

クラウド OSS に対するジャンプ拡散モデルに基づく 最適メンテナンス時刻の推定に関する一考察

山口大学大学院・理工学研究科 田村 慶信 (Yoshinobu Tamura) [†]

[†]Graduate School of Science and Engineering, Yamaguchi University

鳥取大学大学院・工学研究科 山田 茂 (Shigeru Yamada) [‡]

[‡]Graduate School of Engineering, Tottori University

1 はじめに

現在, OS, サーバ, およびアプリケーションレベルまで, 様々な種類のオープンソースソフトウェア (open source software, 以下 OSS と略す) が世界中で開発・公開されている. 近年, 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, オープンソースを利用したクラウド環境の構築に注目が集まっている. しかしながら, ソフトウェアの設計図にあたるソースコードが世界中に公開されているため, 最近のパブリッククラウドに対するクラッキングによる情報事故のように, 悪意のあるサイト攻撃や情報流出の標的になり易く, なかなか導入に踏み切れないのが現状である.

今後ますますクラウド環境採用の動きが活発化するものと考えられ, 低コスト・短納期にクラウド環境を構築できるオープンソースを利用したクラウド環境が注目されている. しかしながら, 「安く早く作れる」というメリットだけが優先してしまい, そのセキュリティや信頼性の問題に多くの企業が悩まされているという現状がある. こうした OSS を利用したクラウド基盤システムの信頼性を定量的に評価する手法は未だ提案されておらず, 職人的・試行錯誤的に運用が行われているのが現状である. 特に, 大規模システムにおいては, ひとたび障害が発生すると個人情報の漏洩だけではなく多大な財産の損失を招くものが多く, セキュリティ・信頼性の評価に関しては重要な要因の 1 つである. クラウド環境に対する最近の研究動向としては, ハードウェア, 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 [1, 2]. しかしながら, そのほとんどがハードウェアやサービス形態の事例研究, データストレージ技術などの性能評価に関するものであり, OSS を利用したクラウド基盤ソフトウェアに対する動的解析に基づく信頼性評価に関する研究は行われていないのが現状である.

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, ソフトウェア信頼度成長モデル (Software Reliability Growth Model, 以下 SRGM と略す) である [3].

本論文では, こうしたクラウド OSS の運用段階を対象とした信頼性評価のためのジャンプ項を含む確率微分方程式モデルを構築する. また, 提案されたモデルからクラウドの利用環境を想定したディペンダビリティ評価尺度を導出する. 特に, 2 種類のジャンプ拡散過程を適用することにより, 提案モデルの適用可能性について考察する. さらに, 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に対する数値例を示す.

2 クラウド OSS に対する信頼性評価のための確率微分方程式モデル

まず, 時刻 $t = 0$ で OSS の運用が開始され, 任意の時刻 t における検出フォールト数 $\{N(t), t \geq 0\}$ は以下の常微分方程式によって記述されるものと仮定する.

$$\frac{dN(t)}{dt} = b(t)\{D(t) - N(t)\}. \quad (1)$$

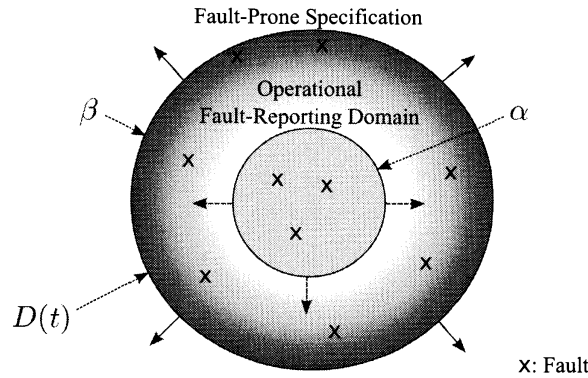


図 1： 要求仕様の変化とフォールト報告領域の概念図。

ここで、 $b(t)(>0)$ は時刻 t におけるフォールト発見率を、 $D(t)$ は要求仕様の変化を考慮した場合における OSS 内に潜在する総フォールト数を示す。

