

ALGORITMOS DE COMUNICACION EN HIPERCUBOS DE BASE B

CARMEN B. MORA

ALEJANDRO TERUEL

Departamento de Computación y Tecnología de la Información

Universidad Simón Bolívar

Apartado 8900, Caracas 1080-A, Venezuela

RESUMEN

Una topología de interconexión de multiprocesadores ampliamente usada por su relación costo-efectividad es la de hipercubos binarios. Una generalización de la estructura de hipercubos binarios es la de hipercubos de base b . En este trabajo se presenta el desarrollo de algoritmos de comunicación para hipercubos de base b . Los algoritmos obtenidos presentan resultados favorables en cuanto a tiempos de comunicación a medida en que aumenta la base. La tecnología disponible en la actualidad limita en cierta medida la implementación física de las estructuras generalizadas dados los costos mayores en nodos de alto grado, sin embargo, no hay que descartar que los avances tecnológicos definan la factibilidad de implementación a futuro permitiendo aprovechar las ventajas de comunicación.

1-INTRODUCCION

La comunicación entre procesos es frecuentemente uno de los principales obstáculos para incrementar el rendimiento de los algoritmos paralelos en multiprocesadores. Saad y Shultz presentan en [5] y [6] algunos algoritmos para intercambio de datos entre procesadores en distintas topologías, destacándose los hipercubos binarios. Las redes de procesadores basadas en hipercubos son atractivas debido a sus propiedades de interconexión entre procesadores.

Lakshmirarahan y Dhall, presentan en [3] una generalización de la topología de hipercubos, llamada hipercubos de base b , los cuales sugieren son más tolerantes a fallas en la medida en que aumenta la base. Esto se fundamenta en el hecho de que hipercubos con mayor base presentan mayor cantidad de caminos paralelos. Otra propiedad importante, es la reducción del diámetro (distancia mínima entre dos nodos

más lejanos) en estas estructuras a medida en que aumenta la base, lo cual sugiere la posibilidad de incrementar la velocidad en la comunicación entre nodos con respecto a los hipercubos binarios. Esto es debido a que los tiempos de comunicación en muchos de los algoritmos sobre hipercubos son proporcionales al diámetro de la estructura.

Los objetivos de este trabajo han sido desarrollar para base genérica b , algunos algoritmos de comunicación existentes para hipercubos binarios, así como algoritmos que surgen como necesidades en la práctica, y analizar el rendimiento de hipercubos de mayor base.

2- LAS ESTRUCTURAS DE HIPERCUBOS

Las definiciones y propiedades que se presentan en esta sección se basan en los trabajos [7] y [3] donde se introducen las estructuras de hipercubos binarios e hipercubos en base b respectivamente. Dadas las restricciones de espacio, se supondrán conocidas las propiedades de los hipercubos binarios.

Extenderemos la definición de distancia de Hamming para considerar las etiquetas en base b , donde $b > 1$. La representación en base b de cualquier entero $0 \leq X < k$ puede ser escrita como $X_1X_2..X_N$, $X_i < b$, donde $\langle b \rangle = \{0,1,2,..,b-1\}$. La distancia de Hamming en base b , denotada $H_b(A,B)$, entre dos nodos A,B es el número de dígitos en base b en que difieren A y B .

Hipercubo de base b . Un hipercubo de base b , y dimensión N , consiste de $k=b^N$ nodos, $2 \leq b \leq k$, donde dos nodos están conectados si y sólo si sus etiquetas en base b difieren en uno y sólo un dígito. Esta definición corresponde a la de hipercubo completo de dimensión N y base b , presentada en [3]. La Figura 1 presenta un ejemplo de hipercubo de 9 nodos, base 3 y dimensión 2.

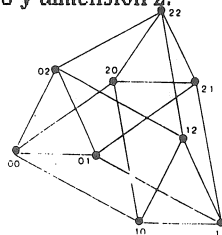


Figura 1. Ejemplo de hipercubo de base 3, $k=9$, $N=2$

3- ALGORITMOS DE COMUNICACION

Los algoritmos que fueron desarrollados para base genérica b , están basados en las presentaciones hechas para el caso de hipercubos binarios en [1], [2] y [5]. Los algoritmos seleccionados para su generalización a base b son: Envío de datos entre dos nodos, Divulgación (Broadcast), Dispersión (Scattering), Recolección (Gathering), Intercambio total, Multienvíos (Multicasting) y Transposición de datos. A continuación se presenta una breve descripción de cada problema de comunicación y seguidamente una selección de algoritmos que representan mejoras sustanciales en tiempos de comunicación con respecto al caso binario. Los algoritmos restantes se encuentran desarrollados en [4].

ENVIO DE DATOS ENTRE DOS NODOS

El envío de datos desde un procesador a otro representa la operación de transferencia de datos más simple entre los algoritmos de comunicación propuestos. La transferencia de datos se lleva a cabo a través del pase de mensajes entre nodos vecinos.

DIVULGACION

El problema de divulgación se refiere a la transferencia de datos desde un nodo en particular a todos los otros nodos del hipercubo.

