分散データストリーム処理における自律Pull制御方式の改善

内藤　一兵衛†　　赤間　浩樹†　　山室　雅司†
†日本電信電話株式会社　NTTサイバースペース研究所

1　はじめに

センサやＩＣタグなどから発生するデータストリームを受け取り，即時，分析等の処理を行うデータストリーム処理システムが注目されている．我々は，分散環境上で外部からの入力を受け止める追記部と，自律的にデータ取得先を決め，処理を行うフィルタ部を分けるアーキテクチャを採用した追記・参照型データ管理システム（以下DMS: Data Management System）[1]を開発している．DMSはセンサデータやネットワークのパケット，システムログなどの小さいデータ系列から，音声，映像データなどの大きなデータ系列までデータサイズの異なる異種データストリームを受け取り，データセンタに設置された数百台規模のPCクラスタで処理することを想定している．

2　DMSでの自律分散Pull制御方式

DMSは図2に示すように冗長化された情報部（I）と複数の追記部（Q），フィルタ部（F），ビュー部（V）から構成される．追記クライアントから送信されるデータを追記部で管理し，フィルタ部はデータを受け付けている追記部群の中からデータ取得先を自ら選択し，データを取得（Pull）し，処理を行いビュー部に送信する．

追記部からフィルタ部にPushする方式も考えられるが，各データの処理時間にばらつきがあった場合に時間のかかる処理を行っているフィルタ部に，さらに他の追記部からデータがPushしてしまいフィルタ部内にデータが滞留し，他のフィルタ部が処理できる状況であっても，負荷が分散されない．フィルタ部が取得対象の追記部を選択してデータを取得し，処理することを繰り返す自律分散Pull制御方式により特定のフィルタ部にデータが滞留することなく，負荷が分散できる．また，全ての機能を止めずに，動的にフィルタ部の機能を更新させることができるという付加効果もある．フィルタ部のデータ取得先追記部選択の方式は文献[2]で提案されている．

2.1　自律分散Pull制御方式の課題

一方で，自律分散Pull制御方式ではフィルタ部間の通信を行わず，個々のフィルタ部が独立に選択先を選ぶ

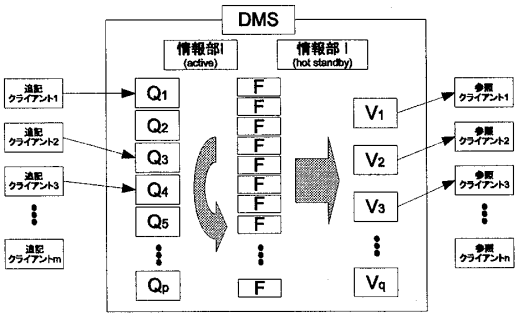

図1: アーキテクチャ概要

為，タイミングによっては特定の一つの追記部にデータ取得を行ってしまう．これにより追記部内のデータキューへのI/Oが集中し，アクセスに来たフィルタ部全てを受け付け，にデータを送信しようとして，追記部のレスポンスが低下する問題が発生する．これをバックプレッシャーと呼ぶ．バックプレッシャーが起こると，フィルタ部自身もレスポンスが低下した追記部の待ち状態となり，系全体のスループットも低下する．大規模なクラスタを構成した場合，フィルタ数も増えバックプレッシャーが系全体へ与える影響は大きくなる．

この事象は追記部が追記クライアントから受け取るデータが増加し，フィルタ部に取得要請を行った場合や，システム起動時にも発生する．

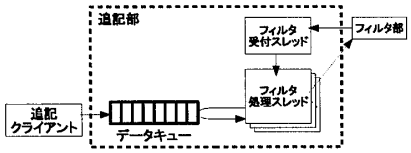

図2: 追記部概要

3　アプローチ

系内のスループットが低下するバックプレッシャー対応のためには以下の３つのアプローチが考えられる．

1. バックプレッシャーが発生しない規模で複数セットのDMSを使って運用する．
2. フィルタ部間で取得先を調整する．
3. 追記部がフィルタ部からの一定以上のアクセスを拒否する．

案１の場合，複数セットのDMSを用意するとそれらの管理コストがかかってしまう．案２の場合，フィル

Improvement of Autonomous Distributed data stream system
†Ichibe NAITO　Hiroki AKAMA　Masashi YAMAMURO
†NTT Cyber Space Laboratories

