

平成 30 年 6 月 19 日現在

機関番号：82645

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00066

研究課題名(和文)ビッグデータを用いたハードウェア障害予測ツールセットの構築

研究課題名(英文)Construction of hardware failure prediction tool set using big data

研究代表者

藤田 直行(FUJITA, Naoyuki)

国立研究開発法人宇宙航空研究開発機構・航空技術部門・主幹研究開発員

研究者番号：70358480

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：大規模化した近年の計算機システムのハードウェアの障害を事前に予測することが求められている。ハードディスクのS.M.A.R.T.情報を機械学習させることによって、ハードディスク障害の予測に関する解析を行った。評価指標として、いわゆる陽性的中率(PRE)を用いた。その結果、壊れる日が近ければ近いほど、PREに影響を与える特有の兆候が表れることが観測された。また、S.M.A.R.T.情報を機械学習させて検出力を高めてあげることによって、かなりの高い精度で障害予測が的中することが示された。故障までの日数と近い期間の判定機による検出力が高いということがわかった。

研究成果の概要(英文)：It is required to predict the failure of the hardware of the computer system, while the system has become large scale. We analyzed the prediction of hard disk failure by machine learning of S.M.A.R.T. information on hard disk. A so-called positive predictive value (PRE) was used as an evaluation index. As a result, it was observed that the closer the day of failure is, the more unique signs affecting PRE appear. In addition, by improving the detection accuracy by machine learning of S.M.A.R.T. information, it was shown that prediction of disability can be expected with considerably high precision. It was found that the detection power of the judging machine during the period close to the number of days before failure was high.

研究分野：並列計算、大規模ストレージ、ネットワークセキュリティ

キーワード：障害検出 原因特定 障害予測 ビッグデータ

1. 研究開始当初の背景

本研究は、コンピュータ内部から発生する内部物理情報(電磁波、音など)に加え、システムから抽出されるパフォーマンス、CPU使用率、メモリ使用率、イベントログ情報、syslogなどの内部状態情報に加え、コンピュータが置かれている環境の温度、湿度、静電気量などの環境ストレスとなりうる外部物理情報からなる膨大なデータを収集し、包括的に精査することによってハードウェア障害を事前に予測するシステムの構築を目的とする。

一般的に、コンピュータシステムは死活監視システムや SNMP (Simple Network Management Protocol) などによって監視されており、問題があれば管理者に通知されるしくみを取り入れていることが多い。このようなシステムは、障害が起きた際にエラーメッセージを出す、冗長化などの影響から障害が顕在化せず、障害検知をすり抜け、障害とは認識されないサイレント障害には十分な効力は発揮しない。そのため、IT 関連のメーカーでは、このような既に起きているサイレント障害を検知するためのソフトウェアを開発し始めている。しかし、システムの停止を伴う明らかな障害であっても、障害が起きる前に予測する障害事前予測に対する問題は依然として残っている。その要因のひとつは、システムから排出される内部状態情報のパラメータが膨大なため、すべてのパラメータを監視することができないことや、個々のパラメータが障害に与える影響、パラメータ間の関係性、評価指標としての重要性などが明らかにされていないことが挙げられる。また、これまでは、計算機の数や性能の観点から、このような膨大なデータを処理することも困難であったため置き去りにされていたという背景もある。そのため、ある日突然システムが停止してしまうという現象は現在でも至るところで起きているのが現状である。宇宙航空研究開発機構においても、20数台程度のサーバに対して、ハードウェア障害も含めて年間10~20回程度の停止が起きている。このような障害が発生すると、管理者は原因の究明と復旧作業などの対応に追われ、一方ユーザも復旧するまでの間、そのサービスを受けられなくなるという事態に陥り、円滑なコンピュータ利用が大きく妨げられることになる。障害の主な原因としては、落雷や停電などの天災や事故、過負荷や製品劣化による内部要因、気温や埃などの外部ストレス、設定ミスなどの人災などが挙げられるが、特にハードウェア障害のように突然の停止が起きると、対応が後手に回ることに加え、管理者も原因の特定に手間や時間がかかり、さらに停止時間が伸びる場合も少なくない。このような現状において、システムの障害を事前に予測するシステムの構築が求められている。しかし、先に述べた通り、障害の起きる原因は多様であり、何

