7H－3

スーパーデータベースコンピュータ SDC2 における データネットワークの動作解析

北村 学　田村 孝之　中村 稔　喜連川 優　高木 幹雄
東京大学生産技術研究所

1 はじめに

「スーパーデータベースコンピュータ SDC2」は、現在我々が開発中の高並列関係データベースサーバである。並列計算機のシステムの動作は複雑で、それを正確に把握することは難しいが、効率の良い動作をさせるためには、その性能を定量的に評価することが必要になる。モジュール内の動作を調べるツールに関しては、すでに研究がなされている [1]。

SDC2 にはネットワークのハードウェア自体にモジュール間の負荷を分散する「バケット平坦化機能」が存在する。この機能は、ネットワークの各スイッチ素子が自律的に、適応的なルーティングを行うことによって実現される。今回、このネットワークの動作をモニタリングする環境を作成したので、それについて報告する。

2 データネットワークの構成

2.1 ネットワーク概要

データネットワークの概要は以下のようなものになっている。各スイッチングユニットは、2×2 の

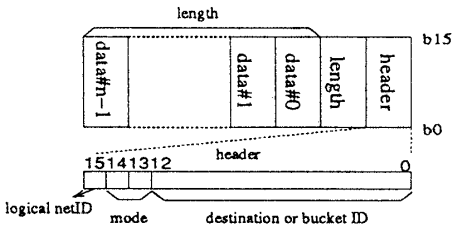

図 1: タプルのフォーマット

クロスバースイッチからなり、クロックごとに各段を入力側から出力側へとパイプライン式にデータを送信する。また、ネットワークに出力するタプルの先頭には、図 1のように接続モード・ID を示すヘッダとタプル長が付加してあり、それによって各スイッチは自律的に経路選択を行える。

また、ネットワーク上にネットワークマネージメントプロセッサが置かれ、コントロールネットワークを通じての処理モジュールとの通信、ネットワークの初期設定などの処理を行っている。

2.2 スイッチングユニットの動作

各スイッチングユニットの動作は以下のようになる。

1. ヘッダがスイッチに入力される (HEAD 信号 ON)
2. ヘッダをレジスタにラッチ (BUSY 信号 ON)
3. 宛先指定モードの場合、宛先アドレスから接続状態決定 (ESTAB 信号 ON)
4. 転送開始 (BUSY 信号 OFF)

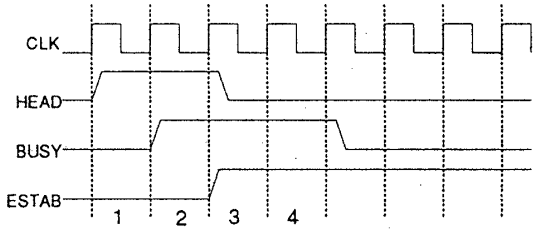

図 2: スイッチングユニットのタイミングチャート

平坦化モードの場合は、バケット ID、バケット長によって接続状態を決定するために 4 の前に数サイクル要する。さらにヘッダを次段に送り、次段から BUSY 信号が到着すると、再び BUSY になり、ブロックが起こることになる。

3 モニタリングシステム

ネットワークから得られる情報で、主に必要なものは、

Behavior Analysis of the Data Network in the Super Database Computer(SDC2)
M.Kitamura,T.Tamura,M.Nakamura,M.Kitsuregawa
Institute of Industrial Science, University of Tokyo

・各スイッチのトラフィック (データの通過している時間)
・各スイッチの輻輳率
である。

これらはスイッチングユニットで、BUSY, ESTAB, HEAD の状態にある時間 B, E, H を用いて計算される。そのため、各スイッチングユニットで B, E, H を計測することが必要になる。

今回のシステムで、各スイッチングユニットの状態をモニタリングする場合には次のような方法が考えられる。

1. マネージメントプロセッサが、各スイッチの状態レジスタを巡回的に取得して、マネージメントプロセッサ側のカウンタに記録
2. 各スイッチにカウンタを用意し、各状態の発生した回数を記憶させておき、マネージメントプロセッサが巡回的に取得する

2 は必要なハードウェアが増加する。しかし、1 の場合はクロックごとの状態変化を記憶できないため、全てのスイッチの状態取得を行う際、はじめに 1 つのスイッチの状態を取得してから再びこのスイッチの状態を取得するまでに大きな遅延が入る。

実際に 1 の方法で行うと、1 つのスイッチの状態を取得して、別のスイッチの状態を取得するまでにかかる時間が約 1.6μsec、12 個のスイッチの状態を取得するまでに要する時間が 17.9μsec かかった。100 バイトデータを平坦化モードで転送した場合、最小の遅延は 5.6μsec のため、この時間遅延は無視できないほど大きい。

そのため、ハードウェアの要素を増やすことになるが、2 の方法が良いということになる。今回は 2 の方法で実装を行った。

4 構成

今回試作したモニタリングソフトウェアは、ネットワークのマネージメントプロセッサで動作するサンプラーと、ワークステーション上で動作するコレクター・ビジュアライザーの 3 つで構成される。

- サンプラー：マネージメントプロセッサから各スイッチングユニットのカウンタ (HEAD, BUSY, ESTAB) レジスタを読む。カウンタを読むと各状態レジスタは 0 にクリアされる。また、コントロールネットワークを通じ、ワークステーションに状態カウンタの値を送信する通信機能も持つ。
- コレクター：ネットワーク側から送信されてきたデータを受信し、計数する。
- ビジュアライザー：各スイッチの状況を GUI でリアルタイムに表示

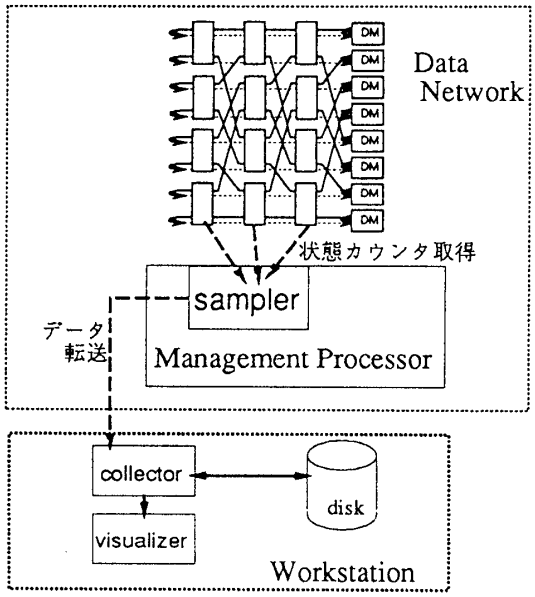

図 3: ネットワークモニタ構成図

5 まとめ

以上のように、SDC2 のデータネットワークのモニタリングシステムを作成した。今後、関係演算処理にネットワークを使用したときの状況を測定し、解析する予定である。

参考文献

[1] 鈴木他: “スーパーデータベースコンピュータ (SDC) における性能評価ツール”, 情報処理学会第 43 回 (平成 3 年後期) 全国大会.

[2] 田村他: “スーパーデータベースコンピュータ (SDC2) におけるデータネットワーク系の実装”, コンピュータシステム研究会, 電子情報通信学会, 1993 年 8 月.