適切なアクセス制御状態にあるネットワークシステムの特徴抽出

○金岡 晃（筑波大）、藤堂 伸勝（筑波大）、
加藤 雅彦（IIJテクノロジー）、
岡本 栄司（筑波大）

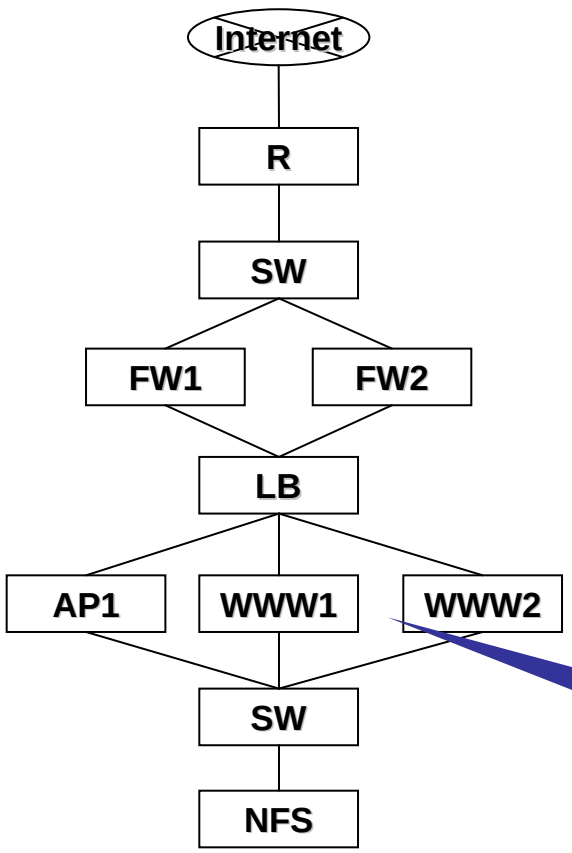
Outline

- 背景と目的
- これまでの成果
  - SCIS 2008
  - 第5回ICSS研究会
- 次数分布の関数近似
  - 候補関数の抽出
  - グループごとの近似
- 固定パラメータでの近似精度測定
- まとめ

本研究の目的と本論文の位置づけ

- 目的
  - ネットワークシステムにおける安全な設計方法論確立のためのシステム定量評価
  - ネットワークシステム（Networked System）:
    - サーバやデータベース、ルータ、ファイアウォール、ロードバランサなど、各機能をそれぞれの機器で行って、全体として1つのサービスを提供するシステム
- 位置づけ
  - 効率よくネットワークアクセス制御が施されたネットワークシステムが持つ共通特性の関数近似

ネットワークシステム設計時の考慮点と現状



現状の設計図例

- ネットワークシステム構築において考慮されるポイント
  - コスト、通信量、拡張性、アクセス制御、耐障害性、脆弱性の影響
- 最適化の困難性
  - 小規模システムでさえ複雑
  - 現状は設計者の経験依存
  - **方法論や理論的アプローチが未開拓**
- 現状の設計における問題点
  - 設計図と構築考慮ポイントの関連性が低い
  - 例：アクセス制御は適切にされるべきだが正確に反映されていない

