

クラウドサービスにおける サーバ群のブロック運用方式の性能解析

森井 広圭, 高橋 豊

キーワード：クラウドコンピューティング, サーバホスティングサービス, ブロック運用方式, マルコフ連鎖

本稿は、森井 広圭による 2015 年度京都大学工学部情報学科に提出した卒業論文をもとに加筆修正したものです。

1. 序論

従来、ユーザは手元にあるハードウェアおよびそこにインストールされたソフトウェアしか利用することができなかったが、近年のコンピュータおよびネットワークの高速化により、インターネット上にあるソフトウェアやハードウェアを必要ときに必要なだけ利用できるようになった。このように、ハードウェアとソフトウェアの両方をサービスとして提供するクラウドコンピューティングが近年注目を集めている。このような環境は仮想化技術に支えられており、大量に集約された計算機資源を仮想的に分割することで、ユーザはその一部を占有的に利用することができる [1]。

ユーザは図 1 のようにレンタルサーバを必要ときに必要なだけ利用することができ、必要でないときは自由に返却することができる。これにより、費用を削減しつつ、サーバへの負荷の増減に柔軟に対応することができる。

レンタルサーバを利用するにあたり、主に二つの費用が発生する。データセンタからレンタルサーバを新たに借り上げる度に発生するセットアップコストと、レンタルサーバを保有している時間に応じて発生するランニングコストである。ランニングコストには消費電力料やメンテナンスコストなどが含まれている。セットアップコストが単位時間当たりに発生するランニ

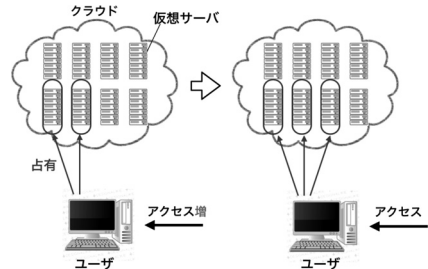


図 1 サーバホスティングサービスの概要

グコストよりも比較的大きく設定されている場合は、頻繁にサーバの借上と返却を繰り返すよりも、レンタルサーバがジョブを実行していない状態（アイドル状態）になった後もしばらく保有し続けるほうが費用を抑えられる場合があることに注意する。

このことを受け、Ezhilchelvan ら [2] は、レンタルサーバを 1 台ずつ取り扱うのではなく、まとまった数をブロックとして取り扱い、ブロック単位で借上と返却を行う運用方式を提案している。また、レンタルサーバの運用により得られる利益とレンタルサーバの維持にかかる費用の両面を考慮した運用利益評価関数を用いて、提案方式を含めたさまざまな方式の比較がなされている。しかしながら、一度に取り扱うレンタルサーバ数（ブロックサイズ）がシステム性能に与える影響は十分に検討されていない。

そこで本研究では、この点について検討を行った。そのために、システム内にアイドル状態のレンタルサーバがない場合はジョブをバッファに一度蓄積し、一定数溜まるとレンタルサーバ群（ブロック）を借り上げて一斉にサービスを開始し、アイドル状態のレンタルサーバが一定数生じた場合はブロックを返却する待ち行列モデルを考案した。このシステムの運用利益を計る運用利益評価関数を提案し、解析で導出した性能評価指標（ジョブ廃棄率、ジョブ待ち時間分布）をもとに数値実験を行った。

もりい ひろき, たかはし ゆたか
京都大学 大学院情報学研究所
〒 606-8501 京都府京都市左京区吉田本町
morii_hiroki@sys.i.kyoto-u.ac.jp
takahashi@i.kyoto-u.ac.jp

2. 数理モデル

本節では、ブロック運用方式の待ち行列モデルについて述べる。システムはレンタルサーバ群とバッファからなる。レンタルサーバはデータセンタからブロック単位で即座に調達される。この際、ブロックサイズを c とする。また、データセンタから借り上げられるレンタルサーバの総数には上限があり、最大 M 台までとする。すなわち、借り上げられるブロック数の上限は $b = \lceil M/c \rceil$ である。 M が c で割り切れない場合、最後に借り上げるブロックは $M \bmod c$ 台のレンタルサーバ群とする。また、各レンタルサーバでのジョブ処理時間は、率 μ の独立で同一な指数分布に従うものとする。このシステムに対し、ジョブは率 λ のポアソン過程に従って到着する。ジョブ到着時のシステムの対応は以下の三つのケースに分類される。

