

モバイルクラウドにおける最適データ分割と適応性能を有した高信頼化技術の開発

研究代表者 田村 慶 信 山口大学 大学院理工学研究科 准教授
共同研究者 山田 茂 鳥取大学 大学院工学研究科 教授

1 はじめに

データの一元管理, 低コスト, 保守・運用が容易といった観点から, OpenStack や Eucalyptus などのオープンソースソフトウェア (open source software, 以下 OSS と略す) を利用したクラウド環境の構築に注目が集まっている. ビッグデータを扱うクラウドコンピューティング環境では, 主に, ビッグデータを支える Hadoop や NoSQL に代表されるデータベースソフトウェアと, Eucalyptus や OpenStack に代表されるクラウドソフトウェアにより運用されている. 最近では, 2014 年 10 月に OpenStack の Juno バージョンがリリースされ, OpenStack のフレームワークの中に Hadoop が組み込まれた. Hadoop は大量データの高速度処理に適したデータベースソフトウェアであり, ビッグデータを扱う多くのシステムで利用されている. このように, クラウドコンピューティング環境においても, 比較的規模の大きなソフトウェアとの連携処理がネットワーク経由で頻繁に行われつつシステムが運用される事例が多くなってきている.

クラウド環境に対する最近の研究動向としては, モバイルクラウド, サービス形態, 性能評価等を対象とした文献はいくつか提案されているが [1, 2], ビッグデータによるデータ肥大化に伴うクラウドを対象とした信頼性評価に関する研究は行われていない. 従来から, ソフトウェア製品の開発プロセスにおけるテスト進捗管理や出荷品質の把握のための信頼性評価を行うアプローチとして, ソフトウェア故障の発生現象を不確定事象として捉えて確率・統計論的に取り扱う方法がとられている. その 1 つが, ソフトウェア信頼性モデル (Software Reliability Model, 以下 SRM と略す) である [3]. これまでに数百におよぶソフトウェア信頼性モデルが提案されてきた [3-5]. しかしながら, 既存のソフトウェア信頼性モデルの多くは大規模ソフトウェア間の通信環境に伴うフォールト発生事象が考慮されていない. クラウド上のビッグデータを想定したソフトウェアに対して既存のソフトウェア信頼性モデルを適用することは可能であっても, クラウドとビッグデータの信頼性評価に関する新たな知見を得ることはできない. 特に, モバイルクラウドの運用環境では, ネットワークに接続した状態で常時運用が行われている. こうしたモバイルクラウドのネットワーク環境から受けるソフトウェア信頼性への影響を考慮することは重要である. 最近では, モバイルクラウドの特徴から, ビッグデータを扱うデータベースソフトウェアがクラウドソフトウェア内に統合されつつある. こうした状況から, データベース上においても, クラウドにアクセスし管理を行うためのクラウドベースデータと, デジタル画像や音楽データなどのコンテンツデータが蓄積されるようになっている. これらのコンテンツデータはストレージ容量の肥大化を招くだけでなく, ソフトウェアシステムの複雑化や運用管理の煩雑化を招く原因の一つとなっている.

本研究では, 信頼性の観点からデータベース上における最適なデータ分割状態を考慮するために, クラウドソフトウェアとデータベースソフトウェアからのフォールト発生状況を分析した. また, モバイルクラウド環境全体の信頼性を評価するためにジャンプ拡散過程モデルを適用した. これにより, 保存されたデータベース上におけるコンテンツデータとクラウドベースデータとの関係性を信頼性の観点から分析することが可能となる. さらに, データベースソフトウェアとクラウドソフトウェアとの最適なデータ分割を考慮した提案モデルに基づく最適メンテナンス問題について議論するとともに, 提案手法をアプリケーションソフトウェアとして実装してきた.