$$D(t) = \alpha e^{-\beta t}. \quad (2)$$

ここで、 α は OSS に潜在するフォールト数を、 β は要求仕様の変更率を表す。本論文では、図 1 に示すように、OSS の要求仕様は運用時刻 t に伴い指数関数的に増加または減少するものと仮定する [4]。また、クラウド OSS の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、式 (2) は、

$$\frac{dN(t)}{dt} = \{b(t) + \sigma\gamma(t)c(t)\} \{D(t) - N(t)\}, \quad (3)$$

となる。ここで、 $\sigma (>0)$ は定数パラメータ、 $\gamma(t)$ は解過程の Markov 性を保証するために標準化された Gauss 型白色雑音、 $c(t)$ は運用環境がクラウド OSS に与える影響度合いを表す。さらに、クラウドの運用段階におけるフォールト発見事象が、ログインするユーザ数やサービスアプリケーション数などの増減により不規則に変動するものと仮定し、ジャンプ項を導入する。式 (3) を、以下の Itô 型の確率微分方程式 [5] に拡張して考える。

$$dN(t) = \left\{ b(t) + \frac{1}{2}\sigma^2 c(t)^2 \right\} \{D(t) - N(t)\}dt + \sigma \{D(t) - N(t)\}dW(t) + d \left(\sum_{i=1}^{M_t(\lambda)} (V_i - 1) \right). \quad (4)$$

ここで、 $M_t(\lambda)$ 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る。また、 V_i は i 回目のジャンプ幅を表す独立な確率変数である。式 (4) の確率微分方程式を Itô の公式を用いて変換すると、

$$N(t) = D(t) \left[1 - \exp \left\{ - \int_0^t b(s)ds - \sigma c(t)W(t) + \sum_{i=1}^{M_t(\lambda)} \log V_i \right\} \right], \quad (5)$$

となる [6]。ここで、 $W(t)$ は Wiener 過程であり、形式的には白色雑音の時間積分 $\int_0^t \gamma(s)ds$ で定義されるものである。

本論文では、フォールト発見率 $b(t) \equiv b_1(t)$ 、 $b(t) \equiv b_2(t)$ 、および $c(t)$ は、次式を満たすものとする。

$$\int_0^t b_1(s)ds = (1 - \exp[-bt]), \quad (6)$$

$$\int_0^t b_2(s)ds = (1 - (1 + bt) \exp[-bt]), \quad (7)$$

$$c(t) = \exp \left[-\frac{t}{c_n} \right]. \quad (8)$$

ここで、 b はフォールト 1 個当りのフォールト発見率を、 c_n はクラウド OSS を構成するコンポーネント数を表す。

3 ジャンプ拡散過程

本論文では、正規分布に従うガウシアン・ジャンプ拡散過程およびラプラス分布に従うラブラシアン・ジャンプ拡散過程の2種類のジャンプ拡散過程を取り扱う。

3.1 ガウシアン・ジャンプ拡散過程

i 番目のジャンプ幅 V_i がガウシアン・ジャンプ拡散過程に従う場合、その密度関数は以下のように与えられる。

$$V_i \equiv f_n(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \quad (9)$$

ここで、 μ および σ は、 i 番目のジャンプに対する平均と標準偏差を表す。

ガウシアン・ジャンプ拡散過程は、ジャンプ幅の平均と標準偏差が大きくなるにつれてジャンプ幅が増大する傾向がある。

3.2 ラブラシアン・ジャンプ拡散過程

また、 i 番目のジャンプ幅 V_i がラブラシアン・ジャンプ拡散過程に従う場合、その密度関数は以下のように与えられる。

$$V_i \equiv f_l(x) = \frac{\exp\left(-\frac{|x-\nu|}{\phi}\right)}{2\phi}. \quad (10)$$