に注目すべきかが明らかではなく、パフォーマンスの低下やトラフィック状況などの内部状態情報を監視しているだけでは予兆を捉えることは難しい。

また、近年、ギガヘルツ帯の微弱な電波や微小な振幅の電気信号を利用した電子機器が広く使われるようになり、電子機器間での電磁干渉による通信障害や機器内部での干渉によるトラブルも増加している。代表者は、以前より多数のサーバや、スーパーコンピュータを管理し運用技術の研究(業績[1,4])に取り組む傍ら、FPGAなどのコンピュータ内部品に関する研究も行っている(業績[3,6,7])。そのような背景から、事前の障害予測の必要性を実感し、本研究への着想にいたった。しかし、通常コンピュータから取得できるログ情報だけでは検知は難しい。そこで、コンピュータ内から発せられる電磁波の微弱な変化などの内部物理情報、システムの設定環境から得られる外部物理情報を同時かつ包括的に精査する必要があるのではないかという着想のもとに、それによって障害の予兆を捉え、事前に障害を予測するシステムを構築することが可能になるのではないかと考えた。そこで、まず、収集したデータから各パラメータの影響を検証し、障害に強く影響を与えていると思われるパラメータを究明した上で、障害の評価指標として分類していくことから始める。また、評価指標となったパラメータの相関関係や因果関係などについてさらに分析し、それらの分析をもとに障害予測のためのモデルを構築する。さらに予測しない障害が発生した場合には、情報をフィードバックし学習機能を持たせることで、予測の精度を高め、さらに精緻な障害予測システムの構築の可能性を探る。

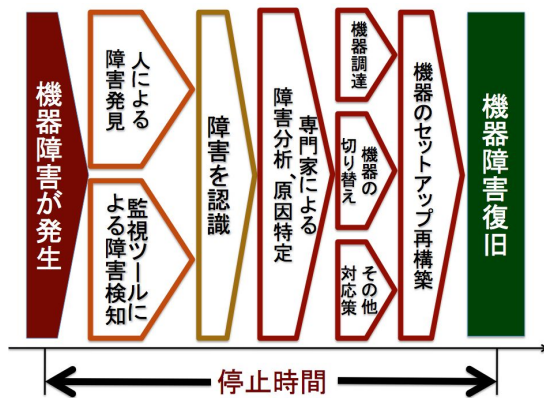


図1 従来の監視による停止時間

2. 研究の目的

本応募研究のテーマは、「ビッグデータを用いたハードウェア障害予測ツールセットの構築」を目的とした研究を行うことである。一般に、コンピュータシステムには、障害の検出、通知、障害原因の特定などの機能を備えた管理ツールが組み込まれているが、障害の検出は、障害が起きた後に有効になるもの

であるため、事前に対応策を打つことは困難である。そこで、本研究では、ネットワーク、サーバ群などの重要かつ大規模なシステムを対象に、ハードウェア内部の物理的現象、システムから抽出できるログ情報、およびシステムを取り巻く物理的環境の状態を捉えた外部的な情報の3つを網羅的に収集し精査することによって、ハードウェア障害が起こる前の予兆を捉え、ハードウェア障害を事前に予測することを目的とする。

3. 研究の方法

本研究で実施する内容は下記の通りである。平成27年度は、内部状態情報、内部物理情報、外部物理情報のデータ収集の実験環境を整備し、データ収集のためのプロファイリングツールを構築する。平成28年度は、障害に影響を与えるパラメータの究明と、異常データ判別ツール、障害予測モデルを構築する。平成29年度は、28年度に構築したモデルを実装し、構築したモデルの検証を行い、予測モデルから漏れた予期しない障害に関する検証を行う。この検証に基づき、さらに精度を高めるための学習モデルを構築する。一連のフィードバックシステムにより、予測モデルの精度を自動的に高めていくことが可能となるシステムの構築を行う。

