障害予兆検知技術のクラウドデータセンターへの適用

Online Failure Prediction in Cloud Datacenters

● 渡辺幸洋　　● 松本安英

あらまし

クラウドコンピューティングは利用者に利便性をもたらす一方，仮想化により多数の機器を集約しているため，障害が発生すると広範囲のサービスが停止するなど多数の利用者に影響が及ぶ。これを避けるため，障害の兆候を捉え，障害が深刻になる前に能動的に対処するアプローチがある。これまで，システムのログメッセージを分析し，障害の予兆を特定するいくつかの手法が提案されてきた。しかし，これらの手法を実環境に適用する場合，多様なメッセージフォーマットに対応する必要があるほか，システム構成の頻繁な変更により分析結果がすぐに陳腐化して使えなくなるなどの課題があった。

本稿では，富士通研究所が開発した新しい障害予兆検知手法を紹介する。この手法では，メッセージをそのフォーマットによらず文字列の類似度によって自動的に分類し，システム構成が頻繁に変更される環境であっても，障害に関連のあるメッセージパターンを常に再学習する。著者らは，実際のクラウドデータセンターの運用環境で本手法のオンライン評価を行った。評価の結果，最善のケースでは，精度 80%，再現率 90%と高い性能で予兆を検知できることが分かった。

Abstract

Once failures occur in a cloud datacenter accommodating a large number of virtual resources, they tend to spread rapidly and widely, impacting many cloud services and their users. One of the best ways to prevent a failure from spreading in the system is to identify signs of a failure before its occurrence and deal with it proactively before it causes serious problems. Although several approaches have been proposed to predict failures by analyzing past logs of system messages and identifying the relationship between the messages and the failures, it is still difficult to automatically predict the failure for several reasons such as variation of log message formats and frequent changes in their configurations. Based on this understanding, we propose a new failure prediction method that Fujitsu Laboratories has developed. The method automatically learns message patterns as signs of failure by classifying messages by their similarity regardless of their format and re-learning the message patterns in frequently-changed configurations. We evaluated our method in an actual cloud datacenter. The experimental results showed that our approach predicted failures with 80% precision and 90% recall in the best case.

ま え が き

近年のクラウドコンピューティングの普及により，ユーザーは自前のサーバを保有しなくとも，必要な計算機資源を必要な期間だけ調達できるようになった。クラウドコンピューティングはユーザーにとって大きな利便性をもたらす一方，クラウドを支える運用管理では新たな課題も発生している。クラウドデータセンターでは，仮想化基盤上で多数のユーザーが計算機資源を共有している。このような環境では，障害が発生すると，その影響が急速かつ広範囲に及ぶほか，仮想化技術でハードウェアを隠蔽しているために障害対処に時間がかかる。したがって，障害を早期に検出し，深刻になる前に速やかに対処することが重要である。高信頼で低コストな運用を実現するためには，障害が発生してから対処するこれまでの事後対処型のアプローチを，障害の発生前に対処して障害の発生を回避する事前対処型のアプローチに変革することが求められる。

富士通研究所では，障害の予兆を事前に検知して対処を迅速化するために，リアルタイムでメッセージパターンを作成・学習し，障害の兆候を検知する手法を開発した。本稿では，メッセージパターンの学習による障害予兆検知技術を紹介するとともに，実際のクラウドデータセンターにおいてオンラインでメッセージを取得して予兆検知を行い，その性能を評価した結果を紹介する。

障害予兆検知における課題

障害を早期に検出するための手段の一つとして，サービスに影響する障害が発生する前に，システムを構成する機器の振る舞いから障害の発生を予測する障害予兆検知がある。この分野ではこれまでにも様々な手法が提案されており，その多くはシステムを構成する機器が出力したメッセージログを分析し，障害に関連のあるメッセージパターンを抽出するものである。Salfnerらは，隠れセミマルコフモデル（HSMM：Hidden Semi-Markov Model）を用いてログに記録されたメッセージの順序を分析し，障害に関連のあるメッセージシーケンスを特定した。[1] しかし，これらの手法をクラウドデータセンターのような大規模なシステムに適用する場合，次のような課題がある。

（1）多様なメッセージフォーマット

大規模なクラウドデータセンターにおいては，システムは様々なベンダーの多様な機種で構成される傾向がある。様々な構成要素から出力されるメッセージのフォーマットはまちまちで，従来の研究の対象であった高性能計算（HPC）環境のログのように統一されていないため，メッセージの分類が困難になる。

