

機関番号：34304

研究種目：若手研究(B)

研究期間：2008～2010

課題番号：20700072

研究課題名(和文) 大規模システム管理のための自律分散型モニタリングシステム

研究課題名(英文) Implementation of Autonomous Distributed Monitoring System for Large-scale Distributed Systems

研究代表者

林原 尚浩 (HAYASHIBARA NAOHIRO)

京都産業大学・コンピュータ理工学部・助教

研究者番号：20397227

研究成果の概要(和文)：

大規模分散システムにおいて信頼性の高いサービスを構築する際にはリソースの管理が不可欠であり、その中でも故障検出は重要な役割を担っている。本研究課題では、高精度の故障検出器の実装とその評価を行い、大規模分散システムにおいて故障検出器の連携と故障情報の提供を行うアルゴリズムの開発を行った。

研究成果の概要(英文)：

Resource management, especially failure detection, is an essential component to build reliable services in large-scale distributed systems. In this project, we implemented the ACCMOS failure detector as a platform-independent implementation and evaluated its performance. We also developed the algorithm to cooperate with several failure detectors and exchange information on suspicion among them.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,300,000	390,000	1,690,000
2009年度	1,100,000	330,000	1,430,000
2010年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学・計算機システム・ネットワーク

キーワード：ネットワーク，分散システム，故障検出器

## 1. 研究開始当初の背景

コンピュータシステムは、社会的な基盤として認知されており、様々なサービスを提供している。特に、サービスを提供するシステムは、ネットワークインフラの充実や汎用計算機の小型化、低価格化によって、大規模化、複雑化する傾向にある。

このような大規模なシステムを運用するためには、常時システムの管理が必要となってくるが、

多くのシステムにおいて、少人数の管理者が人手でシステムの管理を行っている。

また、提供しているサービスによっては、システムが故障し、サービスが提供できなく

なること によって、多大な被害を被るものもあり、システム管理者の負担は増大している。

しかし、ノードやサービスを監視する際、クライアント・サーバ型の通信を行うため、数千台 規模の企業、大学、官庁などのネットワークリソースを網羅的に監視するためには、これらのシステム監視ツールは不向きであるといえる。また、監視対象の追加や脱退による局所的なネットワーク構成の変化が頻繁に起こる場合、これらをシステム管理者が常に監視することは非常に困難であり、この変更によって新たに設定ファイルを書き直す必要が生じる。

既存のツールによるサービスの監視は、サービスを継続的に提供するという目的では有効である。しかし、近年問題視されている、組織内の計算機による情報漏洩などのような、意図しない サービスの提供や禁じられたサービスの不正利用によるシステムの脆弱性を監視するためには効果がない。

## 2. 研究の目的

本研究課題では、大規模システムを網羅的に管理するための自律分散型モニタリングシステムの開発を目的とし、このシステムを実現するための要素技術に関する研究を行う。

このシステムでは、耐故障分散システムの研究で培われた故障検出器、分散アルゴリズム、自己安定アルゴリズムの理論的な要素を取り入れることにより、故障検出精度やスケラビリティの向上が期待できる。また、モニタリングシステムにおいて構成するネットワークは、物理的なトポロジに依存しないネットワークであることから、一種のオーバーレイネットワークであると捉えることができる。従って、システムの初期化や監視対象の加入、脱退によるシステムの構成変更を自律的に行うアルゴリズムは、オーバー

レイネットワークのトポロジに関する研究とも非常に親和性が高い。よって、本システムの実現の過程で開発されるアルゴリズムやプロトコルは多くの関連分野に寄与する可能性がある。

## 3. 研究の方法

本研究課題では、モニタリングシステムを実現するために以下の手順で研究を行った。

### (1) 高精度故障検出器 ACCMOS の実現

Accrual Failure Detectors モデルを採用した故障検出器 ACCMOS を実装する。従来の Accrual Failure Detectors モデルの故障検出器はハートビート型の通信により監視対象をモニタリングしていたが、ICMP による監視を行うことにより、多種のネットワークリソースを監視することが可能となる。大規模システムにおいては、様々な種類のネットワークリソース(例えば、プリンタ、NAS、ネットワークスイッチなど)を監視する必要があるため、この故障検出器の実現は本研究課題の目的を満たすために必要不可欠である。

故障検出器 ACCMOS は Java 1.6 によって実装されている。ICMP による監視については、Java 1.6 において適切に ICMP パケットを送受信する機能が実装されていないため、独自のライブラリを C 言語で実装し、JNI 経由で Java から呼び出すようにしている。

基本的な構造は従来の Accrual Failure Detectors ベースの実装である  $\phi$ -Failure Detector と非常に共通点が多いが、故障している度合い(suspicion level)を計算するために用いていた sliding window を廃止し、フィードバックループを用いることによって積算した Round Trip Time の傾向を計算するため、メモリ効率が飛躍的に向上した。

### (2) 自動初期化、自律再構成機構

大規模分散システムにおいては、管理すべきネットワークリソースが膨大な数になる可能性がある。そのため、複数の故障検出器によって、システム全体のリソースを管理する必要がある。これらの故障検出器間の連携を行うためのアルゴリズムを開発する。

このアルゴリズムは、複数の故障検出器が相互に通信を行うリンクをオーバーレイネットワークとして捉え、そのネットワークの構成(自動初期化)とネットワーク上のノード(故障検出器)の参加、脱退が起こった際のネットワーク再構成を行うアルゴリズムを開発した。

各故障検出器が持つ監視対象ノードの情報は出版/購読モデルによって交換される。つまり、あるプロセスが監視対象ノード  $p$  の情報を必要としているならば、ノード  $p$  を監視している故障検出器にその suspicion level の提供を行うように登録する。そのプロセスは実行を終了するまでノード  $p$  に関する suspicion level を故障検出器から断続的に受け取ることができる。

