グループコミュニケーションにおけるSmall Group Multicastの有用性

松澤智史† 中山雅哉††

本稿ではグループコミュニケーションを行う上で，このSmall Group Multicast技術が有用であることを評価するために、グループ内コミュニケーションツールであるIRC(Internet Relay Chat)のグループ単位での利用状態を調査し，ユーザ数やグループの規模または生起・継続時間，ユーザの接続時間といった観点からの実データを用意し，従来のS/Cモデルや従来のIP Multicastに比べて，Small Group Multicast利用が有用であることを示した．

1. はじめに

近年，同一のデータを複数の対象に効率良く配送するメカニズムとしてIPマルチキャストの研究が行われてきた．IPマルチキャストはマルチメディアデータなどの大容量データが一般的になるにつれて，期待度も高まり，数多くの関連研究が盛んに行われるようになってきた．当初IPマルチキャストは，大規模なデータを多数のユーザに効率良く配送することを念頭に研究開発されてきたが，最近では，必ずしも多数ユーザに配信を行うものだけではなく，小規模のメンバーで構成される小さなグループでの情報共有やコミュニケーションを行うアプリケーションも増えてきており，そのような環境で使用される従来のIPマルチキャストのメカニズムでは，必ずしも効率の良いデータ配信とはならない．そこでこのような小規模グループのIPマルチキャスト専用のメカニズムとして，SGM(Small Group Multicast)という技術が研究開発されて注目を集めている．SGMは名前の通り小規模のグループでのマルチキャスティングを非常に効率良く実行できるが，従来のIPマルチキャストが想定した多数のユーザを対象にした場合は，逆に効率が悪くなる．本稿では，グループコミュニケーションを行おうとした場合に，Small Group Multicastの使用が有用であることを示すために，実際のグループコミュニケーションツールであるIRCを使用し，そこで通常行われているコミュニケーションをモニタしたデータを使用して，ユーザ数，グループの生起・継続時間，ユーザの接続時間と言った点からSmall Group Multicast利用の有用性を示していく．第2章ではIRCとSmall Group Multicastの解説，第3章で関連研究、第4章でIRCによるSGMの有用性を示す調査の結果，そして最後の5章でまとめを述べる．

† 東京大学大学院 工学系研究科
School of Engineering, University of Tokyo
†† 東京大学情報基盤センター
Information Technology Center,The University of Tokyo

2. IRCとSmall Group Multicastについて

2.1 IRCとは

IRC(Internet Relay Chat)[1]は1988年にフィンランドで開発され，1990年には日本でも利用が開始されたネットワークリアルタイム会議システムであり，複数のサーバが相互接続されて1つのIRCネットワークを形成している．

IRCネットワークにはチャンネルというグループのコミュニケーション単位を持ち，同一のチャンネルにJOINしているユーザ同士で会話を行う．チャンネルはチャンネル名という文字列によって識別され，IRCネットワーク内で同一のチャンネル名を持つチャンネルは1つしか存在しない．

本稿ではこのIRCのチャンネルをグループとして見たて調査を行った．

2.2 Small Group Multicast

従来のマルチキャスト[2)3)]は、巨大なマルチキャストグループを扱うように設計された。この設計では、世界規模の配送を行う場合にうまく動作をする。しかしながら、非常に多数のグループを扱うことになると