（2）メッセージの順序が厳密に保証されない

大規模システムの運用では，システムを構成する多数の機器からメッセージを収集する。機器間の時刻ずれや，メッセージを収集する際のネットワーク遅延の差などにより，収集・記録されたメッセージの順序は，各メッセージが実際に出力された順序とは異なる場合がある。このため，メッセージ順序を考慮する従来の手法では，障害の予兆をうまく学習できない。

（3）学習結果の陳腐化

クラウドデータセンターでは，機器の入替えやソフトウェアの改版など，常に一部の機器が更新される状況にあるため，予兆検知のために分析した結果が短期間で陳腐化してしまう。このような環境で分析結果を最新に保つためには，障害の予兆をリアルタイムで分析し，分析の結果を速やかに予兆検知に反映し，最新の状況を保つことが重要である。

オンライン障害予兆検知

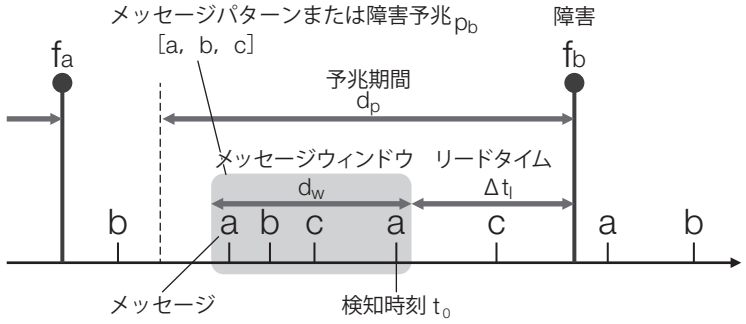
著者らは，前章の課題を解決するために，次に示す手順でリアルタイムにメッセージパターンを作成・学習し，障害の予兆を検知する手法を開発した。なお，本稿での用語は，参考文献（2）をベースとして本手法に合わせて改変したものを用いる（**図-1**）。

（1）メッセージ分類

最初に，**図-2**に示すように，単語を単位として取得したメッセージを分割し，メッセージ辞書の各エントリーと比較する。メッセージに含まれる単語と一致した単語が最も多いエントリーにメッセージを分類する。

（2）メッセージパターン学習

次に，ある時刻における過去数分のメッセージ



障害	：サービスに影響を与える事象
メッセージ	：時刻と本文で構成される，機器が出力したイベント
メッセージウィンドウ	：ある時刻から一定期間前までの時刻幅
メッセージパターン	：メッセージウィンドウ内のメッセージの種類一覧
障害予兆	：障害と強い共起関係を持つメッセージパターン
予兆期間	：障害の予兆があるとみなす，障害発生前の期間
リードタイム	：障害予兆を検知してから障害が発生するまでの猶予期間

図-1　用語定義

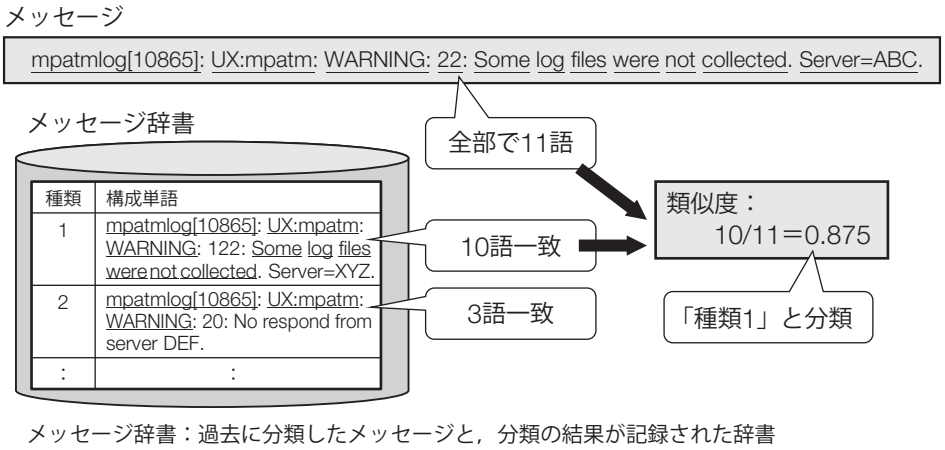


図-2　メッセージ分類

の種類の集合を「メッセージパターン」とし，メッセージパターンと障害の関連をベイズ確率を用いて学習する｛**図-3**（a)｝。メッセージパターンPが障害Tの発生前一定期間に発生する確率を，次の式で求める。