WWW1-WWW2間の通信は？

各機器はそれぞれが持つ機能により特性が異なるが、それを1つの平面上に表現されているため

ネットワークシステム表現モデル（NSQモデル）

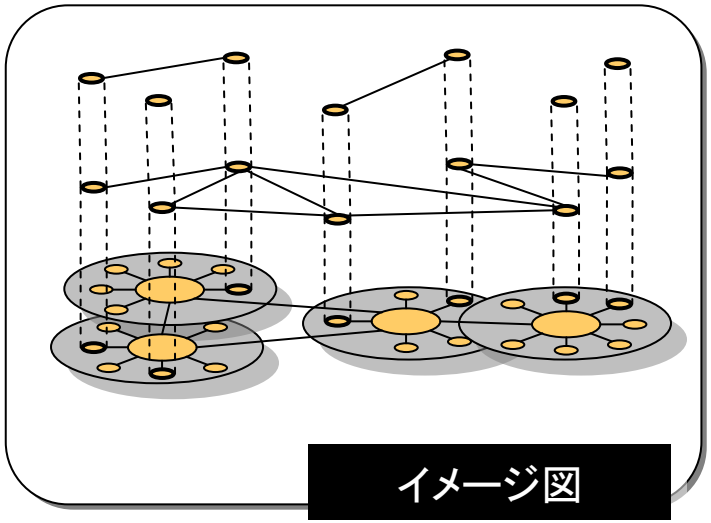
基本戦略

現状

物理的接続のみを反映したネットワーク表現

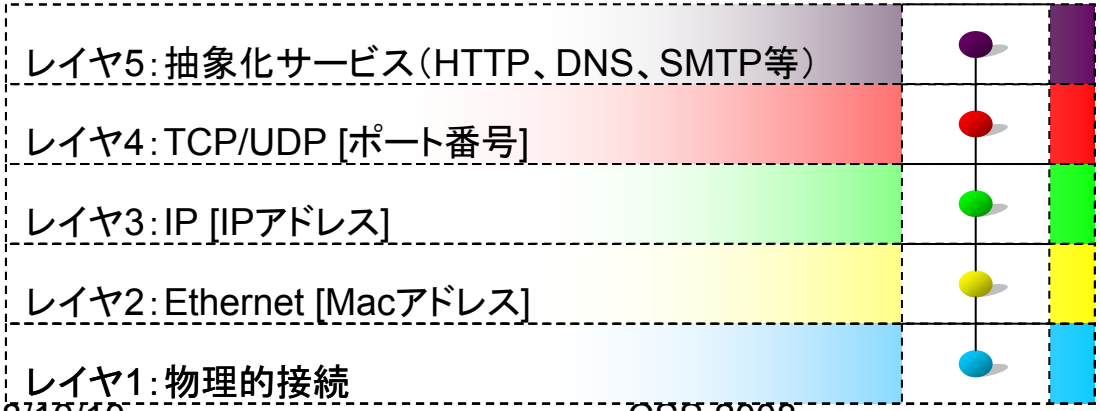
NSQモデル

TCP/IPの階層ごとに作られる論理ネットワーク
＋
階層ごとのネットワークの接続



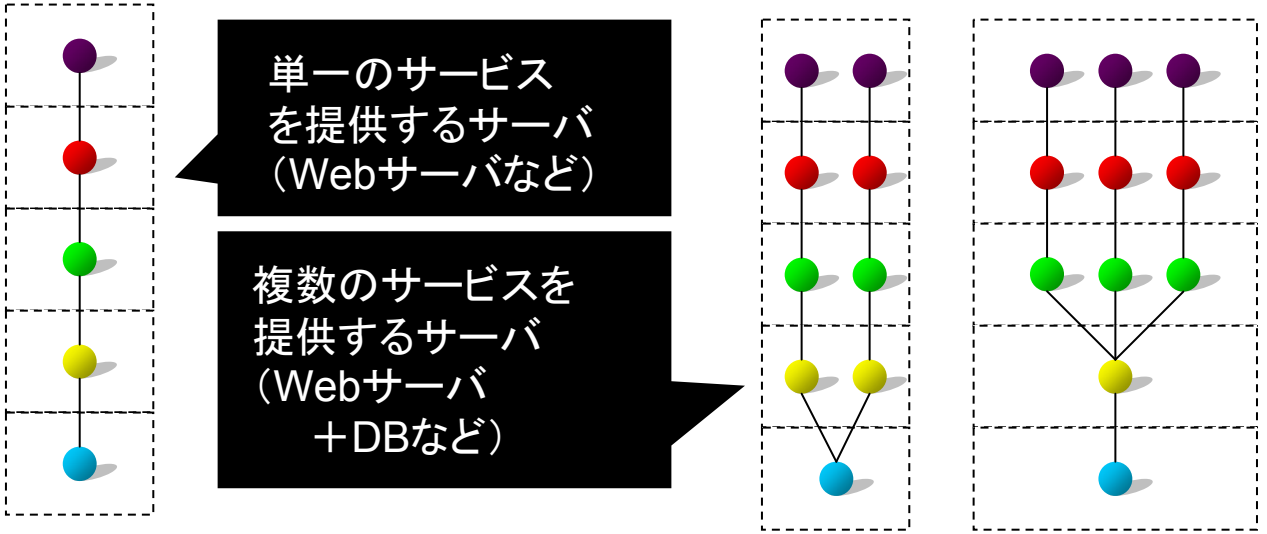