2 ニューラルネットワークに基づくソフトウェア構成比率の推定

モバイルクラウド環境下におけるソフトウェア構成比率は, クラウドのソフトウェア構成, 利用形態, ユーザ数, およびハードウェア構成要素など, 様々な要因により影響を受ける. こうした種々の環境要因をパラメータとして考慮し, 物理的な観点からモデル化することは可能であっても, それを実際のソフトウェア運用環境に適用することは困難であると思われる. ここでは, モバイルクラウドを構成するクラウドソフト

ウェアとデータベースソフトウェアにおけるソフトウェア構成比率を信頼性の観点から評価するために、ノンパラメトリックな手法であるニューラルネットワークに基づく時系列分析手法を利用する。

本研究では、簡単のために3層ニューラルネットワークを適用する。このとき、クラウドソフトウェアの一部としてビッグデータを扱うことが可能なデータベースソフトウェアが組み込まれているものと仮定し、クラウドソフトウェアに対するデータベースソフトウェアの累積フォールト発見数の比率を入力データとして適用する。ここで、教師パターンには、実際に観測されたクラウドソフトウェアに対するデータベースソフトウェアの累積フォールト発見数の比率の正規化された値を採用する。すなわち、時刻 t までにおける実際に観測された累積フォールト発見数の比率に基づいて、各時点における累積フォールト発見数の比率の結合状態の特徴をニューラルネットワークの結合係数に蓄積させ、時刻 $t+1$ における累積フォールト発見数の比率の推定・予測が可能なモデルを考える。

3 ソフトウェアフォールトデータに対するクラスタ分析

ビッグデータを扱うクラウドコンピューティング環境では、主に、ビッグデータを支える Hadoop や NoSQL に代表されるデータベースソフトウェアと、Eucalyptus や OpenStack に代表されるクラウドソフトウェアにより運用されている。最近では、2014年10月に OpenStack の Juno バージョンがリリースされ、OpenStack のフレームワークの中に Hadoop が組み込まれた。Hadoop は大量データの高速処理に適したデータベースソフトウェアであり、ビッグデータを扱う多くのシステムで利用されている。このように、クラウドコンピューティング環境においても、比較的規模の大きなソフトウェアとの連携処理がネットワーク経由で頻繁に行われつつシステムが運用される事例が多くなってきている。特に、クラウドコンピューティングの運用環境における信頼性評価について考えた場合、クラウドソフトウェアに起因するソフトウェア故障だけではなく、データベースソフトウェアに起因するソフトウェア故障も同時に考慮する必要がある。本研究では、ビッグデータを扱うクラウドコンピューティングの運用上における信頼性を評価するために、ビッグデータソフトウェア、クラウドソフトウェア、HTTP サーバソフトウェアにおける3種類のソフトウェア故障の関係性をフォールトデータの観点から分析する。

本研究では、ビッグデータソフトウェア、クラウドソフトウェア、およびクラウドへの接続窓口となる HTTP サーバソフトウェアにおける3種類のソフトウェア故障の関係性をフォールトデータの観点からクラスタ分析[6-8]により評価する。クラスタリング手法として、階層的クラスタ分析と非階層的クラスタ分析に大別される。特に、階層的クラスタ分析の場合、個体数が多い場合は計算量も大きくなる傾向がある。近年、一般的に信頼性評価のために取り扱われるフォールトデータは、バグトラッキングシステムのようなデータベースにより管理されていることが多く、そのデータ量も比較的大きくなる傾向があることから、本研究では離散型数値データを分析するために、非階層的クラスタリング手法として広く知られている k -means 法を適用する。 k -means 法は、アルゴリズムが非常にシンプルであり拡張性が高い。

ビッグデータを扱うクラウドコンピューティング環境を想定したフォールトデータに対してクラスタリングを行うにあたり、データベースソフトウェア、クラウドソフトウェア、およびクラウドへの接続窓口となる HTTP サーバソフトウェアから採取されたフォールトデータを取り扱う。