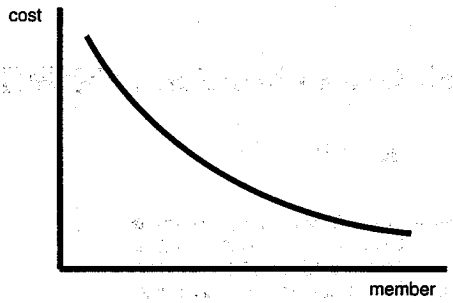

図 1　メンバー毎の従来の Multicast のコスト

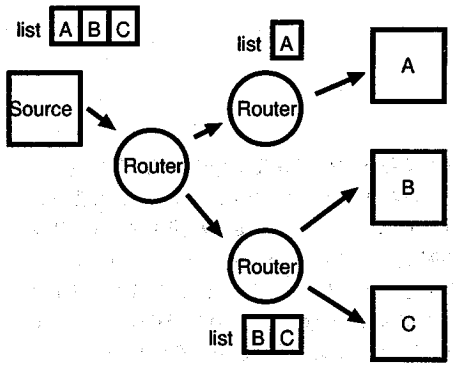

図 2　Xcast の配送メカニズム

スケーラビリティの問題が発生する。

図 1 は、メンバー数に対する従来の Multicast のコストを表したものであり、少数のグループではメンバ毎のコストが非常に高くなる。

このような従来のマルチキャストでは想定していない少人数グループの Mulitcasting を効率良く行うために、SGM(Small Group Multicast)[4] と呼ばれる技術が研究されてきた。SGM は、従来のマルチキャストのメカニズムとは根本的に異なり、ユニキャストのルーティングをベースに配送が行われる。SGM ヘッダと呼ばれる 3.5 層のヘッダ情報を用いてデータ配送を行う。このため従来の通常のルータでは SGM の配送は可能ではない。

SGM の代表的なプロトコルとして Xcast(Explicit Multicast)[5] がある。Xcast は IP ヘッダに宛先リストを入れ、必要に応じてルータでコピーを行いマルチキャストを実現する。このため、特殊なマルチキャストルーティングは必要なく、ユニキャストのルーティングテーブルに従い、宛先リストに存在するノードに応じて隣接のルータへ配送していく。(図 2)

このようにルータは、セッションやグループ毎の状態を維持する必要がなく、特殊なマルチキャストアドレスも不要であり、ユニキャスト以外の特別なルーティングも不要であるため、少人数グループでは従来のマルチキャストよりも効率良くデータ配送が可能になっている。

3. 関連研究

Rick Boivie らの"Small Gorup Multicast: A New Solution for Multicasting on the Interet"[6] では，既存の IP Multicast で使用されるプロトコルの欠陥としてマルチキャストのアドレス空間の問題や，ドメイン間にまたがるマルチキャスト通信の困難さの問題を PIM-SM☆などの具体的なルーティングプロトコルを例にあげて提起している．そしてその解決方法として Small Group Multicast のアーキテクチャを紹介しているが，具体的なユーザやグループ規模の実例を元に Small Group Multicast の需要を述べた研究事例はない．そもそも IP Multicast 自体が今現在もそれほどポピュラーとなっておらず，Small Group を対象としたルーティングの研究も，ある程度アプリケーションを意識して行っているため，グループの規模などの調査も目的のアプリケーションに依存している．

4. IRC による有用性調査

4.1 グループの規模調査

4.1.1 チャンネルごとのユーザ数推移

文献[7] で示した方法により取得した各 IRC ネットワークにおける測定開始時からのチャンネル (グループ) への参加ユーザ数を図 3 と図 4 に示す．

図 3 は GimpNet のユーザ数推移，図 4 は OpenProjectNet のユーザ数推移である．x 軸には，Channel 各自に付与したチャンネル ID を示し，y 軸は 10 分単位の計測開始時からの経過時間，z 軸はそれに対応するユーザ数を表している．