イメージ図

レイヤ定義

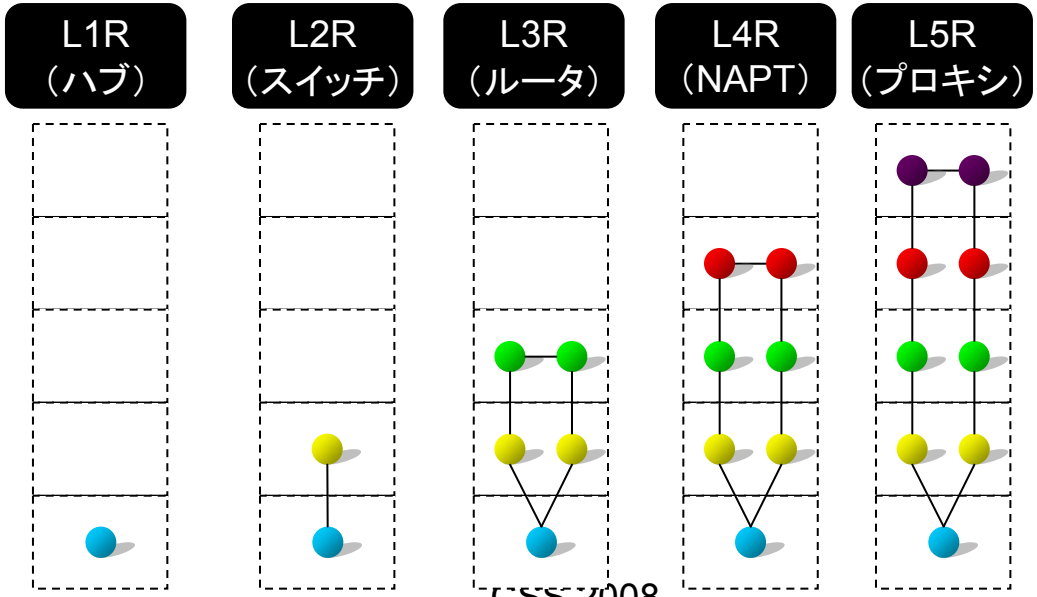


モジュール例

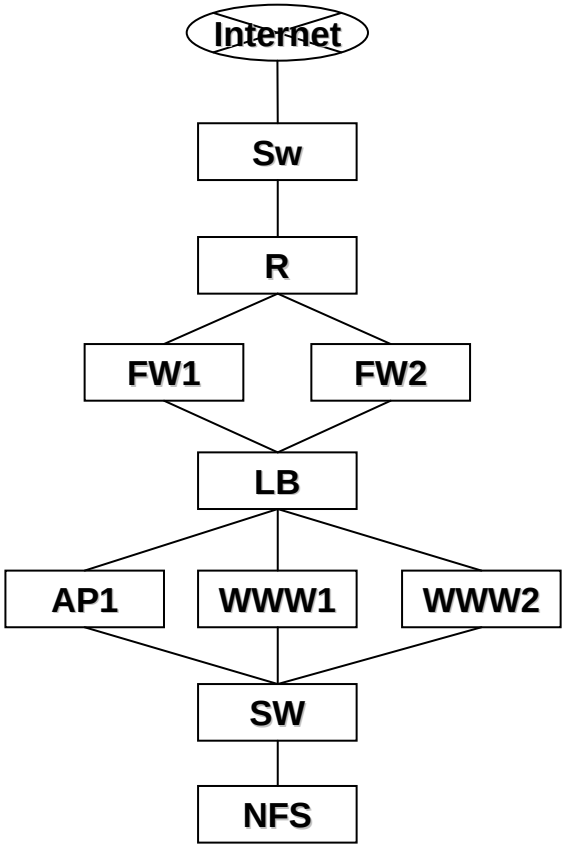
サーバ



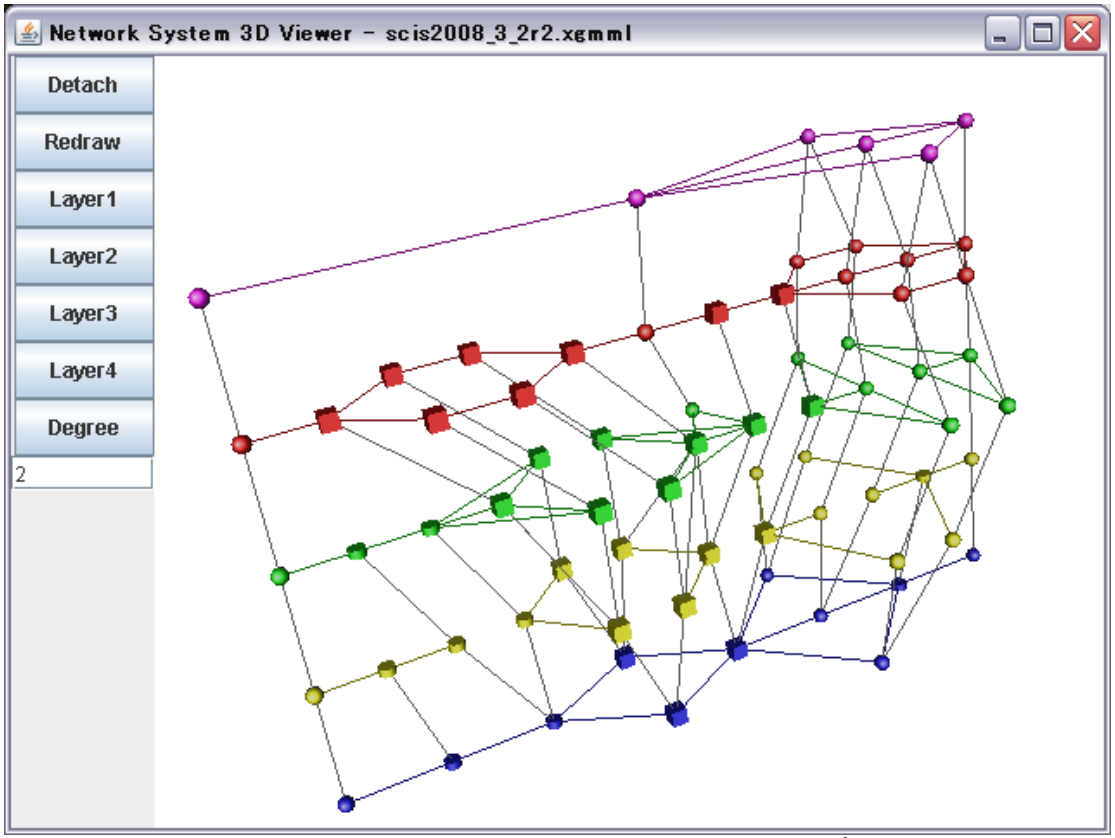
中継機器（機能）



NSQモデルによるネットワーク表現例



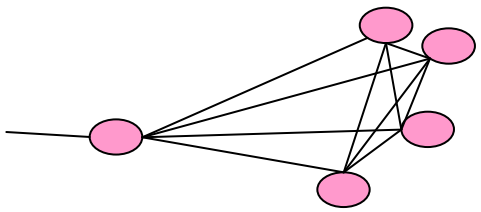
従来の表現



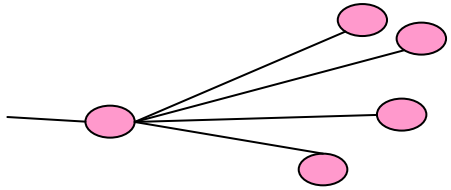