4 ジャンプ拡散過程モデル

モバイルクラウド環境の変化に伴う運用形態の特徴を考慮するために、フォールト発見率 $b(t)$ に不規則性を導入するとともに、クラウドソフトウェアとデータベースソフトウェア間の相互作用に伴う外的な影響をジャンプ拡散項により表現する[9-12]。このとき、任意の時刻 t におけるソフトウェア内の残存フォールト数 $N(t)$ は、以下の式により与えられる。

$$N(t) = a(1+bt) \exp \left\{ -bt - \sigma W(t) - \sum_{i=1}^{M_i(\lambda)} \log V_i \right\}. \quad (1)$$

ここで、 a はソフトウェア内に潜在する総フォールト数、 b はフォールト1個当りのフォールト発見率、 $\sigma (> 0)$ は定数パラメータを表す。また、 $W(t)$ は Wiener 過程であり、形式的には白色雑音の時間積分で定義されるものである。さらに、 $M_i(\lambda)$ は $W(t)$ とは独立な強度パラメータ λ をもつポアソン過程であり、時

刻 t までにジャンプが発生した回数を表す。 λ はジャンプ事象が生じる確率的な頻度であり、 V_i は i 回目のジャンプ幅を表す独立な確率変数を意味する。

5 アプリケーションの開発

5-1 要求仕様定義

開発されたツールの要求仕様の一例を以下に示す。

- 信頼性評価に使用するデータは、実際の運用段階において各主要ソフトウェアから採取された実測データを用いる。
- 採取された実測データに基づいて信頼性評価を行い、各推定結果をグラフで表示する。
- 各主要ソフトウェアに対する相互作用を考慮するために3次元フォールトクラスタリングを適用する。
- クラウド環境全体の信頼性を評価するためにジャンプ拡散過程モデルを適用する。
- 採用されたジャンプ拡散過程モデルに含まれる未知パラメータを推定するために、最尤法を適用する。
- パラメータ推定の際には、統計言語 R と動的リンクによって推定結果を取得する。
- 主要ソフトウェアコンポーネント間の相互作用を把握するために、3D フォールトクラスタリングによる推定結果を用いる。
- 信頼性評価尺度として、および残存フォールト数のサンプルパスを用いる。
- ツールの操作には GUI を使用し、マウスを用いて行う。
- ツールの開発言語に Flex[13] および ActionScript を使用する。
- グラフの描画には Flex のチャートコンポーネントおよび Papervision3D[14] を使用する。

5-1 実行手順

信頼性評価のための実行手順の一例を以下に示す。

- クラウドの運用段階において、クラウド環境を構成する主要ソフトウェアコンポーネントからフォールト発見数データを採取する。
- フォールトデータと未知パラメータの初期値をアプリケーションのメイン画面上で入力する。
- 最尤法および遺伝的アルゴリズムにより未知パラメータを推定する。
- 信頼性評価尺度をグラフ表示する。

6 アプリケーションの実行例

オープンソースソフトウェアとして開発および公開されている、クラウドソフトウェアである OpenStack[15]、データベースソフトウェアである Hadoop[16]、および HTTP サーバである Apache HTTP サーバ[17]におけるバグトラッキングシステム上に登録されたフォールトデータを適用したアプリケーションの実行例を示す。

開発されたニューラルネットワークに基づくアプリケーションソフトウェアの実行例を以下に示す。まず、図 1 は、開発されたアプリケーションソフトウェアのメイン画面を示す。また、図 2 は、アプリケーションにより実行されたソフトウェア構成比率の推定結果である。さらに、ジャンプ拡散過程モデルによる推定された残存フォールト数のサンプルパスを図 3 に示す。

同様に、フォールトデータクラスタリングによるアプローチに基づくアプリケーションソフトウェアの実行例を示す。図 4 は、開発されたアプリケーションソフトウェアのメイン画面を示す。また、図 5 は、アプリケーションにより推定されたクラスタリングの推定結果である。さらに、ジャンプ拡散過程モデルによる推定された残存フォールト数のサンプルパスを図 6 に示す。