RECOLECCION

La recolección o "Gathering", se refiere a la transferencia de la misma cantidad de datos desde todos los nodos a un nodo en particular. En este caso cada nodo envía un paquete de datos distinto al nodo receptor.

DISPERSION

El problema de dispersión o "scattering", consiste en la transferencia de datos distintos desde un nodo en particular a todos los otros nodos del hipercubo. A diferencia del problema de divulgación, los datos que recibe cada nodo son distintos, por lo que hay que controlar en el algoritmo, la distribución correcta de los datos a los nodos correspondientes.

INTERCAMBIO TOTAL

El intercambio total de datos se refiere al envío de paquetes iguales desde cada nodo hacia todos los otros nodos. Esta operación puede verse como multidivulgación puesto que cada nodo necesita hacer divulgación de su dato a todos los otros nodos.

MULTIENVIOS

La forma de comunicación de datos por multienvios, se refiere al tipo de comunicación entre procesos en el cual un nodo fuente desea enviar un mismo mensaje a m distintos nodos destinos. Esto constituye una generalización con respecto a la comunicación referida al envío de datos entre dos nodos ($m=1$) y a la de divulgación ($m=2^{N-1}$).

TRANSPOSICION DE DATOS

El problema de transposición de datos se refiere al envío de datos distintos desde cada procesador hacia todos los otros procesadores. Esto puede ser considerado como una generalización del problema de dispersión o en forma análoga, del problema de recolección.

RESULTADOS

A efectos del diseño y análisis de rendimiento de los distintos algoritmos de comunicación desarrollados, se consideró el siguiente modelo:

- Todos los paquetes tienen igual longitud
- El tiempo para transmitir un paquete desde un nodo cualquiera a un nodo vecino es $\beta + \mu$ donde:
 - β = Latencia o tiempo de inicio de comunicación
 - μ = Tiempo que tarda un paquete en atravesar un canal de comunicación.
- El espacio de almacenamiento de cada nodo es infinito. Esto implica que no se consideran retardos ocasionados por insuficiencia de espacio de almacenamiento en nodos receptores.
- La transmisión de un paquete desde un nodo a cualquier número de nodos vecinos se efectúa en la misma cantidad de tiempo.
- Todos los canales de comunicación incidentes a un nodo pueden ser usados simultáneamente para transmisión y recepción de paquetes (MLA).

- Dado un número k de nodos, las estructuras de hipercubos a comparar son de la forma:

- 1) Base 2, dimensión $n = \log_2 k$, ($k=2^n$)
- 2) Base b , dimensión $N = \log_b k$, ($k=b^N$)

Los distintos algoritmos pueden ser agrupados según cuatro técnicas distintas que representan la forma en que fue atacado cada problema. En la Tabla 1 se presentan los resultados obtenidos en cuanto a tiempos de ejecución, en forma comparativa para base 2 y base b .

ALGORITMO	TECNICA	BINARIO	BASE b
Envío entre dos nodos Divulgación Recolección Dispersión	Envío Simultáneo	$n(\beta+\mu)$ $n(\beta+\mu)$ $n\beta+(k/2)\mu$	$N(\beta+\mu)$ $N(\beta+\mu)$ $N\beta+(k/b)\mu$
Divulgación Recolección Dispersión Transposición	Recursión	$n(\beta+\mu)$ $n\beta+(k-1)\mu$ $n\beta+(k-1)\mu$ $(k-1)$	$N(\beta+\mu)$ $N\beta+(k-1/b-1)\mu$ $N\beta+(k-1/b-1)\mu$ $(k-1)/(b-1) - 1$
Intercambio Total Transposición	Partición	$n\beta+(k-1)\mu$ $n(\beta+(k/2)\mu)$	$N\beta+(k-1/b-1)\mu$ $N(\beta+(k/b)\mu)$
Multienvíos	Dimensión Común	$n(\beta+\mu)$	$N(\beta+\mu)$

Tabla 1. Algoritmos desarrollados/resultados

La técnica de envío simultáneo para la construcción de algoritmos de comunicación se refiere a la idea de transmitir cada paquete de datos haciendo uso de la mayor cantidad de canales de comunicación disponibles y enviar en forma simultánea por cada uno de ellos.

La técnica de recursión se refiere a la construcción del algoritmo siguiendo los caminos de comunicación que resultan de la construcción recursiva de hipercubos a partir de hipercubos de dimensión menor.

La construcción de algoritmos de comunicación según la técnica de partición, se refiere a utilizar en cada paso los canales de comunicación que definen la partición del hipercubo según alguna dimensión.

La técnica de dimensión común busca enrutar los paquetes de datos por los enlaces de comunicación correspondientes a la dimensión del hipercubo donde hay mayor cantidad de nodos receptores.

A título de ejemplo, se presenta a continuación uno de los algoritmos desarrollados según la técnica de partición.