提案モデルによる表現

モデルが持つ特徴解析

- リンクがある＝ノード間が通信可能
- リンクの多さ＝必要となる通信パターンが多い 又は 不必要な通信が許可されている
- リンクの多さを見ることで、特徴がつかめないか
  - **各ノードが持つリンク数の分布（次数分布）を解析**



次数	ノード数
4	4
5	1



次数	ノード数
1	4
5	1

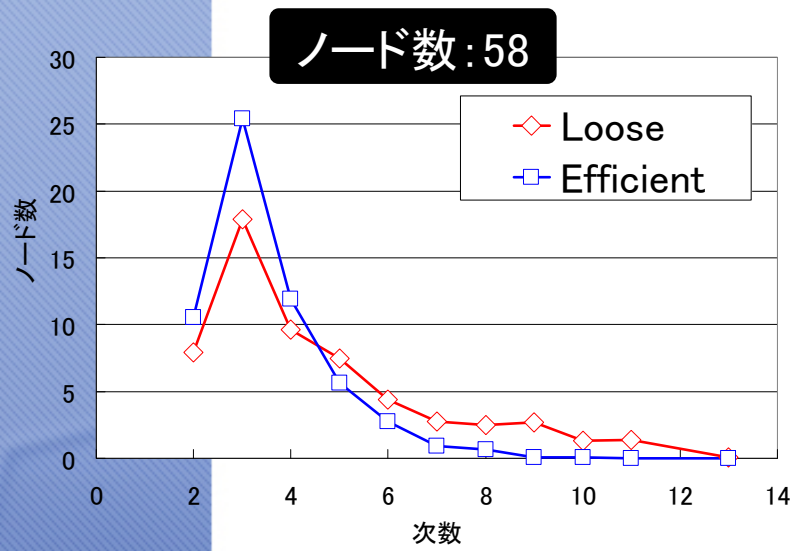
次数分布比較

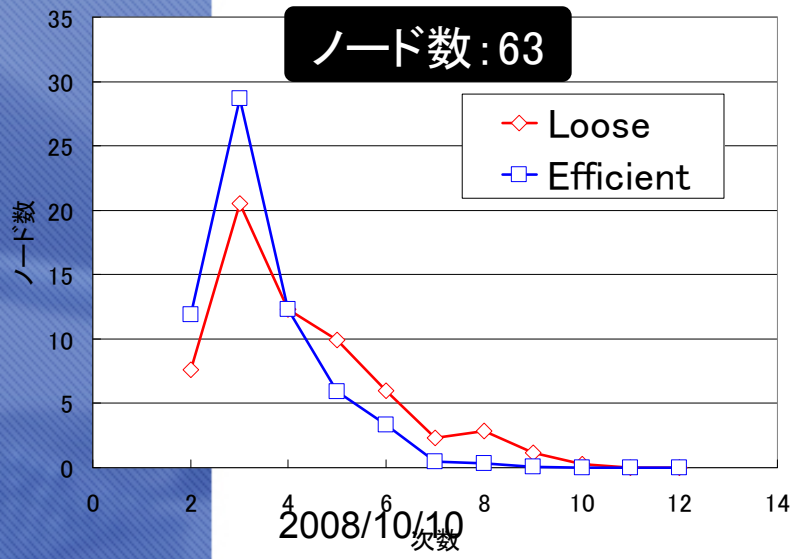
- 比較対象
  - アクセス制御の適用有無による違い
    - NSの構築手法（Efficient、Loose）により次数分布に差が現れるか
  - モジュール有無の違い
    - 特定モジュールの存在が次数分布にどのような違いを生じさせるか
    - L1RとL3Rについて提示
- 比較方法
  - さまざまなネットワークシステムの構成を含んだデータセットを準備
  - 全データセットを、ノード数ごとにグループ分けを行いそれぞれの次数分布を求める
  - グループに含まれるNS数が少ない（<300）のものを除き比較（ノード数が35以上）

