

サービス工学におけるシステム応答時間の分析

○魚住光成(三菱電機(株)) 浅間一(東京大学)

Analysis at system response time in service engineering

*Mitsunari UOZUMI, MITSUBISHI Electric Corp., Hajime ASAMA, The Univ. of Tokyo

Abstract —It is necessary to evaluate the performance of the computer system that provides service at response time. In a complex application, response time becomes a synthesis of two or more normal distribution. The effect of cache is not impartial depending on the characteristic of the application. It is thought one caused by the influence.

Key Words: Service engineering

1. はじめに

人にサービスを提供するシステム、たとえばATMや券売機、ECサイトなど多種多様なサービスが、コンピュータと通信ネットワークを使って提供されている。

これら機械が人にサービスを提供するとき、システムの負荷状況によって時間性能は変動し、サービス品質の変動として利用者への影響が現れる。

我々は、サービスを提供するシステムの応答時間の分析を行なった。

2. サービス工学とサービスメディア

サービス工学では、サービスを提供するシステムをサービスメディアと位置づけ、これがサービスを提供するだけでなくサービスの評価計測を行い、利用者の満足度としてシステムにフィードバックするモデルを提唱している。^[1]

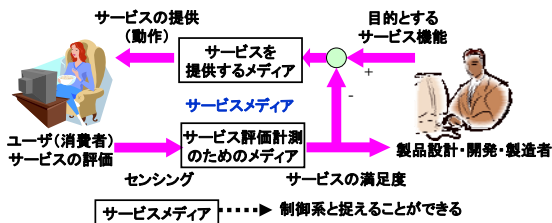


Fig.1 Service media in service engineering

このモデルにおいて、利用者の満足度は、サービスのコンテンツによる変化と、サービスメディアの品質によって変化すると考えられる^[2]。サービスメディアの品質は、サービスコンテンツを利用者の要求に従って引き出すコンピュータなどの機器の時間性能と、利用者までサービスコンテンツを伝達する通信ネットワークの通信品質によって変動する。

本研究では、このサービスを提供する機器の時間性能の分析を行なった。

3. サービスメディアの時間性能品質

コンピュータの処理性能の向上により、1台のコンピュータが複数の利用者に対し、同時にサービスを提供する事は一般的になっている。このときの処

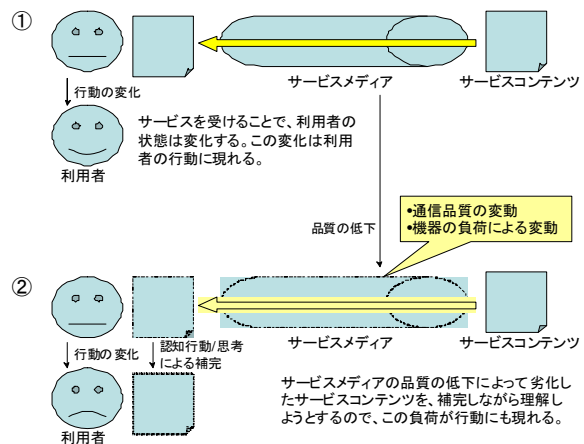


Fig.2 Quality of Service media

理性能は、スループットとして表されることが多い。スループットは、単位時間当たりの処理件数や出力したデータ量であり、コンピュータの処理能力をわかりやすく表す指標である。

しかし、スループットは、コンピュータの処理能力の指標となるが、利用者個々の応答時間を推定することは難しい。多重度が上がるにつれスループットが上昇しても、個々の要求の応答時間については、均一であるとは限らない。

サービスメディアの品質は、利用者からみたシステムの品質であるため、スループットではなく、応答時間に基づく指標であるべきである。

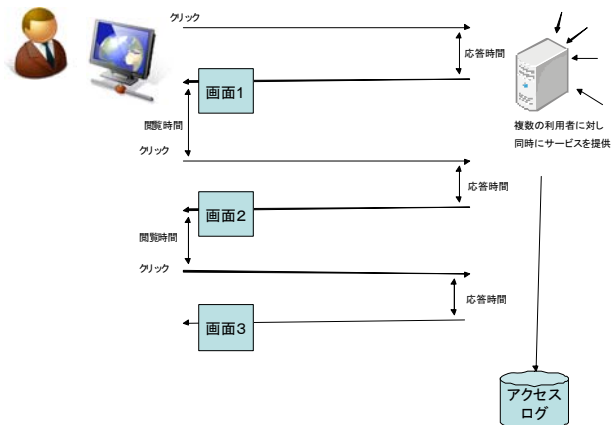


Fig.3 A system and response time

4. 応答時間の実際

コンピュータは、外部からの要求を単純に逐次処理するだけでなく、可能な場合は、複数の処理を同時並行に実行する。また、キャッシュなど処理を高速化するメカニズムも備えており、その為、要求を受け付けてから結果がでるまでの時間を正確に表現することは難しい。しかし、統計量としては多重度が上がるにつれ、応答時間の平均と分散が上昇することが考えられる^[3]。

