Access Grid の評価実験
VIA を用いた MVICH の性能評価

About Access Grid
An evalution of the MVICH using VIA

田村 隆一，小池 政輝
Ryuichi TAMURA，Masaki KOIKE

Abstract: This paper describes 2 themes; Access Grid,VIA. First we give an acount of AG; a collaboration tool which eneble to communicate between remote places. And next VIA. It is an architecture that realizes high performance communication by means of cuting the overhead caused by intervention of OS kernel. We evaluate the performance of MVICH using VIA.

1 Access Grid

Access Grid [1](以下 AG) は，物理的に離れた空間におけるコラボレーションを可能にするためにデザインされたツールであり，情報の交換やコミュニケーションを行なう新たな媒体である． Grid のミドルウェアを用いることにより，Grid 間のコミュニケーションをサポートし，遠隔コラボレーションを実現する Compelling Interactive System を作ることができる．

AG は，従来よりも大規模な TV 会議，共同作業，セミナー，講義そしてチュートリアルを行なうことが出来る．つまり個人対個人の空間ではなく，グループ対グループの共有空間をネットワーク上に構築する事ができ，自然かつシームレスな face-to-face のコミュニケーションを実現することが可能である．

1.1 Access Grid の仕組み

AG の各ノードは，Display System，Video System，Audio System そして Network の４種類のコンポーネントから構成される．

各部屋の設計を Fig. 1 に示す．

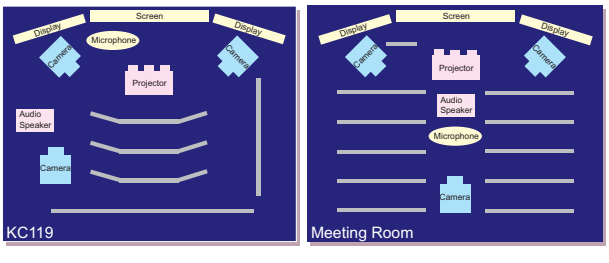

Fig. 1　AG Node Room and Equipment Placement

Video System は Video Capture Machine に複数台のカメラを接続し，発表者，聴衆，スクリーンの映像をキャプチャする．

Audio System は Video System での視覚的な不十分さを補うため，Audio Machine にマイクやスピーカを接続し聴覚的な役割を担う．

Display System は，Display Machine からプロジェクタ，スクリーン，プラズマディスプレイ等に接続され Video System でキャプチャされた映像を映し出す．

1.2 Access Grid に必要なハードウェア

まずコンピュータ４台（Display Machine，Video Capture Machine，Audio Machine，Control Machine）が必要である．この時４台のマシンを１組のディスプレイ，キーボード，マウスで操作するためのデュアル・コンソールの KVM スイッチがあれば便利である．各マシンのスペックは 100Mbps Ethernet adapter，256MB のメモリ，9GB のハードディスクの容量が最低基準となる．

また各マシンの環境は，Display Machine では OS として Windows2000 を用い，複数の出力を行なうことができるグラフィックカードが必要である．同様に Video Capture Machine，Audio Machine は OS として RedHat Linux[2]，必要デバイスとして Video Capture Machine ではカメラ台数分のビデオキャプチャカード，Audio Machine ではサウンドカードを使用する．そして Control Machine では OS に Windows98 もしくは WindowsME を使用する．

４台のマシン及び各メディア機器の接続状況を Fig. 2 に示す．

1.3 Access Grid に必要なソフトウェア

米国 Chicago 大学 Argonne National Laboratory の MCS Division が開発したソフトウェア[3]が一般的に用いられる．AG の各ノードにそれぞれ異なったソフトウェ