INTERCAMBIO TOTAL (Técnica: Partición)

El algoritmo de intercambio según dimensiones presentado en [5] para el caso binario como ADEA (Alternate Direction Exchange Algorithm), consiste en dividir en cada paso, según una dirección distinta, el cubo de dimensión n en dos cubos de dimensión $n-1$ y realizar el intercambio de datos entre los nodos opuestos correspondientes de los dos cubos de dimensión menor. La correspondencia entre nodos referida como "nodos opuestos" según la dirección i -ésima se define de la siguiente manera: Dos nodos vecinos A y B son opuestos según la dirección i , si las etiquetas A y B difieren en la posición i -ésima. Sin pérdida de generalidad, se toma como sentido de dirección desde el bit menos significativo al más significativo (derecha a izquierda). La Figura 2 presenta un esquema de como se realiza el intercambio para el caso de un hipercubo binario de dimensión 3.

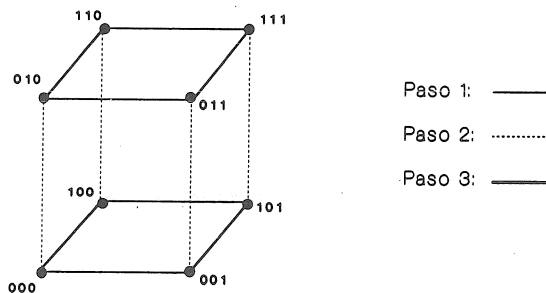


Figura 2. Intercambio según ADEA en hipercubo de dimensión 3.

Para el algoritmo correspondiente a base b , debe considerarse que ahora cada nodo tiene $b-1$ vecinos opuestos según cada dirección. Esto se debe a que en base b , cada dígito puede variar según b posibles valores.

ADEA en hipercubos base b

Sea $N = \log_b k$

For $i=1$ to N do

Cada nodo intercambia sus b^{i-1} datos acumulados con los $b-1$ nodos opuestos en la i -ésima dirección

Cabe observar que el intercambio del lazo interno no es un intercambio entre dos nodos sino entre b nodos. Sin embargo, dado que estos b nodos son vecinos entre sí (todos difieren en la misma posición i), y haciendo uso de la suposición MLA, este intercambio entre b nodos puede realizarse en 1 sólo paso. Bajo esta consideración el ciclo interno tendrá entonces una duración asociada a la cantidad de datos a intercambiar y no a la cantidad de nodos que intervienen en el intercambio.

CONCLUSIONES

El aporte fundamental del presente trabajo ha sido el desarrollo de las rutinas básicas de comunicación para hipercubos de base genérica b , basado en los algoritmos existentes para hipercubos binarios. Dichos algoritmos han sido diseñados para ser usados sin cambios con cualquier valor de base $b \geq 2$, aprovechando la disponibilidad de canales de comunicación adicionales según aumenta el valor de la base y logrando en consecuencia mejoras en los tiempos de ejecución.

Como resultados del desarrollo de algoritmos para base genérica b , se pudo comprobar la hipótesis de incidencia favorable de dimensiones menores en los tiempos de comunicación entre los procesadores.

Los algoritmos desarrollados constituyen un punto de partida para la construcción de una librería de subrutinas que realicen las tareas de comunicación más comunes en sistemas de multiprocesadores. Aunque hasta el momento no se conoce la existencia de estas arquitecturas, queda pues para futuros trabajos la implementación y prueba de

dichas rutinas en un sistema simulado. Todo parece indicar que la tecnología disponible limite la conveniencia de altos grados en los nodos, y en consecuencia las ventajas de una mayor base. La tendencia es una vuelta hacia estructuras más sencillas. La disponibilidad en la actualidad de conmutadores de reconfiguración que permiten modificar la topología de interconexión en forma dinámica, permitiría decidir según la aplicación, el tipo de configuración que más se adecúe a la naturaleza del problema. No hay que descartar sin embargo que los avances tecnológicos en los próximos años permitan disponer a bajo costo de nodos de alto grado de comunicación, y es de esperarse la oportunidad para las estructuras de hipercubos generalizadas.

REFERENCIAS

- [1] D. P. Bertsekas, C. Ozveren, G.D. Stamoulis, F. Tseng y J. N. Tsitsiklis, "Optimal Communication Algorithms for Hypercubes", *Journal of Parallel and Distributed Computing*, 11, p. 263-275, (1991).
- [2] A. H. Esfahanian y M. L. Ni, "Multicast in Hypercube Multiprocessors", *Journal of Parallel and Distributed Computing*, 8, p.30-41, (May 1988).
- [3] S. Lakshminarayanan y S. K. Dhall, "A New Hierarchy of Hypercube Interconnection Schemes for Parallel Computers", *The Journal of Supercomputing*, 2, p. 81-108, (1988).
- [4] C. Mora , "Algoritmos de Comunicación en Hipercubos de Base b", Tesis de Maestría en Ciencias de la Computación", Universidad Simón Bolívar, (Enero 1992).
- [5] Y. Saad y M. H. Schulz, "Data Communication in Hypercubes", *Journal of Parallel and Distributed Computing*, 6, p. 115-135, (1989).
- [6] Y. Saad y M. H. Schulz, "Data Communication in Parallel Architectures", *Research Report YALEU/DSC/RR-461*, Yale University, (March 1986).
- [7] Y. Saad y M. H. Schulz, "Topological Properties of Hypercubes", *Research Report YALEU/DSC/RR-389*, Yale University, (June 1985).