最初に…
比較結果から得た1つの仮説を

最適なアクセス制御状態にあるネットワークシステムは
NSQモデル上の次数分布において一定の分布形状を持つ

アクセス制御状態による違い
ノード数：58、63





Loose（アクセス制御無し）、
Efficient（アクセス制御あり）共通

・次数は3で最大
・次数4以上は減少傾向

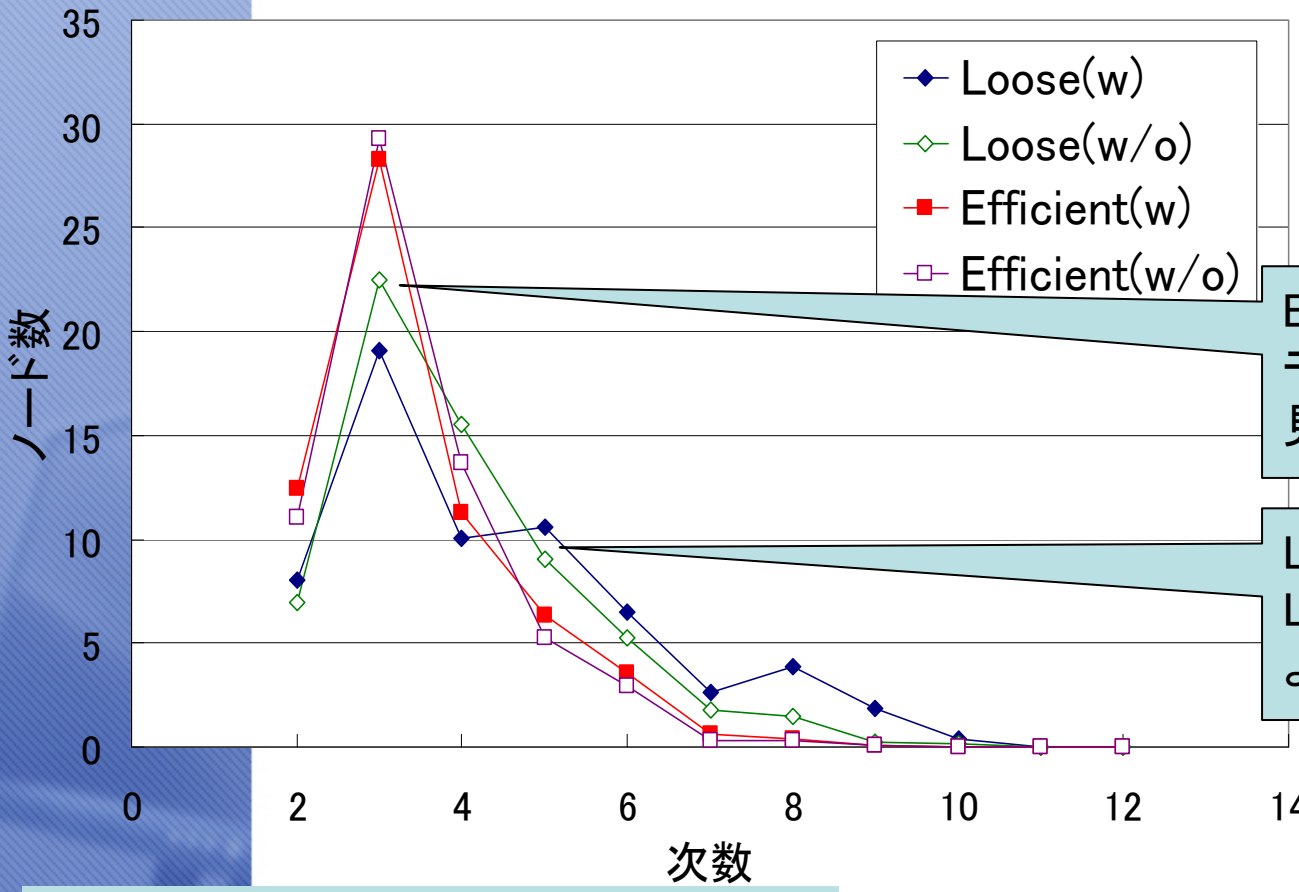
Loose、Efficient差異

・Efficient：次数4以上の減少がなだらか
・Loose：次数4以上の値が上下

・Loose：Efficientと比較して平坦

他のノード数においても
同様の傾向

特定モジュールの影響(L1R)
ノード数:63



Efficientケースでは
モジュール有無の差が
見られない

Looseケースでは
L1Rを持つNSは
より平坦な分布を持つ

他のノード数においても
同様の傾向

最適なアクセス制御状態にあるNSは
次数分布において一定の分布形状を持つ

⇨ 仮説を支持する結果

関数近似

- 仮説を基に関数で次数分布を近似する
  - 仮説「最適なアクセス制御状態にあるネットワークシステムは、NSQモデル上の次数分布において一定の分布形状を持つ」
- 近似に用いる関数のターゲット
  - 統計学で用いられる各分布から、同様の形になる可能性のある分布を抽出
  - パラメータに自由度を持たせ、候補関数を挙げる
- データ
  - 作成したデータセットのうち、アクセス制御が適切にされているネットワークシステム群のデータを利用

近似式のターゲット

- タイプA
  - ベキのみの利用（F分布タイプ） $f(x) \propto \frac{x^{\alpha}}{(1+\beta x)^{\gamma}}$
- タイプB
  - 指数(exp)*ベキのタイプ