[1] http://www.accessgrid.org

[2] 現バージョンは 6.2 ベースで AG 仕様である．

[3] ttp://www-fp.mcs.anl.gov/fl/accessgrid/agpackaging.htm でダウンロードが可能．

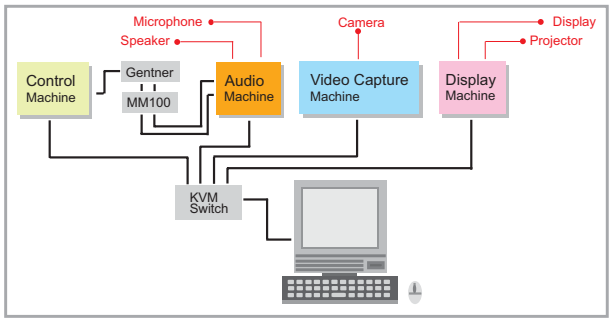

Fig. 2　AG Hard Ware Configurations

アがインストールされる．

Display Machine で用いるソフトウェアは以下の３種類である．

◆ event server

event server は resource manager(arm，vrm) に Virtual Venue[4]情報を供給する．event server は AG ノードにとって重要なソフトウェアであり，VV server[5]や resource managers を相互接続する．

◆ Distributed PowerPoint

Distributed PowerPoint（以下 DPPT）は client，master，server(registry，agserv) の３つのコンポーネントから成る．DPPT では，VV 上にいる参加者が主催者の PowerPoint をダウンロードし，ある主催者側が始めに master を実行し，全ての参加者がその後 client を実行する．server が master と client を同期させ，リアルタイムにプレゼンテーションが行なえる仕組みになっている．また client は Display Machine で実行されるので，PowerPoint は Display Machine に接続されたスクリーンやプラズマディスプレイに映し出される．

◆ moo

moo は基本的にはテキストベースのチャットツールであり，異なった virtual room（VV の各部屋）にいても利用できる．同じ VV 上にいるならば，どの参加者がどの部屋にいるのかが分かったり，過去のログの内容を読み返したり，ネットワークさえつながっていれば Video Capture Machine や Audio Machine にトラブルが起きても moo を利用し，会話できる．

また Video Capture Machine では video resource manager（以下 vrm）を Audio Machine では audio resource manager（以下 arm）をソフトウェアとして用いる．どちらも Display Machine の event server からの Virtual Venue 情報を検知し，vrm は映像を arm は音声を入力する．

[4]Virtual Venue は物理的には実在しない仮想会議室のことである．複数の同時発生的な会議などをサポートするメカニズムを持ち，AG における出入り口となっている．

[5]現在 ANL が管理しており予約を入れることによって使用可能である．

2　VIA を用いた並列処理の高速化

2.1　はじめに

クラスタシステムの高速化の要求に伴ってクラスタノード間の相互接続を行う System Area Network(SAN) がネットワーク研究分野において注目されている．しかし，最も広く利用されているネットワークプロトコルである TCP/IP は遅延が大きく，高速な通信を行う際のボトルネックとなってしまう．そこで近年，高速なプロセス間通信を行うために設計された新しい標準規格 VIA が提唱された．本報告では，この VIA についての説明を行い，これを導入したクラスタの性能を評価，検討する．

2.2　VIA とは

VIA とは Virtual Interface Architecture の略であり，クラスタを用いた並列処理を行う際のプロセス間通信の高速化を実現することを目的に，Intel 社，Microsoft 社，COMPAQ 社が共同で発表したネットワークアダプタのアーキテクチャ仕様である．

VIA は主に以下の構成要素からなる．

◆ VI Provider

VIA に対応した NIC とカーネル内ドライバからなる．

◆ VI Consumer

VI を利用するアプリケーションプログラムのことである．

◆ Virtual Interface(VI)

データの転送操作ををを行うために，VI Consumer が VI Provider に直接アクセスすることができる仕組みであり，ネットワークインターフェースカード (NIC) がユーザプロセスに提供する．送信用と受信用の 2 つの Discriptor Queue と Doorbell と呼ばれる機構を持つ．

各ユーザプロセスはそれぞれの VI を独占的に保有することができる．あるプロセスが他のプロセスに対してメッセージを送受信する必要が生じた場合，まずディスクリプタをキューに書き込む．ディスクリプタとは VI Provider が行うべき処理情報を含むポインタのような構造体である．ディスクリプタがキューに書き込まれたという事実はアドレスなどの情報とともに Doorbell によって NIC に通知され，VI NIC はディスクリプタを参照しながらプロセスに DMA することでデータを転送する．

