P2P

上川 純一

1　はじめに

High Performance Computing (HPC) の分野において耐障害性を提供するシステムとして，P2P システムが考えられる．P2P では，メッセージが集中的に管理されることなく目的地に到達する．例えば freenet，gnutella，napster 等はファイル交換を目的とする，代表的な P2P のシステムである．これらは，各ノードがクライアントであると同時に，ファイルサーバとしての機能を果たしている．P2P では単一サーバに負荷が集中せずサーバを増強することなくより大規模なシステムに対応できる．しかし，P2P には実装しているアプリケーションの動作が遅いという問題がある．例えば，gnutella ではデータの検索に長時間を要する．アプリケーションのアイドル時でもネットワークの維持をするためだけに，ネットワークの帯域を大量に消費する．

大きな PC クラスタを構築し，その上にアプリケーションを構築するためには，P2P 的な設計を導入したシステムが必要であると考えられる．そこで，そのようなシステムに特化したシステムで，スムーズに効率良く動作できるネットワークの構成を考える．現存のシステムの多くは目的が特化しすぎているため他の用途に応用しにくいため汎用のシステムを構築する．

2　実装

PC クラスタは一般的な PC をネットワーク接続することにより構築する並列計算機である，PC クラスタ上で稼働するアプリケーションには静的なノード数が必要なものだけでなく，ノードの負荷状況や故障状態に応じて柔軟に使用するノード数を変更することが可能なものも存在する．より大きなシステムを P2P の特徴をもったシステムを PC クラスタ内に実装する．提案システムでは，ツリー形式のトポロジを利用して情報を伝達する Fig. 2．これは，一般的な P2P システムのノードがグラフを構成しているトポロジとは大きく異なる．しかし PC クラスタでは必ずマスターノードが存在するため，ツリー構造を利用することは有効である．

ツリー構造を導入すると，中間で情報を中継するノードというものが存在する．それらのノードが停止すると，下流の情報が上流まで到達しないという問題が起きる．その問題を解決するために，ノードが停止した場合に上流を迂回するための手法を作成した．各ノードは自分の上流にあるノードの情報を「Route-To」として取得

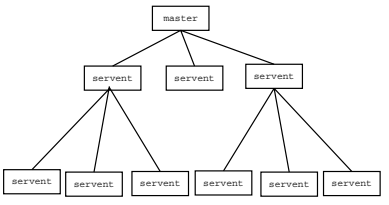

Fig. 1　Tree structure within cluster

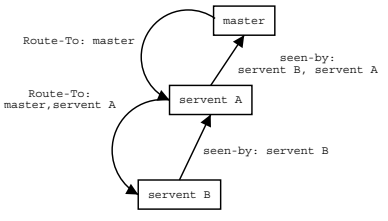

Fig. 2　Seen-by and Route-To

する．その情報を利用すると，自分の直接上にいるノードだけでなく，その上のノードの情報等も分かる．ここで，自分の直接上にあるノードに接続できなかった場合に，そのさらに上のノードに対して接続する，という手法をとることにより，ツリー構造の場合の中間ノードの問題というものが解決できる．また，そのようにして上流のノードに過多に接続した場合に上流のノードが負荷を軽減できるような手法も作成した．各ノードは自分に接続している下流のノードがどれであるか，という情報をもっている．下流のノードが多すぎるのなら，その一つに対して，他の下流のノードを上流にするように指定することができる．この二つのメカニズムにより，動的に変化するツリー構造がつくり出せる．

3　成果

PC クラスタ内における階層的な構造を持つ P2P 的システムを提案した．このシステムでは高い耐障害性と，効率の良い通信を両立することを目標としている．通信量の多い数値計算クラスタ等の計算のためのデータ通信の枠組として利用されることを想定する．このシステムは，階層的な構造を取りリレー方式で情報をやりとりする．さらに，木構造を動的に変化させて，耐障害性を実現している．PC クラスタには代表的ノードが存在するという性質を応用してツリー構造の通信トポロジが実現できた．