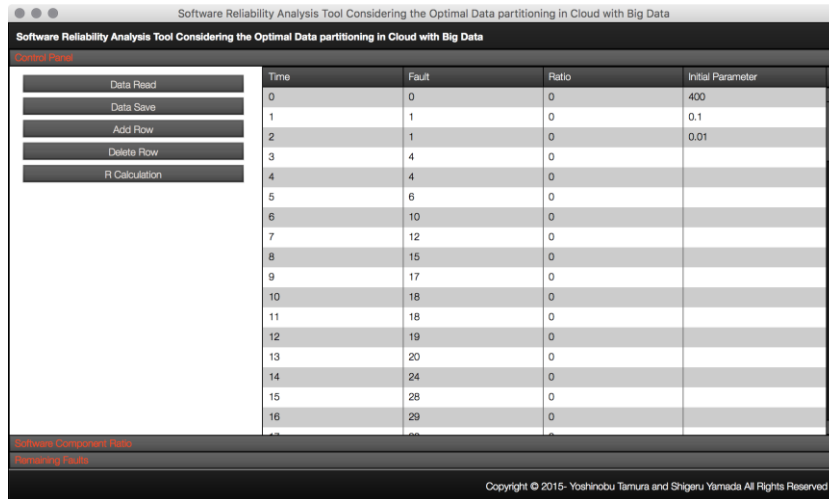


図 1 : アプリケーションのメイン画面.

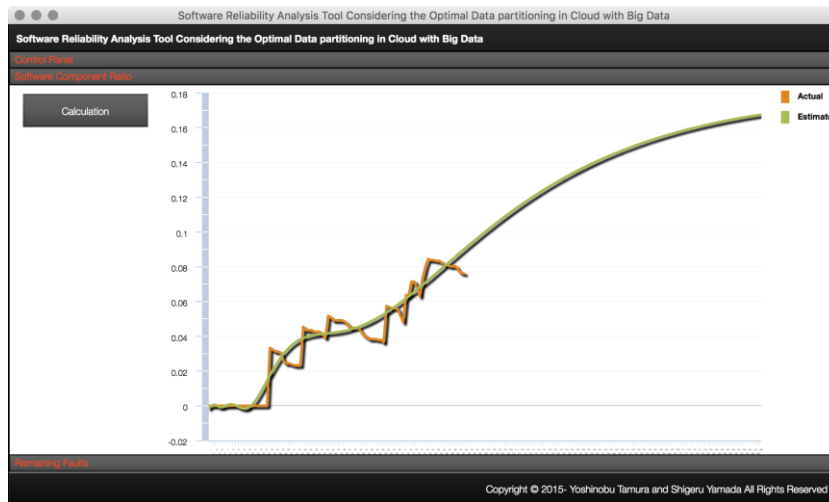


図 2 : ソフトウェア構成比率の推定結果.

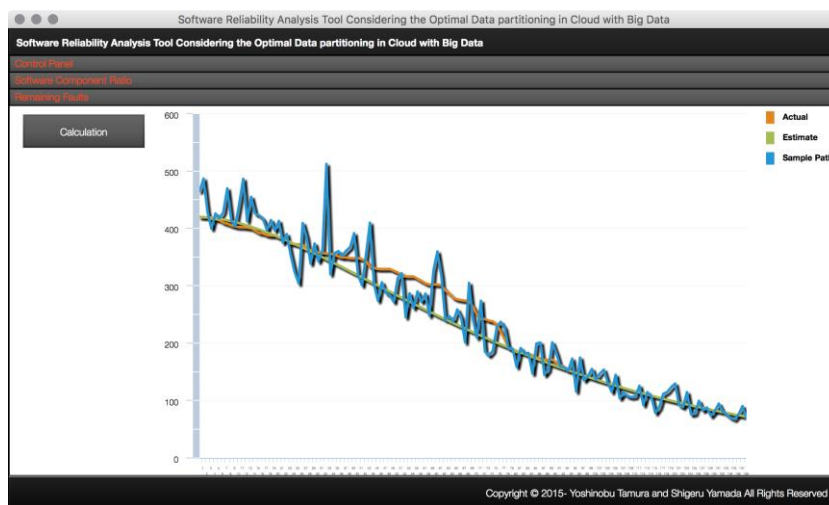


図 3 : 推定された残存フォールト数のサンプルパス.