- (1) アイドル状態のレンタルサーバが存在する場合、ジョブは即座にサービスを受ける。
- (2) アイドル状態のレンタルサーバが存在しない場合、ジョブはバッファの最後尾に蓄積され、次の二つの事象のいずれかが起きればサービスが開始される。
 - (i) 順番が回ってきた場合。
 - (ii) バッファ内ジョブ数がブロックサイズ c に達し、新たにブロックを借り上げた場合。
- (3) M 台のレンタルサーバがすべて稼働中に到着した場合、ジョブは廃棄される。

また、アイドル状態のレンタルサーバ数が c 個に達すると 1 ブロック分のレンタルサーバ群を直ちにデータセンタに返却する。

3. 解析

本節では、解析の概要を述べる。 $I(t)$ と $J(t)$ をそれぞれ時刻 t におけるシステム内のブロック数とバッファ内ジョブ数を表す状態変数とする。確率過程 $\{(I(t), J(t)); t \geq 0\}$ は状態数が有限のマルコフ連鎖となり、モデルの仮定より既約であるため、唯一の定常分布 π を持つ。 π が得られると、ジョブ廃棄率とジョブ待ち時間分布が導出可能になる。これらの性能評価指標を用いて、運用利益評価関数を設定する。

運用利益評価関数を構成する要素は以下のとおりである。

- ・一つのジョブを処理した際に得られる報酬
- ・待ち時間がサービス品質保障で設定された値を超えた際に支払うペナルティ

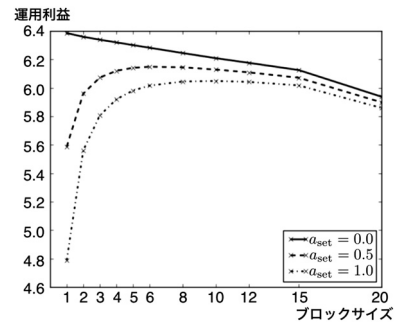


図2 ブロックサイズと運用利益の関係

- ・セットアップコスト
- ・ランニングコスト

これらの要素を考慮すると、運用利益評価関数は次のように表される。

$$R = \lambda(1 - q)\{r - u\Pr(W > v \mid \mathcal{E}_{\text{accepted}})\} - \{a_{\text{set}}S + a_{\text{run}}L\}.$$

右辺の第1項は得られる報酬を、第2項は発生する費用を表している。 q はジョブ廃棄率、 r は一つのジョブを処理した際に得られる報酬、 u はジョブの待ち時間がある閾値 v を超えた際に支払うペナルティ、 $\Pr(W \leq t)$ はジョブ待ち時間分布、 $\mathcal{E}_{\text{accepted}}$ はジョブの到着時に廃棄が発生しないという事象、 a_{set} はセットアップコスト、 a_{run} はランニングコスト、 S は単位時間当たりに新たに借り上げるレンタルサーバ数の期待値、 L はシステム内に存在するレンタルサーバ数の期待値である。

4. 数値結果と考察

数値実験においては、セットアップコストが増加した際に、ブロックサイズの変化が運用利益に与える影響について検討を行い、図2のような結果が得られた。パラメータ設定は $M = 120$, $\lambda = 1.6$, $\mu = 1/60$, $r = 10$, $u = 5$, $v = 10$, $a_{\text{run}} = 0.1$ であり、 $a_{\text{set}} = 0, 0.5, 1.0$ と変化させている。これより、システムの運用利益は、ブロックサイズが大きいほど、セットアップコストの増減による影響を受けにくいことが判明した。これは、ブロックサイズが大きくなるにつれて、ブロックの借上頻度が減少することに起因する。

参考文献

- [1] 浦本直彦, “クラウド・コンピューティング一動向と今後の展望,” オペレーションズ・リサーチ: 経営の科学, **56**(6), pp. 305–310, 2011.
- [2] P. Ezhilchelvan and I. Mitrani, “Static and dynamic hosting of cloud servers,” *Computer Performance Engineering, Lecture Notes in Computer Science*, **9272**, pp. 19–31, 2015.