このグラフによって常時多数ユーザを抱えるチャンネルと，短時間で少数のユーザで構成される一時的なグループチャンネルの傾向を視覚的にとらえることができる．

4.1.2 ユーザ数とチャンネル数の変移

前節の図 3 と図 4 に示したように，同じチャンネルでも時刻に応じて参加する人数が変動するため，計測期間中の各時刻における各チャンネルの参加人数を

☆ 従来の IP マルチキャストの一般的なルーティングプロトコルであり，sparse な環境での使用を目的にしたプロトコル

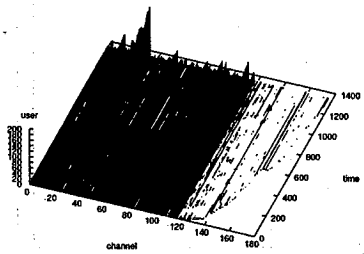

図 3 GimpNet のチャンネルごとのユーザ数推移

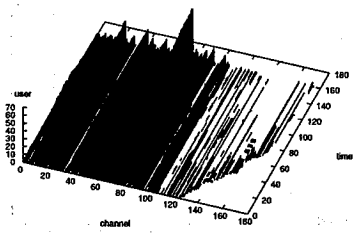

図 4 OpenProjectNet のチャンネルごとのユーザ数推移

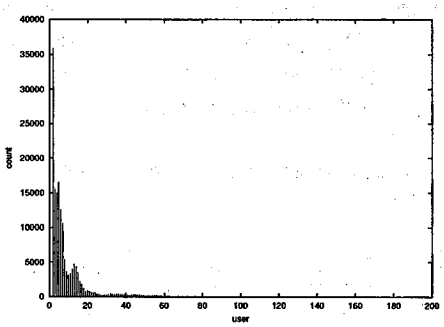

図 5 GimpNet のユーザ数とその度数

別々のグループとしてとらえた時のユーザ数を示すと図 5 と図 6 のようになる．

図 5 は GimpNet におけるユーザ数とそのユーザ数で構成されたチャンネルの度数を表しており，x 軸がユーザ数で y 軸はユーザ数ごとのチャンネル数を示している．1 チャンネルの最大数は 3 月 12 日の 2 時に記録した # gnome の 189 人であり，チャンネルあたりの平均ユーザ数は 10.3 人，標準偏差は 17.11 人となった．総度数は 164213 個であった．表 1 は図 5 におけるそのユーザ数以上の累積度数が 50%,60%,70%,80%,90%を

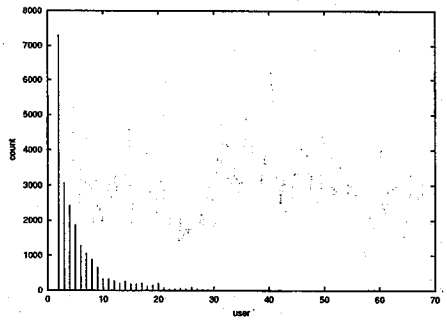

図 6 OpenProjectNet のユーザ数とその度数

超えたものを示したものである．この結果から 20 人以下のチャンネルの累計が全体の 90%以上となることがわかる．

表 1 GimpNet の累計とその割合

user	count の累計	全体の割合
5	84534	51%
7	108167	66%
9	117346	71%
13	132858	81%
20	148192	90%

図 6 および表 2 は，OpenProjectNet での同様の集計結果である．1 チャンネルの最大数は 3 月 12 日の 2 時に記録した # slashdot の 68 人であり，チャンネルあたりの平均ユーザ数は 6.12 人，標準偏差は 7.29 人となった．チャンネル総数は 21863 個であった．

表 2 OpenProjectNet の累計とその割合

user	count の累計	全体の割合
4	12798	59%
5	14673	67%
6	15950	73%
8	17904	82%
13	19721	90%

また 13 人以下のチャンネルの累計が全体の 90%以上を示すことがわかる．これらの結果から，ほとんどのチャンネルは非常に少ないユーザ数で形成されたグループであり，その規模は 20 人以上で約 90%以上をカバーすることができる．