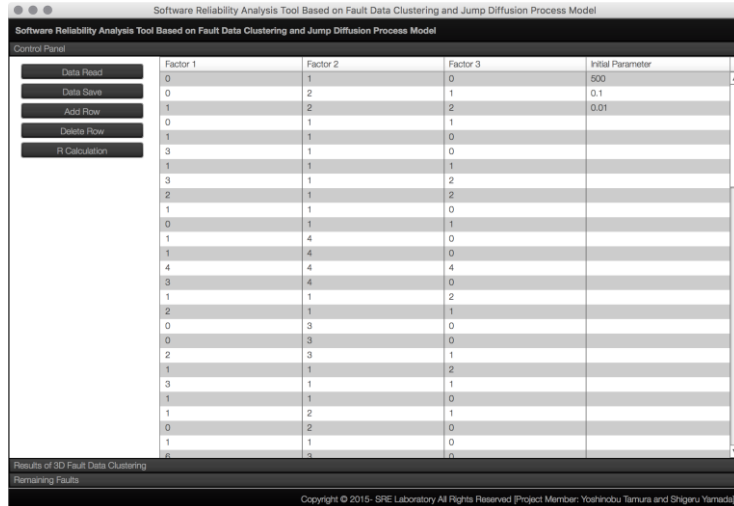


図 4 : 3D アプリケーションのメイン画面.

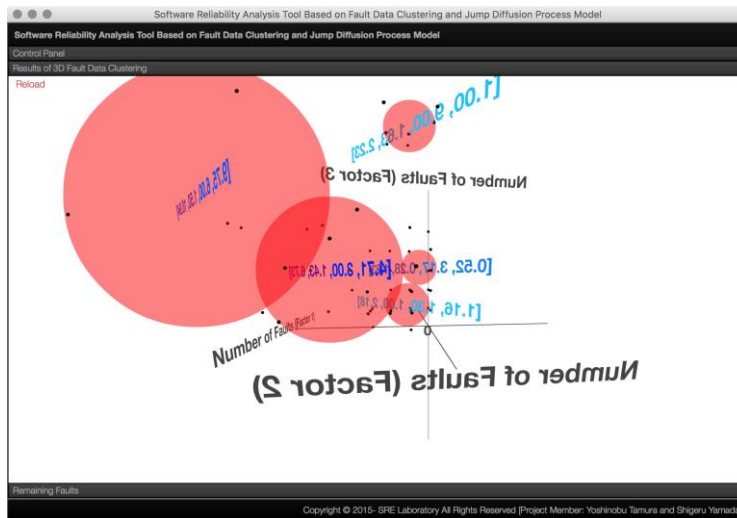


図 5 : フォールトデータクラスタリングによる分析結果.



図 6 : ジャンプ拡散過程モデルに基づく残存フォールト数の推定結果.

7 おわりに

本研究では、モバイルクラウドの実利用環境を想定し、信頼性の観点からデータベース上における最適なデータ分割を判断するために、クラウドソフトウェアとデータベースソフトウェアからのフォールト発生状況に基づき、ソフトウェア構成比率をニューラルネットワーク手法により分析した。さらに、モバイルクラウド環境全体の信頼性を評価するためにジャンプ拡散過程モデルを適用した。また、クラウドコンピューティングの最適メンテナンス時刻を推定するために、従来のコスト評価基準に基づくソフトウェアの最適リリース問題を応用することにより、データベースソフトウェアとクラウドソフトウェアとの通信環境を考慮した提案モデルに基づく最適メンテナンス問題について議論した。