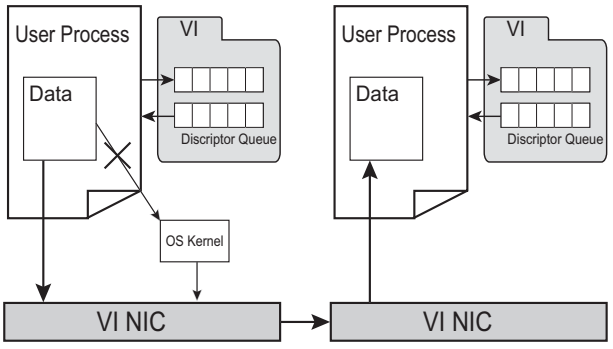

Fig. 3　The Flow of Data on VIA

NIC にメッセージを送受信するする際，従来はシステムの物理メモリーを使う必要があり，この要求が OS カーネルを経由することで処理の遅れが生じていたが，VIA ではプロセス間のデータの送受信をカーネルの介在を必要とすることなく行うことができるため，飛躍的にオーバーヘッドの低い通信を実現できる (Fig. 3).

今回はこの VIA を Linux 上のモジュールとして組み込んだものである M-VIA(Modular-VIA)[6]を導入し，その性能の評価を行った.

2.3　計測結果

計測に用いたクラスタのノードは 2 台であり，それらのマシンの主な仕様は Table 1 に示す.

Table 1　Status

	Node 1	Node 2
CPU	233MHz	344MHz
Memory	64MByte	64MByte
OS	Debian GNU/Linux	Debian GNU/Linux
Kernel	kernel2.2.17	kernel2.2.17
Ethernet Card	DEC Tulip	DEC Tulip

M-VIA 環境で MPI を使用するために MVICH[7]をコンパイルした．ベンチマーク測定にはネットワーク性能のみを評価するプログラム[8]を用いた．計測結果を以下 Fig. 4，Fig. 5，Fig. 6 に示す．これらの結果を見ると，バンド幅には大きな変化は見られないが，レイテンシは著しく向上していることが分かる.

2.4　今後の課題

以上の結果から VIA はクラスタシステムの性能の向上に有効であることが分かった．今後は，これらを Cambria などの大規模クラスタシステムにも導入し，並列処理環境をより快適なものとすることが考えられる.

[6]http://www.cs.berkeley.edu/ philipb/via/
[7]http://www.nersc.gov/research/FTG/mvich/
[8]http://www.aragriculture.org

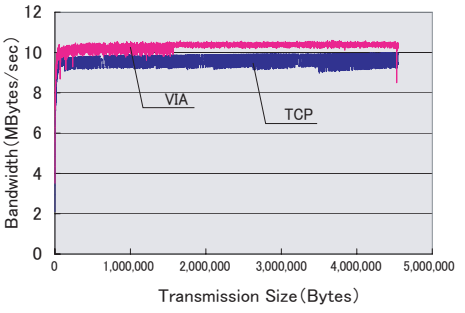

Fig. 4　Comparison of Bandwidth

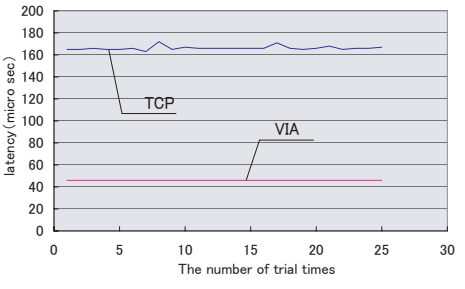

Fig. 5　Comparison of Latency(Minimum)

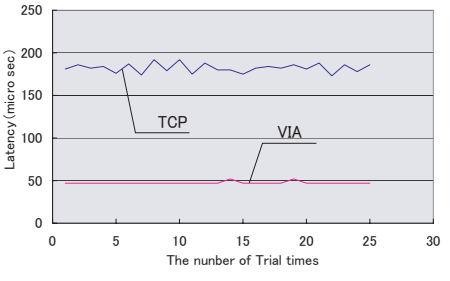

Fig. 6　Comparison of Latency(Average)

参考文献

1) http://www-fp.mcs.anl.gov/fl/accessgrid/
2) http://www.accessgrid.org/agdp/
3) http://foxtrot.ncsa.uiuc.edu:8900/public/AGIB/
4) http://www.nersc.gov/research/FTG/via/
5) http://www.pds-flab.rwcp.or.jp/comet/paper/JSPP99-koba.pdf