4. 研究成果

本研究においては、内部状態情報、内部物理情報、外部物理情報の3つの要素からデータを収集するため、初年度は必要な機器の整備を実施した。すなわち、高周波用アンテナや高周波解析用のスペクトラムアナライザなどの内部物理情報測定機器を設置し、気温、湿度計などの外部物理情報収集に必要な機器を解析用PCに繋ぐことでデータ収集に必要な測定環境を整備した。また、整備した測定環境で実際の測定を試行し、測定内容の確認やセンサー設置に関する調整や検討を行った。

整備した測定環境で得られたデータを、高周波、低周波および使用CPU数、メモリなどの大量のデータを収集・保持するプロファイリングツールを構築するための準備を開始した。具体的には、データ自動収集プログラムの作成、観測自動化のためのプログラムの作成などを実施することによって、大量のデータを自動的に収集、保持するシステムを構築するための設計とプロトタイピングを実施した。また、プロファイリングツールにはsyslog解析のための処理プログラムも組み込むことで、内部、外部物理情報、内部状態情報の3つの情報を一箇所にまとめて一括で管理し、解析することを可能とする設計とした。

二年目は、初年度に準備したシステムをもとに、コンピュータ内から発せられる電磁波の微弱な変化などの内部物理情報やシステムの設置環境から得られる外部物理情報を

包括的に精査することによって障害の予兆を捉えるシステムを構築し、障害が発生したハードウェアの実データの代わりに、人工的に生成したデータを用いた場合の異常検知手法について、ROC (Receiver Operating Characteristic) 曲線を用いて比較を行った。

最終年度は、ハードディスクのS.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology) 情報を機械学習させることによって、ハードディスク障害の予測に関する解析を行った。評価指標として、機械学習による予測で壊れると推測しかつ実際に壊れた個体ハードディスクの個数を全ハードディスク個数で割ったもの、いわゆる陽性的中率(PRE)を用いた。その結果、壊れる日が近ければ近いほど、PREに影響を与える特有の兆候が表れることが観測された。また、S.M.A.R.T. 情報を機械学習させて検出力を高めてあげることによって、かなりの高い精度で障害予測が的中することが示された。さらに、学習と故障までの期間という観点から、故障までの日数と近い期間の判定機による検出力が高いということがわかった。このことから、学習に用いたデータの中には、壊れるまでの期間に応じて、その期間特有の傾向が見られるのではないかと考えられる。したがって、データに応じて学習データを使い分けることで、全体として、より検出力の高いシステムの構築が期待できることが改めて確認された。

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3件)

Rika ITO, Naoyuki FUJITA, A Hardware Failure Prediction System Using Big Data Analysis, WIP session of International Conference on Parallel, Distributed and Network-Based Processing(976-3-902457-45-5), No peer review, Vol.1, 2016, 1-2

Rika ITO, Naoyuki FUJITA, An Experimental Detection for Hardware Failure Using Big Data Analysis, WIP session of International Conference on Parallel, Distributed and Network-Based Processing(978-3-90245704806), No peer review, Vol.1 2017, 9-10

Rika ITO, Naoyuki FUJITA, Prediction of Hardware Failure Using S.M.A.R.T. Information, Proceedings of the Work inProgress Session PDP2018, No peer review, Vol.1, 2018, 7-8

[学会発表](計4件)

Rika ITO, A Hardware Failure Prediction System Using Big Data

Analysis, International Conference on Parallel, Distributed and Network-Based Processing, 2016

伊藤利佳、藤田直行、ハードウェア障害検知システムのための実験的検討、電子情報通信学会、2017

Rika ITO, Naoyuki FUJITA, An Experimental Detection for Hardware Failure Using Big Data Analysis, International Conference on Parallel, Distributed and Network-Based Processing, 2017

Rika ITO, Prediction of Hardware Failure Using S.M.A.R.T. Information, PDP2018, 2018

6. 研究組織

(1) 研究代表者

藤田直行 (FUJITA, Naoyuki)

国立研究開発法人 宇宙航空研究開発機構・

航空技術部門・主幹研究開発員

研究者番号：70358480

(4) 研究協力者

伊藤利佳 (ITO, Rika)