実際のクラウドOSSのソフトウェアフォールト発見数データを適用し、ジャンプ拡散過程モデルに対する数値例を示すことにより、保存されたデータベース上におけるコンテンツデータとクラウドベースデータとの関係性を信頼性の観点から分析できることを示した。特に、デジタル画像や音楽データなどのコンテンツデータを扱うデータベースソフトウェアに対する管理工数の配分の見直し時期を推定するようなソフトウェア運用管理上における評価指標を示した。また、モバイルクラウド環境全体に対する信頼性評価尺度として、ジャンプ拡散過程モデルに基づく信頼性評価法を提案するとともに、最適メンテナンス時刻の推定法について議論した。さらに、提案手法に基づくアプリケーションソフトウェアを実装し、開発されたアプリケーションソフトウェアの実行例を示した。これにより、クラウドベースデータとコンテンツデータとの最適データ配分状態を信頼性の観点から考慮した総合的な信頼性評価が可能となるものと考えられる。

【参考文献】

- [1] H. Suo, Z. Liu, J. Wan, and K. Zhou, Security and privacy in mobile cloud computing, Proceedings of 9th International Wireless Communications and Mobile Computing Conference, pp. 655-659, 2013.
- [2] A. Khalifa and M. Eltoweissy, Collaborative autonomic resource management system for mobile cloud computing, Proceedings of Fourth International Conference on Cloud Computing, GRIDs, and Virtualization, pp. 115-121, 2013.
- [3] S. Yamada, Software Reliability Modeling: Fundamentals and Applications, Springer-Verlag, Tokyo/Heidelberg, 2013.
- [4] M.R. Lyu, ed., Handbook of Software Reliability Engineering, IEEE Computer Society Press, Los Alamitos, CA, 1996.
- [5] P.K. Kapur, H. Pham, A. Gupta, and P.C. Jha, Software Reliability Assessment with OR Applications, Springer-Verlag, London, 2011.
- [6] P. Bellot and M. El-Beze, A Clustering Method for Information Retrieval, Technical Report IR-0199, Laboratoire d'Informatique d'Avignon, France, 1999.
- [7] P.S. Bradley, K.P. Bennett, and A. Demiriz, Constrained k -means Clustering, Technical Report MSR-TR-2000-65, Microsoft Research, Redmond, WA, 2000.
- [8] A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Prentice Hall, New Jersey, 1988.
- [9] R.C. Merton, Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics, vol. 3, pp. 125-144, 1976.
- [10] L. Arnold, Stochastic Differential Equations-Theory and Applications, John Wiley & Sons, New York, 1974.
- [11] E. Wong, Stochastic Processes in Information and Systems, McGraw-Hill, New York, 1971.
- [12] S. Yamada, M. Kimura, H. Tanaka, and S. Osaki, Software reliability measurement and assessment with stochastic differential equations, IEICE Transactions on Fundamentals, vol. E77-A, no. 1, pp. 109-116, 1994.
- [13] The Apache Software Foundation, Apache Flex, <http://flex.apache.org/>
- [14] papervision3d, Open Source realtime 3D engine for Flash, <https://code.google.com/p/papervision3d/>

[15] The OpenStack project, OpenStack, <http://www.openstack.org/>

[16] The Apache Software Foundation, Apache Hadoop, <http://hadoop.apache.org/>

[17] The Apache Software Foundation, Apache HTTP Server, <https://httpd.apache.org/>