$$(\text{障害Tの発生確率}) = \frac{\text{Tの予兆期間におけるPの観測回数}}{\text{全ての期間におけるPの観測回数}}$$

求めた確率をメッセージパターン辞書に記録すると同時に，メッセージパターンPと障害Tとの時刻差（リードタイム）を求めて記録する。

(3) オンライン障害予兆検知

システムから出力されるメッセージを分類してパターンを作成し，学習した結果と照合することで，各種障害の発生確率をリアルタイムで評価する。障害の発生確率が設定したしきい値より高い場合，障害の予兆として運用管理者に通報を行う｛図-3（b)｝。

本手法は，次の三つの特徴を持つ。

(1) フォーマットに依存しないメッセージ分類

クラウド環境では様々なフォーマットのメッセージが混在するが，本手法ではメッセージを構成する単語の一致数で分類するため，様々なフォーマットのメッセージを一様に扱うことができる。更に，メッセージの意味を解釈せずに機械的に分類するため，メッセージ辞書を人が定義する必要がない。

(2) 順序に依存しないメッセージパターン作成

クラウド環境ではメッセージの順序が必ずしも

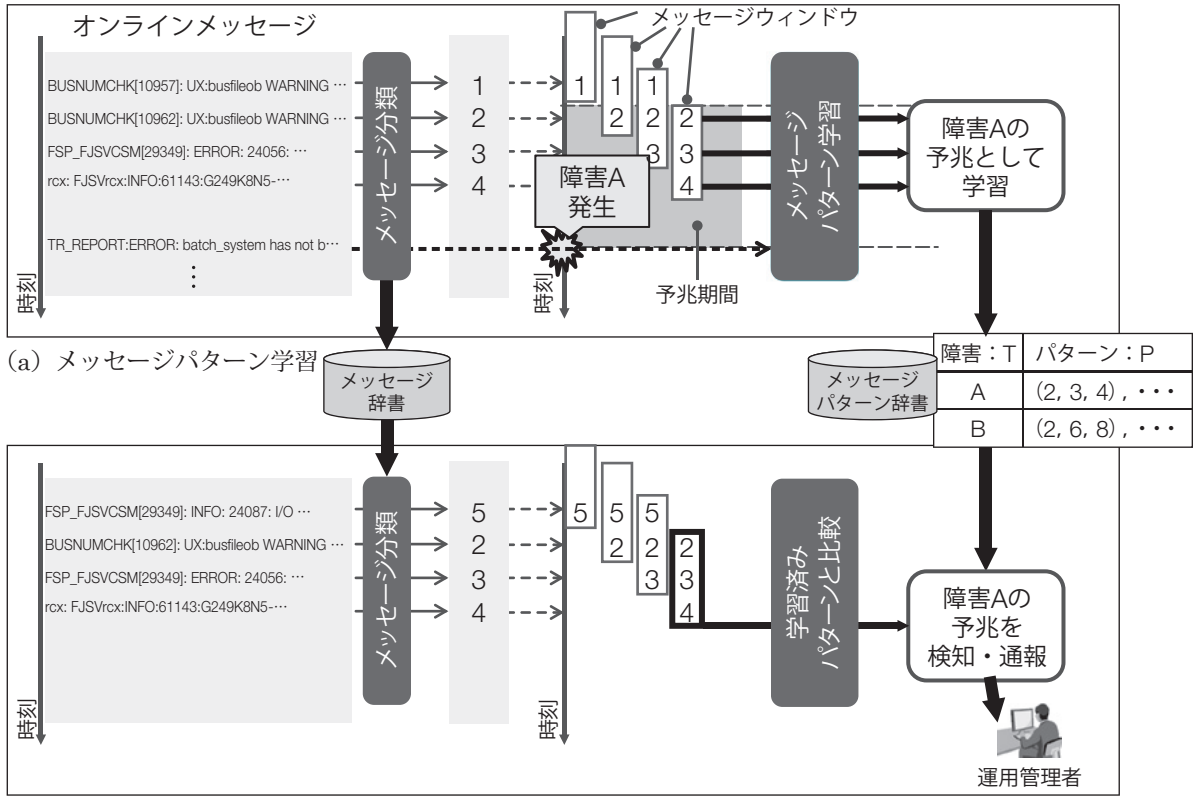


図-3　予兆学習・検知

保証されないが，本手法では，メッセージパターンを作成する際，順序を考慮せずにメッセージ種類の集合として扱うため，メッセージの順序が多少入れ替わっても学習結果に影響しない。