#### 4. 研究成果

故障検出器 ACCMOS の実装については、異なる環境において実験を行い、評価を行った。実験環境は以下の3つで行った。

- (1) 無線 LAN (メッセージロス率 0.41%)
- (2) 有線 WAN (都市間)(メッセージロス率 0.09%)
- (3) 有線 WAN (同一都市内)(メッセージロス率 0.05%)

実験結果はChenらによって2002年に提案されたメトリクスの中から正検出率(Query Accuracy Probability)と誤検出レート(Mistake Rate)によって評価した。

故障検出器のパラメータは suspicion level に対するしきい値を 0.7, 1.0, 2.0,

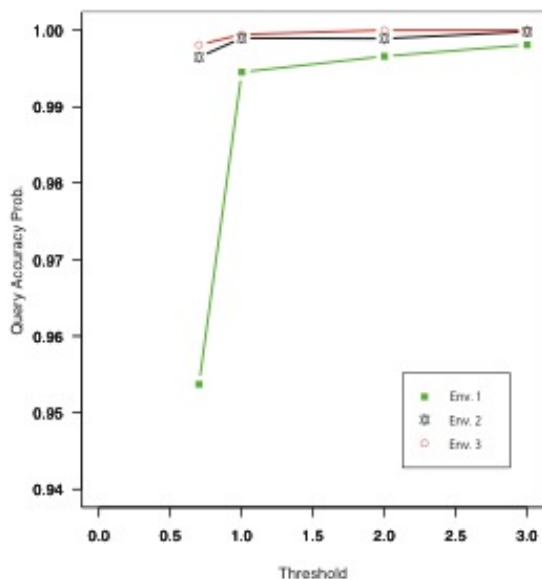


図 1 正検出率 vs. しきい値

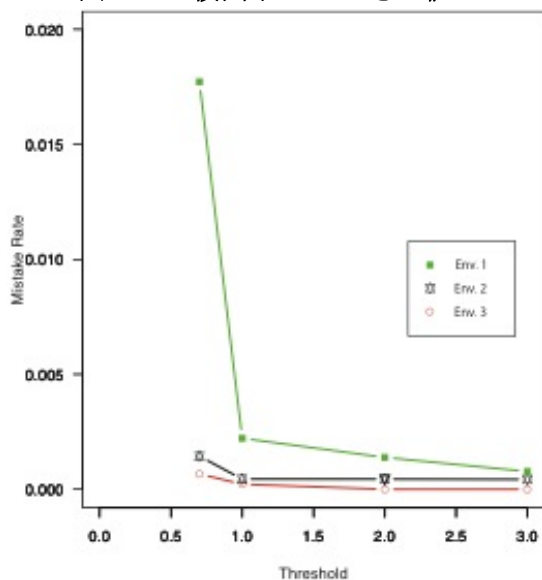


図 2 誤検出レート vs. しきい値

3.0 と設定し、その時の正検出率、誤検出レートを計測した。

図 1 によると、実験環境 2, 3 においては全てのしきい値において、99%以上の精度で故障検出を行っており、非常に高い精度での故障検出を実現できているといえる。また、メッセージロス率が高い実験環境 1 においても、しきい値が 1.0 以上では 99%以上の故障検出を実現できている。

図 2 においては、誤検出(擬陽性, false positive)の値を計測している。これも全体的に低い値となっており、誤検出によるアプリケーションへの影響が少ないことを示している。

自動初期化, 自律再構成機構アルゴリズムについては, 分散アルゴリズム開発評価環境 Neko を用いてシミュレーションを行った。故障検出器の数が 10, 20, 40, 80, 160, 320, 640 となったときのメッセージ数は線形に増加するが, シミュレーションのステップ数よりは小さく抑えられることが分かった。

故障検出器 ACCMOS と共に開発を進めてきた Accrual Failure Detectors モデルは, 現在, Facebook など多くのプラットフォームで用いられている分散データベース Cassandra の一部としても実装されており, 大規模なシステムにおいても信頼性のある故障検出器として認知されている。

本研究課題で実装した故障検出器 ACCMOS を中心としたモニタリングシステムも同様に信頼性の高いアプリケーションやシステムの構築に寄与する能力があると考えられる。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

- ① 林原尚浩, アクルーアル故障検出器 ACCMOS の実装と評価, 情報処理学会論文誌, 査読有, Vol. 51, No. 12, pp. 2310-2318, 2010
- ② Hayashibara, N., Takizawa, M., Design of the Notification System for Failure Detectors, International Journal of High Performance Computing and Networking, 査読有, Vol. 6, No. 1, pp.

25-34, 2009,

[学会発表] (計 2 件)

- ① 林原尚浩, アクルーアル故障検出器を用いた故障検出精度の制御, 情報処理学会 141 回 DPS 研究会, 査読無, 2009 年 11 月 27 日, 広島県, 呉市, 大和ミュージアム
- ② 林原尚浩, アクルーアル型故障検出器 ACCMOS の実装と評価, 情報処理学会 マルチメディア通信と分散処理ワークショップ 2009, 査読有, pp. 247-254, 2009 年 10 月 9 日, 北海道, 上川郡上川町, 層雲峡温泉

[その他]

ホームページ等

<http://rudds.kyoto-su.ac.jp/jp/wiki.cgi?page=Research>

## 6. 研究組織

### (1) 研究代表者

林原 尚浩 (HAYASHIBARA NAOHIRO)  
京都産業大学・コンピュータ理工学部・助教  
研究者番号: 20397227

### (2) 研究分担者

なし

### (3) 連携研究者

なし