〈発表資料〉

題名	掲載誌・学会名等	発表年月
Software reliability analysis considering the fault detection trends for big data on cloud computing	Industrial Engineering Management Science and Applications 2015, Lecture Notes in Electrical Engineering 349, Springer-Verlag Berlin Heidelberg, pp. 1021-1030	May 2015
Reliability analysis based on a jump diffusion model with two Wiener processes for cloud computing with big data	Entropy, Vol. 17, No. 7, Multidisciplinary Digital Publishing Institute, Switzerland, pp. 4533-4546	June 2015
Three dimensional Wiener processes model and optimal software maintenance planning	Proceedings of the Ninth International Conference on Mathematical Methods in Reliability, Tokyo, Japan, pp. 863-870	June 1-4, 2015
ビッグデータとクラウドを想定したフォールトデータクラスタリングに基づくソフトウェア信頼性評価法	電子情報通信学会 技術研究報告 [ソフトウェアサイエンス], 札幌市教育文化会館, Vol. 115, No. 153, pp. 119-124	2015年7月22-24日
Optimal maintenance problem based on hazard rate model for mobile clouds	Proceedings of the 21st ISSAT International Conference on Reliability and Quality in Design, Philadelphia, Pennsylvania, USA, pp. 84-88	August 6-8, 2015
Reliability assessment based on jump diffusion model for cloud computing under the influence of big data	Proceedings of the 21st ISSAT International Conference on Reliability and Quality in Design, Philadelphia, Pennsylvania, USA, pp. 99-103	August 6-8, 2015
Hazard rate model considering various factors for mobile clouds	Proceedings of the 21st ISSAT International Conference on Reliability and Quality in Design, Philadelphia, Pennsylvania, USA, pp. 203-207	August 6-8, 2015
Software reliability assessment tool based on fault data clustering and hazard rate model considering cloud computing with big data	Proceedings of the IEEE International Conference on Reliability, Infocom Technology and Optimization, Amity University, Uttar Pradesh, Noida, India	September 2-4, 2015
モバイルクラウドに対するフォールトデータクラスタリングとハザードレートモデルに基づく信頼性評価法	日本オペレーションズ・リサーチ学会 平成27年秋季研究発表会アブストラクト集, 九州工業大学 戸畑キャンパス, pp. 142-143	2015年9月10-11日
Cost optimization based on decision-making and reliability modeling for big data on cloud computing	An International Journal of Communications in Dependability and Quality Management, Vol. 18, No. 4, pp. 5-19	November 2015

Mobile application for reliability assessment based on three noisy models for cloud computing	Proceedings of the 2nd East Asia Workshop on Industrial Engineering, Seoul, Korea, pp. 104-109	November 6-7, 2015
モバイル OSS の運用環境特性を考慮した信頼性評価のためのモバイルアプリケーション	第17回 IEEE 広島支部 学生シンポジウム論文集, 岡山大学, CD-ROM (A-50)	2015 年 11 月 21-22 日
モバイルクラウドに対する変動要因を考慮したハザードレートモデルに基づく信頼性評価法	第17回 IEEE 広島支部 学生シンポジウム論文集, 岡山大学, CD-ROM (A-53)	2015 年 11 月 21-22 日
Software reliability model considering the optimal data partitioning for a mobile cloud	Proceedings of the 8th Japan-Korea Software Management Symposium, Chungnam, Korea, pp. 30-42	November 27, 2015
A method of reliability assessment based on hazard rate by clustering approach for cloud computing with big data	Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, CD-ROM (Reliability & Maintenance Engineering 2)	December 6-9, 2015
A method of reliability assessment based on neural network and fault data clustering for cloud with big data	Proceedings of the 2nd International Conference on Information Science and Security, Seoul, Korea, pp. 181-184	December 14-16, 2015
Reliability computing and management considering the network traffic for a cloud computing	Annals of Operations Research (Article first published online: DOI: 10.1007/s10479-016-2140-5, 15 February 2016), Springer US, Vol. 236, No. 2, pp. 1-14	January 2016
既存 SRGM に基づく最適リリース時刻推定のためのモバイルアプリケーション	IEEE Consumer Electronics Society West Japan Joint Chapter 研究会, 海峡メッセ下関, pp. 8-13	2016 年 1 月 22 日