4.2 グループ数の調査

次に前節の測定で得られたデータより、ある時刻に同時に存在するグループ数を調べる．得られたデータ

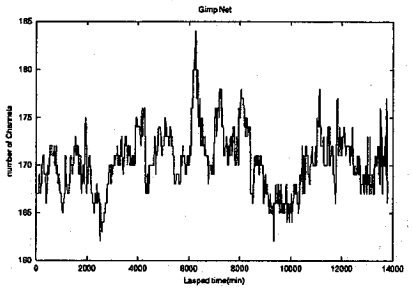

図 7　GimpNet のグループ数の推移

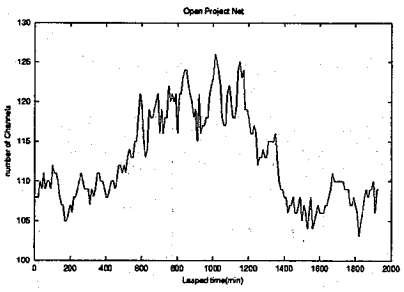

図 8　OpenProjectNet のグループ数の推移

は図 7 と図 8 のグラフとなった．

4.3　グループの生起・継続時間調査

次に，グループがどの程度の時間の間構成されているかを調査する．調査データの入手方法時間等は規模調査のものと同じである．ここでのグループは，あるチャンネルが作成されてからユーザがいなくなるまでに形成されるコミュニティを 1 つとして考えることにした．すなわち同じチャンネルでも一旦ユーザがいなくなり，また後ほど形成された場合は別グループとみなすことにしている．

図 9 は，GimpNet においてグループが生起してからの継続時間とその頻度を表したグラフである．また図 10 は OpenProjectNet のグループ継続時間とその頻度を表したグラフである．x 軸は継続時間を 10 分単位で表した数値であり，y 軸はそれに対応するグループ数である．

各々の図において右端に現れている頻度の高いデータは計測開始から終了までずっと継続しているグループを示している．このようなグループ (チャンネル) は，GimpNet では全 613 グループ中の 84 グループ

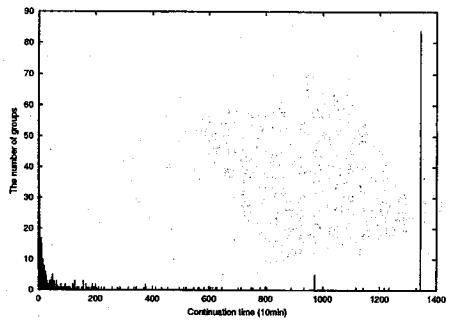

図 9　GimpNet のグループの継続時間

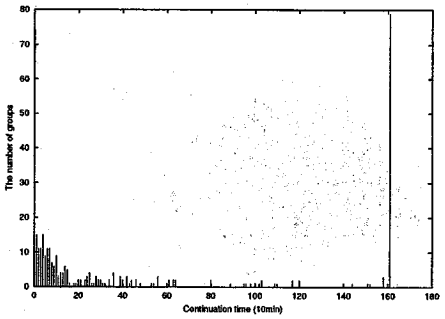

図 10　OpenProjectNet のグループの継続時間

がこれに該当し，OpenProjectNet では 286 グループ中 79 グループが該当していた．GimpNet の残り 529 グループを対象に，全体の 50-90%を占めるグループ数に対応する継続時間は表 3 の通りである．

表 3　GimpNet の継続時間以内に該当するグループの累計とその全体の割合

継続時間	グループの累計	全体の割合
150	271	51%
220	319	60%
450	371	70%
830	424	80%
2060	477	90%

830 分以下の継続時間のグループ数は，全体の 8 割を占める 424 グループ存在していた．

OpenProjectNet の残り 207 グループを対象に同様のことを行うと，表 4 に示す結果を得られる．

表 4　OpenProjectNet の継続時間以内に該当するグループの累計とその全体の割合

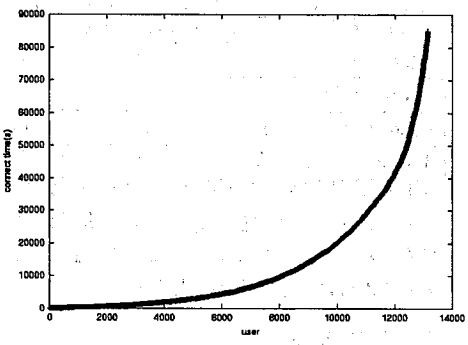

図 11　IRC ユーザの接続時間調査

継続時間	グループの累計	全体の割合
100	105	51%
150	127	61%
270	145	70%
420	168	81%
800	187	90%

継続時間が 800 分以下のグループは 187 グループで全体の 9 割を占める.