(3) リアルタイムでのメッセージパターン学習

本手法では，入力されたメッセージをリアルタイムで分類してメッセージパターンを作成し，障害予兆の学習と検知を行う。一般的なバッチ処理によるパターン学習と異なり，システムに構成変更があった場合でも即座にメッセージパターン辞書を更新でき，常に最新の学習結果を用いて障害予兆を検知できる。

クラウド環境での評価

本手法の性能を評価するために，商用のクラウドデータセンターでオンライン評価を行った。

(1) 対象システム

このシステムは数百台の物理サーバで構成され，1万以上のVM（Virtual Machine）を提供する。この環境で90日間，メッセージログを採取して障害予兆の通報を試行した。この期間中，約945万件のメッセージが発生し，509種類に分類された。また，障害は112件発生し，20種類に分類された。発生した障害の例を**表-1**に示す。

表-1　発生した障害の例

障害種別	内容	発生回数
a	バッチシステム障害 #1	21
b	プロセス稼働率異常	10
c	しきい値異常	10
d	ストレージノード停止	21
e	バッチシステム障害 #2	7
f	バッチシステム障害 #3	6
g	予期せぬノード再起動	5
h	ディスクコピー障害	7
そのほか（12種類）		25
合計		112

(2) 実装

オンラインでの障害予兆検知システムを試作し，実際の商用クラウドデータセンターの運用環境内の

VMに設置した（**図-4**）。試作にはJavaとMySQLを用いた。VMの性能はCPUがXeon 2.0 GHz，メモリが3.4 Gバイトに相当する。

（3）評価指標

評価を行う際の指標には，障害予兆の分野の研究で一般的な次の三つを選んだ。

・精度：全ての予兆検知の件数に対する，予兆検知後に実際に障害が発生した予兆検知の件数の比
・再現率：全ての障害の件数に対する，予兆検知によって検知された障害の件数の比
・F値：精度と再現率の調和平均

一般に，精度と再現率はトレードオフとなる傾向がある。もしも障害予兆の見逃しを避けようとすると，多数の誤検知を引き起こす。システムの運用管理において，誤検知が多発すると，管理者の作業時間が増加してコストが上昇するため，実運用においては精度を許容可能な範囲に収める必要がある。

（4）結果

表-1に示すそれぞれの種類の障害について，しきい値を0.99とした場合の予兆検知性能を**図-5**に示す。最も良い値を示す例をとると，種別aの障害についての予兆検知の結果は，精度が80％（24/30），再現率が90％（19/21），F値が0.85であった。

（5）考察（障害の性質に起因する予兆検知性能の違い）

評価を通して，精度と再現率は，障害の種類によって大きく異なることが分かった。この原因を探るため，学習したメッセージパターンのリードタイムと障害予兆の正解率の関係に着目して分析を行った。分析の結果，障害が次に示す三つのカテゴリに分類できることが分かった。

・逓減型

リードタイムが長くなるにつれ，精度が低下する。この傾向を示す障害には，プロセスのハングなど，小さな異常が積み重なった結果発生すると思われるものが多かった。これらの障害では，対処に必要な時間と精度のバランスを考慮する必要がある。例えば前述のプロセスハング障害では，リードタイムが0-10分，10-20分，20-30分のとき，精度はそれぞれ77％，52％，17％であった。また，この障害を回避するための対処（プロセス再起動）には10分程度を要する。この場合，リードタイムが20分以上では予兆の正解率が低く，10分未満では対処が間に合わないため，実際の運用ではリードタイムが10分以上20分未満の予兆のみ通報する

