfaga máxima para un determinado porcentaje es la ráfaga de mayor extensión en el tiempo para la cual la utilización de la CPU permanece por debajo de dicho porcentaje. La ráfaga promedio para un determinado porcentaje corresponde al tiempo promedio de

duración de todas aquellas ráfagas para las cuales la utilización de la CPU permanece por debajo de dicho porcentaje.

Los siguientes gráficos muestran las curvas formadas por las duraciones en segundos de las ráfagas máximas y promedio para ambas trazas en forma acumulativa. Es decir, las ráfagas máximas y promedio por debajo del 5% de utilización de CPU, por debajo del 10%, del 15% y del 20%.

Tanto en el gráfico que ilustra las ráfagas máximas como el que ilustra las ráfagas promedio muestran una importante desigualdad en el funcionamiento de los servidores.

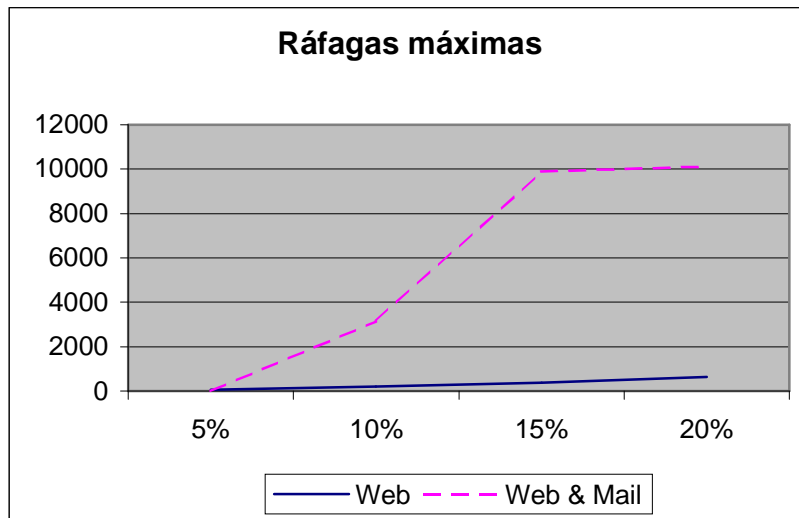


Figura 5: Ráfagas máximas comparadas (seg.)

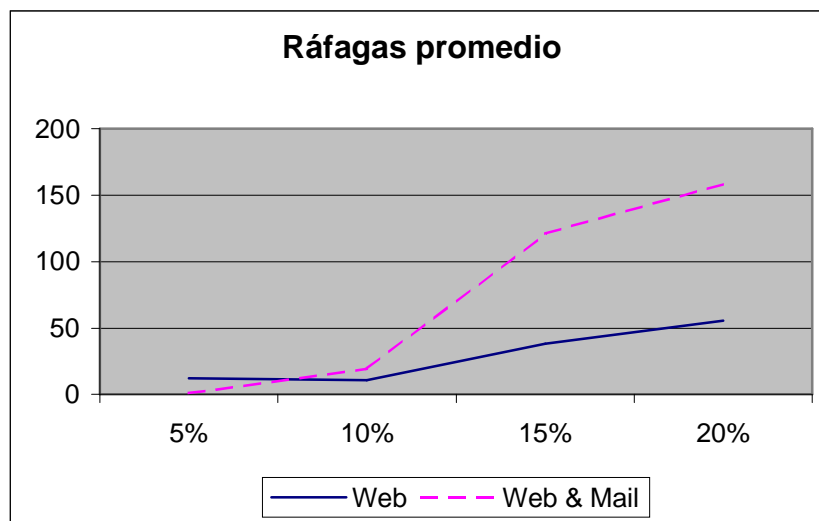


Figura 6: Ráfagas promedio comparadas (seg.)

No se descubre nada nuevo al ver en las figuras 5 y 6 que los requerimientos a los servidores Web son satisfechos en mucho menos tiempo que los requerimientos a los servidores de mail. Sin embargo es importante conocer estos valores para un aprovechamiento correcto y eficiente.

En otras palabras, es posible aprovechar la potencia de cálculo desaprovechada por ambos tipos de servidores, aunque en el caso de los servidores Web se amplía el abanico de las aplicaciones posibles y adecuadas, debido a que las ráfagas poseen en general una mayor duración.

5. Conclusiones y trabajos futuros

Si bien este trabajo constituye apenas la punta de un iceberg, es un paso necesario hacia la utilización de la potencia de cálculo que desaprovechan los servidores de Internet.

Los datos obtenidos y analizados ponen en evidencia un potencial factible de utilizar. Esta utilización podrá ser concretada por medio del desarrollo de sistemas distribuidos que requieran de breves ráfagas de ejecución que puedan ser altamente paralelizadas, es decir con baja o ninguna dependencia de datos. Puede contribuir en gran medida la tolerancia a fallas como forma de adaptación a los cambios de disponibilidad.

Un buen ejemplo de este tipo de aplicaciones puede ser un sistema de búsqueda de números primos que envíe requerimientos de cálculo a distintos servidores. En este sentido se dirigen los trabajos futuros.

Se pretende desarrollar un sistema compuesto por un servidor de requerimientos y un conjunto de clientes (que a su vez son servidores de Internet). En cada uno de estos clientes correrá en background un proceso que medirá permanentemente la utilización de la CPU. Cuando dicho proceso detecte que la utilización de la CPU esté por debajo de un parámetro configurable por el administrador del equipo, pedirá al servidor de requerimientos un conjunto de datos a procesar. Finalizado ese procesamiento, el cliente devolverá los resultados al servidor.

Si por cualquier motivo el nivel de utilización de la CPU supera al valor del parámetro mencionado, se abortará automáticamente el procesamiento actual de forma tal de no interferir en las necesidades de servicio de Internet. Después de transcurrido cierto tiempo, el servidor de requerimientos da por perdido ese trabajo y envía los mismos datos al próximo cliente que los requiera.

6. Referencias

[1]	Shoch J.F., Hupp J.A. The “Worm” programs—Early Experience with a Distributed Computation. <i>Comm. ACM</i> 25,3. 1982
[2]	Hagmann R. Process Server: Sharing Processing Power in a Workstation Environment. <i>Proceedings of the Sixth International Conference on Distributed Computing Systems</i> . 1986
[3]	Litzkow M. Remote Unix: Turning Idle Workstations into Cycle Servers. <i>Proceeding of the Summer 1987 Usenix Conference</i> . 1987
[4]	Theimer M. Preemptable Remote Execution Facilities for Loosely-Coupled Distributed Systems. Ph.D. Th., Stanford University. 1986
[5]	Nichols D.A. Using Idle Workstations in a Shared Computing Environment. <i>Proceedings of the 11'th ACM Symposium on Operating Systems Principles</i> . 1987
[6]	Douglis F., Ousterhout J. Transparent Process Migration: Design Alternatives and the Sprite Implementation. <i>Software Practice and Experiences</i> . 1991
[7]	Mutka M., Livny M. The Available Capacity of a Privately Owned Workstation Environment. <i>Performance Evaluation</i> . 1991
[8]	Theimer M., Lantz K. Finding Idle Machines in a Workstation-based Distributed System. <i>IEEE Transaction on Software Engineering</i> . 1989
[9]	http://www.seti.org/
[10]	Arpaci R.H., Dusseau A.D., Vahdat A.M., Liu L.T., Anderson T.E., Patterson D.A. The Interaction of Parallel and Sequential Workloads on a Network of Workstation. <i>Proceedings of the 1995 ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems</i> . 1995.