ここで、 ν および $\sqrt{2}\phi$ は、 i 番目のジャンプに対する平均と標準偏差を表す。

ラブラシアン・ジャンプ拡散過程に従う確率変数の変動分布の裾の減衰は、ガウシアン・ジャンプ拡散過程よりも緩やかとなる傾向がある。

4 ディペンダビリティ評価尺度

クラウド環境が障害なく継続的に利用できる状態かどうかを確認することは、運用段階における保守コストの削減にもつながることから非常に重要となる。本論文では、継続した利用可能時間を評価するための尺度として稼働率を以下のように定義する。

$$PER(t) = \frac{t - MTTR(t) \cdot N(t)}{t}. \quad (11)$$

ここで、 $MTTR$ は時刻 t におけるクラウド OSS の不動作時間、すなわち平均修復時間を表す。本論文では、平均修復時間 $MTTR$ は修復率 λ をもつ指数分布に従うものと仮定する。

5 最適メンテナンス時刻の推定

クラウド OSS を運用する際における総コストを定式化するために、以下のパラメータを定義する。

c_1 : 単位時間当りの運用コスト ($c_1 > 0$),

c_2 : フォールト 1 個当りの修正コスト ($c_2 > 0$),

c_3 : 運用段階におけるフォールト 1 個当りの保守コスト ($c_3 > c_2$).

ここで、 c_2 はバグトラッキングシステム上に登録されたフォールトを対象とし、 c_3 は実際のクラウドの運用環境に起因するフォールトを対象とする。このとき、総コストを以下のように定義する。

$$C(t) = c_1 t + c_2 N(t) + c_3 \{D(t) - N(t)\}. \quad (12)$$

この式 (12) を最小にする時刻 t^* が、クラウド OSS の最適メンテナンス時刻となる。

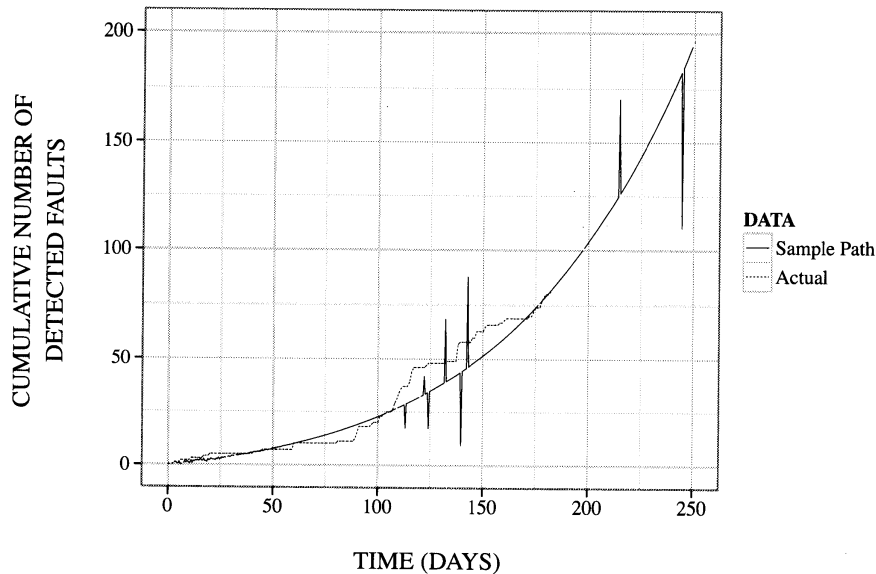


図 2： 推定されたガウシアン・ジャンプ拡散過程に基づく累積フォールト発見数のサンプルパス ($b(t) \equiv b_1(t)$ の場合)。

6 数値例

実際のクラウド OSS のオープンソースプロジェクトである OpenStack [7] 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数値例を示す。

推定されたガウシアン・ジャンプ拡散過程に基づく累積フォールト発見数のサンプルパスを図 2 および図 3 に示す。また、推定されたラプラシアン・ジャンプ拡散過程に基づく累積フォールト発見数のサンプルパスを図 4 および図 5 に示す。図 2～図 5 から、ジャンプの幅と大きさが異なり、ラプラシアン・ジャンプ拡散過程に従う確率変数の変動分布の裾の減衰は、ガウシアン・ジャンプ拡散過程よりも緩やかとなる傾向がある様子が確認できる。