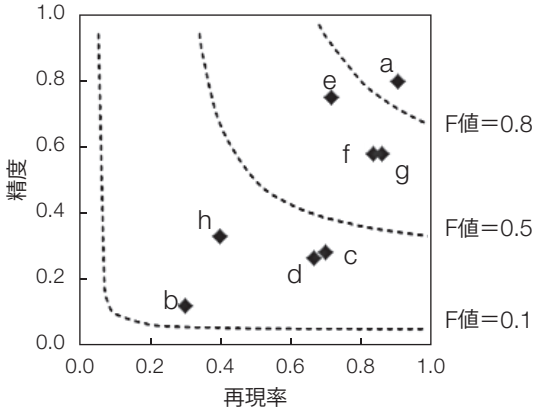


図-5　障害別の予兆検知性能

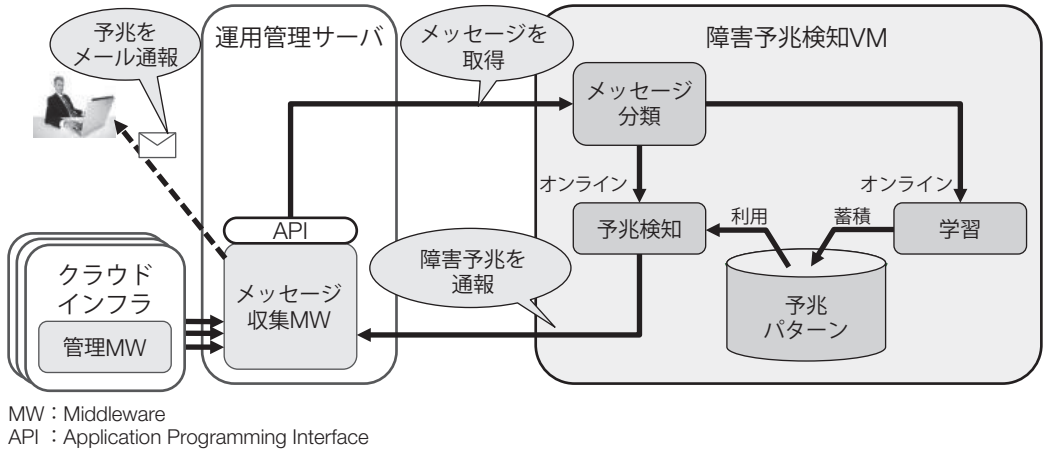


図-4　障害予兆検知の試行

ことで，「正解する可能性が高く，対処が間に合う予兆」のみを通知できる。

・長期型

リードタイムが長くなっても，正解する予兆検知が一定数存在する。この傾向を示す障害には，ストレージ装置の停止など，ハードウェアの異常を伴うものが多かった。これらの障害については，精度は50％程度と低いものの，障害の1時間以上前に障害予兆を検知できていることから，この中で，重大かつ回避措置に時間がかかるものについては，積極的に予兆を通報して対処することで，重大障害の発生件数を削減できると考える。

・直前型

リードタイムが10分未満とごく短い範囲にのみ，正解する予兆が存在する。この傾向を示す障害には，VMの移動（マイグレーション）の失敗など，人やプログラムにより機器に対して何らかの操作を行った直後に発生するものが多い。これらの障害については，障害予兆検知ではなく，操作手順の事前検証などの技術により障害の発生を予防するなどの対策が有効だと考える。

実際のクラウドデータセンターの運用に障害予兆検知を適用する際は，このような障害の誤検知を避けるために，障害の特性により予兆の通報を抑止する機能の実装が必要である。

む　す　び

本稿では，深刻な障害が発生する前に対処するために富士通研究所が開発した障害予兆検知手法と，実際のクラウドデータセンターにおいてオンラインで評価した結果について紹介した。評価の結果，メッセージの集合と障害の関連を算出する本手法は，従来の手法の適用が困難な大規模なクラウドコンピューティング環境でも障害予兆を検知できることが示された。また，障害の性質と予兆検知の精度の特性を3種類に分類し，特性を考慮することで効率的に予兆検知に対処できることが示唆された。

本手法では，メッセージパターンと障害の間にあるメカニズムに踏み込まず，統計的に関連を抽出する。本手法をクラウドインフラの障害対処プロセスに統合するとともに，ほかの分析手法や構成情報，インシデント記録と連携させて利用することで，クラウドデータセンターのインフラの運用管理を改善できると考える。

参考文献

(1) F. Salfner et al.：Using Hidden Semi-Markov Models for Effective Online Failure Prediction. In Proceedings of the 26th IEEE International Symposium on Reliable Distributed Systems. Washington, DC, USA, IEEE Computer Society, p.161-174（2007）.

(2) F. Salfner et al.：A survey of online failure prediction methods. ACM Comput. Surv. New York, NY, USA, March, 2010. Vol.42, p.10:1-10:42. ACM（2010）.

著者紹介

渡辺幸洋（わたなべ　ゆきひろ）
システムソフトウェア研究所システムマネジメント研究部 所属
現在，クラウドコンピューティング環境の運用管理技術の研究に従事。

松本安英（まつもと　やすひで）
システムソフトウェア研究所システムマネジメント研究部 所属
現在，運用保守についての研究開発およびクラウド関連技術の標準化についての業務に従事。