タ間で取得先を調整するアルゴリズムをフィルタ部へ追加することは可能であるが，大きなオーバーヘッドが予測されるため，本論文では案3を選択した．

3.1 設計

追記部ではクライアントからのデータストリームの入力を受け付け，キュー（データキュー）で管理し，フィルタ部からのデータ取得リクエストに応じてキューからデータを取得しフィルタ部に送信する．フィルタ部からのデータ取得リクエスト拒絶の流れについて以下で説明する．

3.2 フィルタ部拒絶方式の設計

図3.2に追記部設計の概要を示す．追記部内でアクセスに来たフィルタ部のディスクリプタ，IPアドレス等の情報をフィルタ情報キューに蓄積する．予め起動したフィルタ処理スレッドはフィルタ部情報をフィルタ情報キューの先頭から取得し，該当フィルタ部にデータキューからデータを取得して送信する．これによりデータキューへのアクセスはフィルタ処理スレッド数により制限される．フィルタ情報キューには最大値を設定しておき，最大値を超えた場合にフィルタ部にアクセス拒否通知を送信することでフィルタ部のデータ取得待ちを防止する．

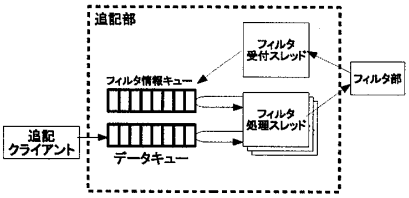

図3: 設計概要

4 評価と考察

4.1 評価方法

バックプレッシャーは多数のフィルタ部が頻繁に少数の追記部へアクセスする状況で起きやすい．そこで，入力データサイズ4KBとしフィルタ部での処理は取得して破棄するだけとし，追記部2台に対して40ストリームを入力し，フィルタ部の台数を増加させた時の系内のスループットを従来方式とフィルタ情報キューの長さを固定にした本方式で測定した．フィルタ部は1台につき4プロセス起動している．

4.2 結果

フィルタ部4台時のスループットを基準としたスループット比を実験結果を図4.2に示す．従来方式ではフィルタ部のマシン台数が16台になるとスループットが落ちているが，これはアクティブな追記部に対してフィルタ部が多いためバックプレッシャーが起きているからである．これに対し，追記部での拒絶方式を採用するとスループットが落ちることなく，スケールアウトすことが確認できた．

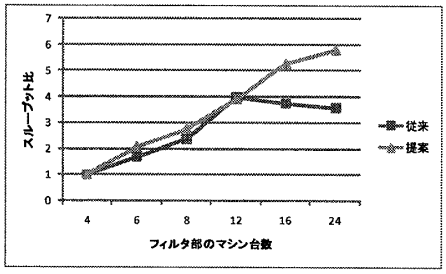

図4: 追記処理レート比較

5 フィルタ情報キュー最大長の検討

上述したフィルタ情報キューを用いたフィルタ部からのアクセス管理において，管理者がフィルタ情報キューの最大長を静的に設定した場合，これを超えた後にアクセスしたフィルタ部に対して即時にアクセスを拒絶する．このフィルタ情報キューの最大長を固定にした場合，追記部への入力が少ないときにはフィルタ部を待たせてしまう．また，追記部への入力が多いときに，フィルタ部からのアクセスを拒絶してしまうと，追記部内での未処理レコードが増加してしまう．データキューの入出力のバランスにより動的にフィルタ情報キューの最大長を変更した方が全体のスループットが向上すると考えられる．

6 まとめ

本論文ではデータストリームの並列処理を実現する自律分散Pull制御方式において，発生するバックプレッシャーに対応するために追記部においてフィルタ部からの一定以上のアクセスを拒絶する方式を提案した．また，これにより一定のバックプレッシャーに対応できることを示した．今後は，フィルタ情報キューの最大長を動的に変化させた場合の評価が必要だと考えている．

参考文献

[1] 赤間　他：追記・参照型データ管理システムの設計と評価，情報処理学会論文誌，Vol.49 No.2, pp.749-764（2008）．

[2] 内山　他：分散データストリーム処理における適応型リソース制御方式の検討，情報処理学会ＤＰＳ研究会，pp.49-54（2008）．