さらに、2 種類のジャンプ拡散過程に基づき推定された総コストのサンプルパスを図 6 および図 7 に示す。図 6 および図 7 から、 $b_2(t)$ の場合においては、ジャンプの幅と大きさが異なり、ラプラシアン・ジャンプ拡散過程に従う確率変数の変動分布の裾の減衰は、ガウシアン・ジャンプ拡散過程よりも緩やかとなる傾向がある様子が確認できる。また、2 種類のジャンプ拡散過程の最適メンテナンス時刻はほぼ同じ値となり、約 200 日でクラウド OSS のメンテナンスを実施すれば良いことが確認できる。

7 おわりに

本論文では、クラウド OSS の実利用環境を想定し、信頼性評価のためのジャンプ項を含む確率微分方程式モデルを利用したディペンダビリティ評価法について議論した。特に、クラウドシステムのクラウドアプリケーション数などが、ノイズとして影響を及ぼすものと仮定し、実際のフォールト発見数データに対する数値例を示した。さらに、ガウシアン・ジャンプ拡散過程およびラプラシアン・ジャンプ拡散過程の 2 種類のジャンプ拡散過程を適用した提案モデルに対する適用可能性について考察した。

クラウド OSS の運用段階に対して、より現実的な信頼性評価および最適メンテナンス時刻の推定手法として利用できるものとする。

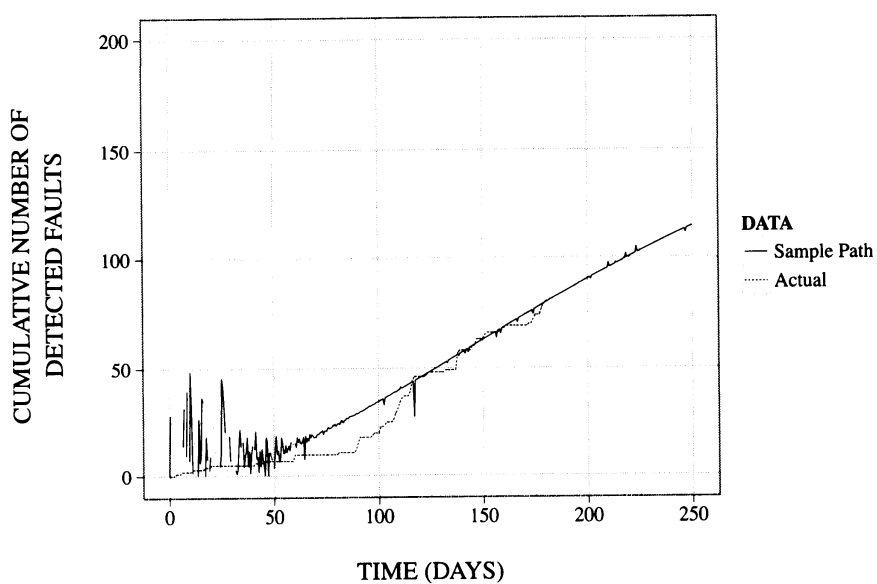


図 3： 推定されたガウシアン・ジャンプ拡散過程に基づく累積フォールト発見数のサンプルパス ($b(t) \equiv b_2(t)$ の場合)。

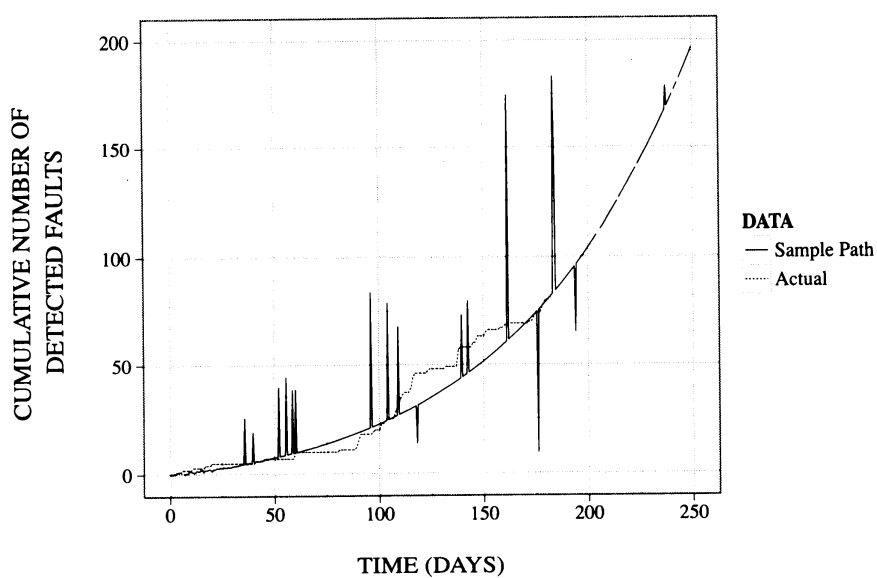