しかし、特定の多重度における応答時間は、システム内の複数の要因により、Fig.4のような複合した正規分布となり、ある多重度における応答時間の分布 $f(y)$ は、

$$f(y) = \sum_i N(\mu_i, \sigma_i^2) \quad (5.1)$$

$$N(\mu_i, \sigma_i^2) = \frac{\alpha_i}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(y-\mu_i)^2}{2\sigma_i^2}\right\} \quad (5.2)$$

と表すことができる。

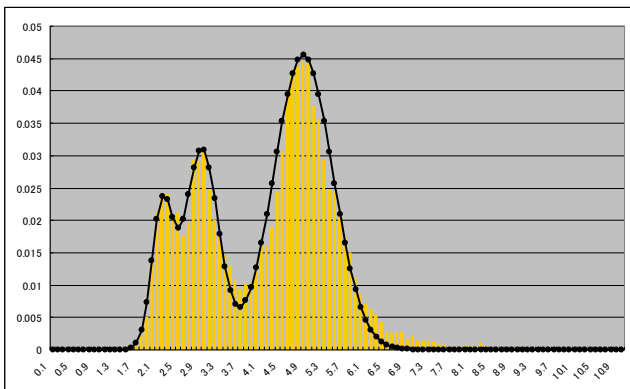


Fig.4 Observed data

5. システムにおけるキャッシュの振る舞い

処理の高速化のためのキャッシュは、要求に対して公平に機能することが期待されている。Fig.5は、100のユニークなページに対し、アプリケーション

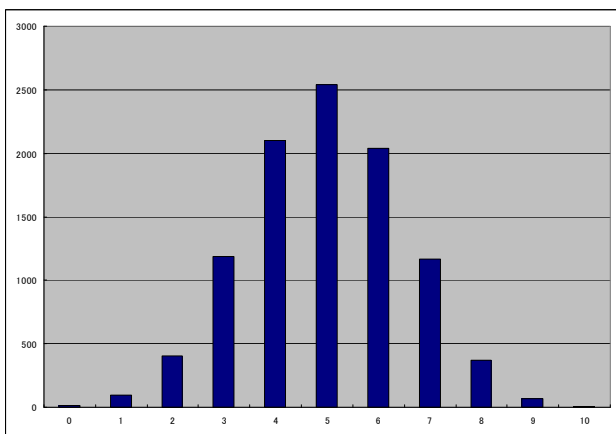


Fig.5 Distribution of random access

がランダムな10ページをアクセスした場合の実行コストの分布を示したものである。キャッシュは、FIFOによる管理が行なわれ、サイズは100ページとしている。キャッシュにヒットした場合は、コスト0、ヒットしなかった場合はコスト1としてアプリケーションの実行コストを積み上げる。

この場合は期待通りキャッシュは公平に機能し、分布は正規分布となっているが、Fig.4に示したような現象は説明できない。

6. アプリケーションの規則性とキャッシュ

アプリケーションが内部で規則的な振る舞いをする場合、例えば、ランダムに決まるページ番号を元に、ページ番号の連続した10ページをアクセスする場合について、同様のシミュレーションを実施した。その結果が、Fig.6である。

FIFOで管理されたキャッシュがあり、また、アプリケーション間にランダム性があっても、アプリケーションの規則によって、処理コストの分布は変わっている。

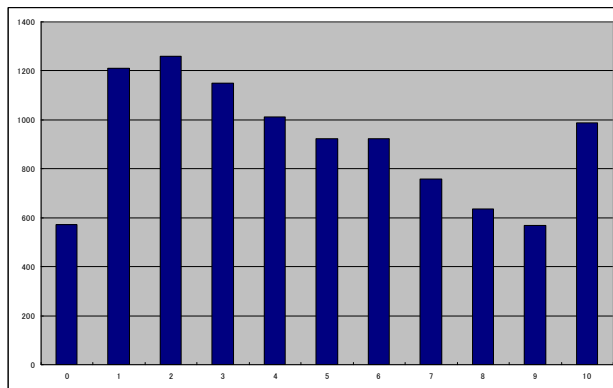


Fig.6 Distribution of random access with same rule

7. 結論

このシミュレーションは、アプリケーションのディスクアクセスとみなすことができ、またコストは応答時間である。

キャッシュなど高速化のメカニズムを持つシステムでは、キャッシュの効果は、アプリケーションの特性によって公平でないことを示しており、これによって複数の分布からなる応答時間が生まれると考えられる。

参考文献

- [1] 浅間一：サービス工学とシステムインテグレーション，計測と制御 278-283 Vo1. 44, 2005.
- [2] 魚住光成，浅間一：サービス工学における利用者のインタラクション評価，第25回日本ロボット学会学術講演会 1E35, 2007
- [3] 魚住光成，浅間一：サービス工学におけるシステム応答時間のモデル化，日本ロボット学会学術講演会 2009, 2009