これらの結果により，IRC でのグループは，ある程度の規模の数だけ常時 (継続時間不明) が存在するが，それ以外のグループは約 13 時間 (800 分) 以下に約 80%が集中することが明らかとなった.

4.4　ユーザの接続時間調査

次に焦点をチャンネルではなく，ユーザに移す．本節では IRC に参加するユーザがどの程度の期間接続をするかという調査を行う．前節まで使用したデータは，チャンネルごとのある時間においての接続数であるため，どのユーザが接続したかまでは判別不能であるため，別データを使用する．本調査での使用データは，IRC 利用者である我々が約 1 年半の間 IRC を利用した際に取得したログデータを使用する．グデータは約 40 程度のチャンネルの 2 年分のデータであり、このログデータで検出されたユーザ数は 13155 ユーザであった。測定は 24 時間以上接続しているユーザは常時ユーザとみなしてカウントせず、1 日単位でのログで join leave を行ったユーザの接続時間を出している。

結果は図 11 のようになった。

x 軸はユーザ 13090 人を接続時間順にソートしたものに番号を振ったものであり、y 軸は接続時間 (MAX24 時間) を秒で表記したものである。ログは分単位での記録のため、各ユーザの接続時間の誤差はそれぞれ最大 1 分になる．この分布を簡略に示したものが表 5 である．

表 5　ユーザの IRC への接続時間の分布

接続時間	該当ユーザ数	割合
15 分以内	2693	20.6%
30 分以内	3876	29.6%
1 時間以内	5439	41.6%
2 時間以内	7181	54.9%
3 時間以内	8220	62.8%
6 時間以内	10121	77.3%

表 5 より，過半数のユーザが 2 時間以内の接続時間で終了していることがわかる.

4.5　評　　価

上記の 4 つの実験を行った結果，グループに関しては，20 人以下のメンバーで構成されるグループが 90%，継続時間が 13 時間以内であるグループが 90%(そのうち 100 分以内が 50%) であることが判明し，ユーザの接続時間に関しては，6 時間以内の接続時間であったユーザが約 80%で (そのうち 2 時間以内は 50%以上) あることが判明した．これらのデータを元に，グループコミュニケーションを行うにはメカニズムで行うのが良いかを考察する.

4.5.1　グループ管理方式

まずはグループの管理方式の比較を行う．従来の S/C モデルは，中央管理型のモデルであり，サーバがすべてのグループに属するユーザのクライアントの情報を管理する必要がある．つまり全ユーザの情報と全チャンネルの情報をサーバで一括管理する必要があるため，サーバにかかる負荷は非常に高い．このモデルを採用している現存の IRC はサーバを複数に分割し負荷分散を行っているが，サーバに直接接続するクライアント数の減少のみ可能で，管理するチャンネル数などは減らせない.

次に従来の IP Multicast でグループコミュニケーションを行う場合，管理の方式はグループごとに Multicast Address を割り当てて行う．しかしながらグループの数の調査で，小さな IRCnetwork である GimpNet や OpenProjectNet でも 8bit のチャンネル識別子が必要であり，比較的大きな IRCnetwork である Friend chat では 10bit でも収まらない．IP Multicast は classD と呼ばれる 224.0.0.0-239.255.255.255 ですべてのサービスのグループを管理する必要があり，

このような少人数グループが大量に出現するサービスでは，将来的には枯渇する問題もあり，とても良いメカニズムとは言えない．

SGM では，中央管理の特別なサーバも不要であり，アドレス空間などにも制約はない．基本的にグループを構成するホスト間で管理情報を共有するだけなので，上記の問題点はすべて当てはまらない．