図 4： 推定されたラプラシアン・ジャンプ拡散過程に基づく累積フォールト発見数のサンプルパス ($b(t) \equiv b_1(t)$ の場合)。

謝辞

本研究の一部は、文部科学省科学研究費基盤研究 (C) (課題番号 24500066 および 22510150) の援助を受けたことを付記する。

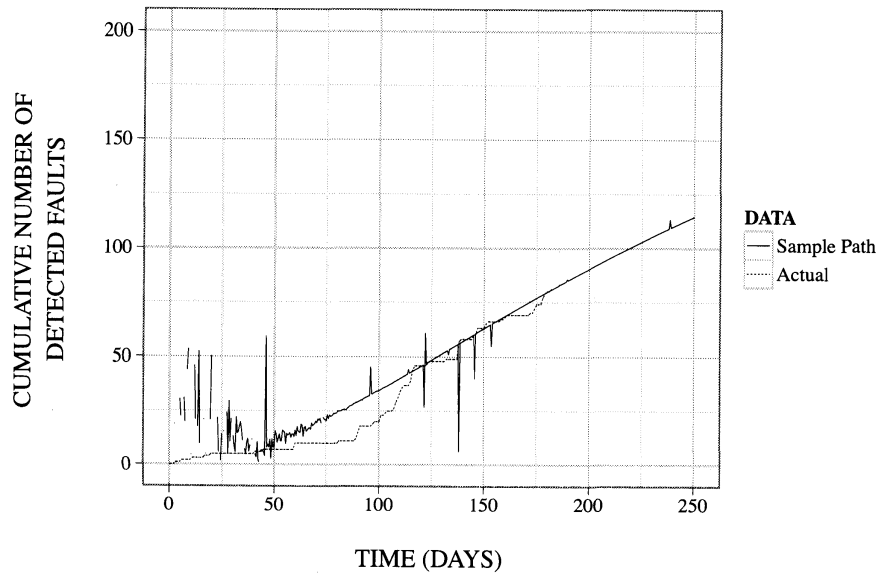


図 5： 推定されたラプラシアン・ジャンプ拡散過程に基づく累積フォールト発見数のサンプルパス ($b(t) \equiv b_2(t)$ の場合)。

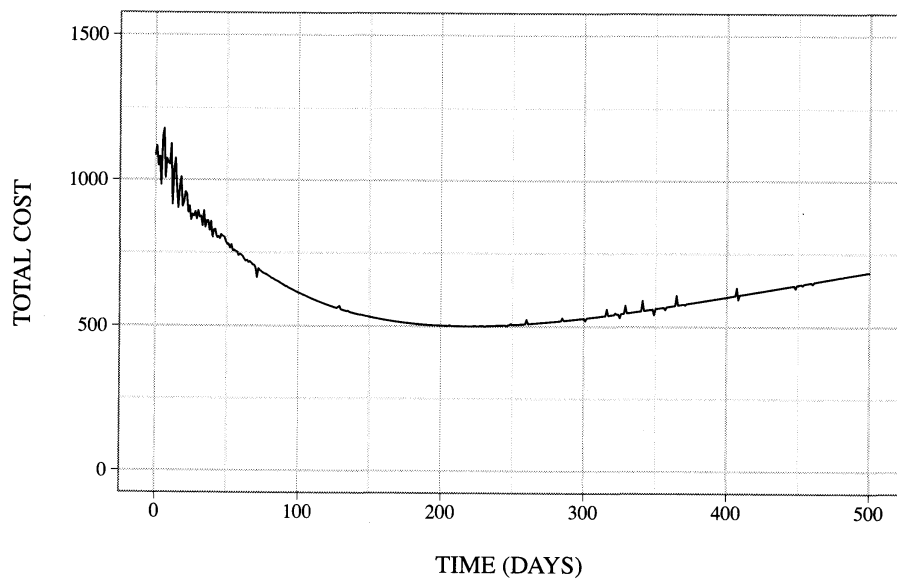


図 6： 推定されたガウシアン・ジャンプ拡散過程に基づく総コストのサンプルパス ($b(t) \equiv b_2(t)$ の場合)。

参考文献

- [1] Bo Yang, Feng Tan and Yuan-Shun Dai, Performance evaluation of cloud service considering fault recovery, The Journal of Supercomputing, Springer, published online: 23 Feb., 2011.

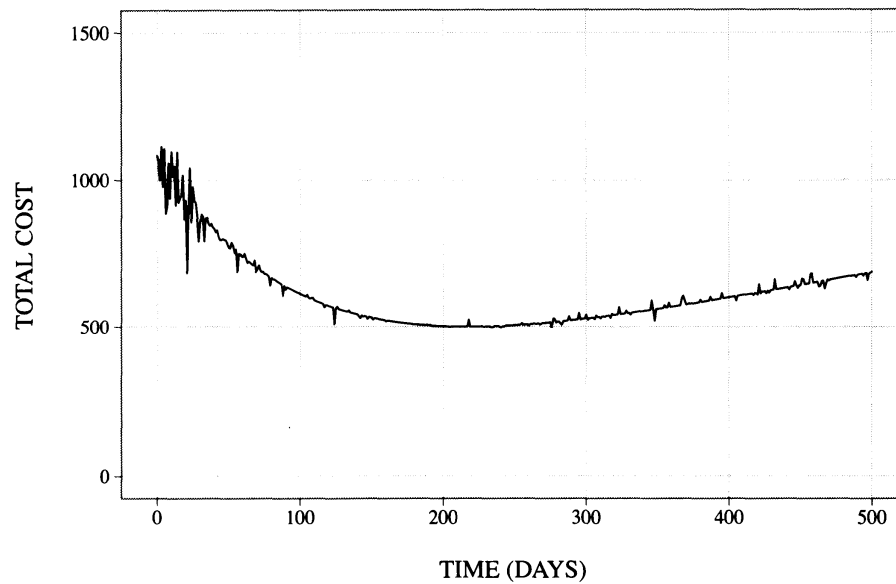


図 7: 推定されたラプラシアン・ジャンプ拡散過程に基づく総コストのサンプルパス ($b(t) \equiv b_2(t)$ の場合)。

- [2] A. Iosup, S. Ostermann, M.N. Yigitbasi, R. Prodan, T. Fahringer, and D.H.J. Epema, Performance analysis of cloud computing services for many-tasks scientific computing, *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 6, 2011.
- [3] S. Yamada, *Software Reliability Models: Fundamentals and Applications* (in Japanese), JUSE Press, Tokyo, 1994.
- [4] S. Yamada and T. Fujiwara, “Testing-domain dependent software reliability growth models and their comparisons of goodness-of-fit,” *International Journal of Reliability, Quality and Safety Engineering*, vol. 8, no. 3, pp. 205–218, 2001.
- [5] L. Arnold, *Stochastic Differential Equations-Theory and Applications*, John Wiley & Sons, New York, 1974.
- [6] 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.
- [7] The OpenStack project, OpenStack. [Online]. Available: <http://www.openstack.org/>