4.5.2 接続にかかる時間

次にグループに参加してコミュニケーションを行うまでの接続にかかる時間についての考察をする．従来の S/C モデルはアプリケーションによる使用プロトコルの差はあるものの，接続にかかる時間は基本的にユニキャストでのコネクションにかかる時間に依存する．SGM も多少の差異はあるものの，基本的には S/C モデル同様ユニキャストでの通信がメインであるのでほぼ同様と言える．しかし従来の IP Multicast では，接続にかかる時間でデメリットが生じる．IP Multicast はグループに参加を行うために，マルチキャストルータと呼ばれるルータを経由していく必要がある．この際経由するマルチキャストルータの DB をすべて更新していく必要があるため，管理元のマルチキャストルータが遠いほど hop 数による接続時間のオーバーヘッドが，ユニキャストの場合に比べて大きい．ユーザの接続時間調査の結果より，15 分以内のユーザで 20%を占めるグループコミュニケーションでは，この接続時間のオーバーヘッドは無視できないものとなる可能性が高い．

4.5.3 ルータにかかる負荷

最後にルータにかかる負荷を比較してみる．

ルータに特別な処理や state 維持が皆無である S/C モデルによるルータの負荷は，基本的に最小であるが，中央管理を行うサーバに近ければ近いほど bandwidth は大きくなる．従来の IP Multicast では，グループの情報を管理する必要があるため，グループ数に比例して state 管理情報が肥大化する．SGM はグループを構成するユーザ数に比例してパケットあたりのルータの処理が増加する．

しかし，本調査で，グループコミュニケーションは非常に少数の構成員によるグループがたくさん作成されることが判明したので，IP Multicast の欠点は増大し，SGM の欠点は非常に少ないと言える．

4.5.4 評価まとめ

これまでの評価を表にすると表 6 となる．

表 6 評価のまとめ

	S/C	IP Multicast	SGM
group 管理	負荷大	address 枯渇	なし
接続時間	小	大	小
router 処理	1	1	M☆
state 維持	0	G☆☆	0
bandwith	大	小	小

表中の M はグループを構成するメンバー数 M であり，G はグループ数の値が入る．M は非常に少なく，G は多くなることが判明したので，この表より SGM 利用の欠点はほぼなくなると見れる．

5. おわりに

本稿では，グループコミュニケーションを行う上で，現状のサーバクライアントモデルでの実装である IRC をターゲットにし，IRC を利用するユーザの傾向や，作成されるグループの傾向を調べ，このようなグループコミュニケーションを行うには，現状のサーバクライアントモデルよりも，Small Group Multicast と呼ばれる少人数グループでの few to few マルチキャストを実現する方式を用いることが有効であることを示した．

参考文献

1) J. OikarinenS, D. Reed: Internet Relay Chat Protocol, RFC1459, May 1993
2) S. Deering, "Multicast Routing in a datagram internetwork", PhD thesis, December 1991.
3) S. Deering, D. Estrin, D. Farinacci, V. Jacobson, C. Liu, and L. Wei. "The Pim Architecture for Wide-area Multicast Routing", ACM Transactions on Networks, April 1996.
4) R. Boivie, et al. "Small Group Multicast, Internet Draft", work in progress, draft-boivie-sgm-02.txt, July 2001.
5) R. Boivie, N. Feldman, Y. Imai, W. Livens, D. Ooms, O. Paridaens "Explicit Multicast (Xcast) Basic Specification" work in progress, draft-ooms-xcast-basic-spec-04.txt, January, 2003.
6) R.Bovie, N. Feldman, C.Metz "Small Group Multicast: A New Solution for Multicasting on the Internet" IEEE Computer Society, Internet Computing Online May June 2000.
7) 松澤智史, 中山雅哉:IRC におけるグループの生起時間と規模の研究, 情報処理学会 研究報告「高品質インターネット」 No.007, May 2003.