  - タイプB-1（カイ2乗分布、ガンマ分布） $f(x) \propto x^{\alpha} \exp(\beta x)$
  - タイプB-2（カイ分布） $f(x) \propto x^{\alpha} \exp(\beta x^2)$
  - タイプB-3（逆ガウス分布（ワルド分布）） $f(x) \propto x^{\alpha} \exp\left(\beta x + \frac{\gamma}{x}\right)$
  - タイプB-4（対数正規分布） $f(x) \propto x^{\alpha} \exp\left\{(\beta \log x + \gamma)^2\right\}$
- タイプC
  - 離散分布タイプ
  - タイプC-1（ポワソン分布） $f(x) = \sum_{k=0}^{x} \frac{\lambda^k e^{-\lambda}}{k!}$
  - タイプC-2（エルミート分布） $f(x) = \sum_{j=0}^{[x/2]} \frac{{\lambda_1}^{x-2j} {\lambda_2}^{j}}{(x-2j)!\, j!}$

近似方法:
遺伝的アルゴリズムの利用

- 遺伝的アルゴリズム(GA)を使う
  - 個体 g (タイプAの場合、$g = (\Theta, \alpha, \beta, \gamma)$)をランダムに複数用意
  - それぞれの適合度を計算
  - 選択、交叉、突然変異を確率的に行う
- GA選択の理由
  - シンプルな最小二乗法を用いることも考えた
  - 対象となる分布についてのより詳細な解析が行っていない
  - 局所解に収束する可能性
  - 厳密解を求めているわけではない
- 適合度計算
  - 2乗誤差の和は、ノード数に応じて値が変わる
  - ノード数に依存しないように修正を加えた S を適合度として採用

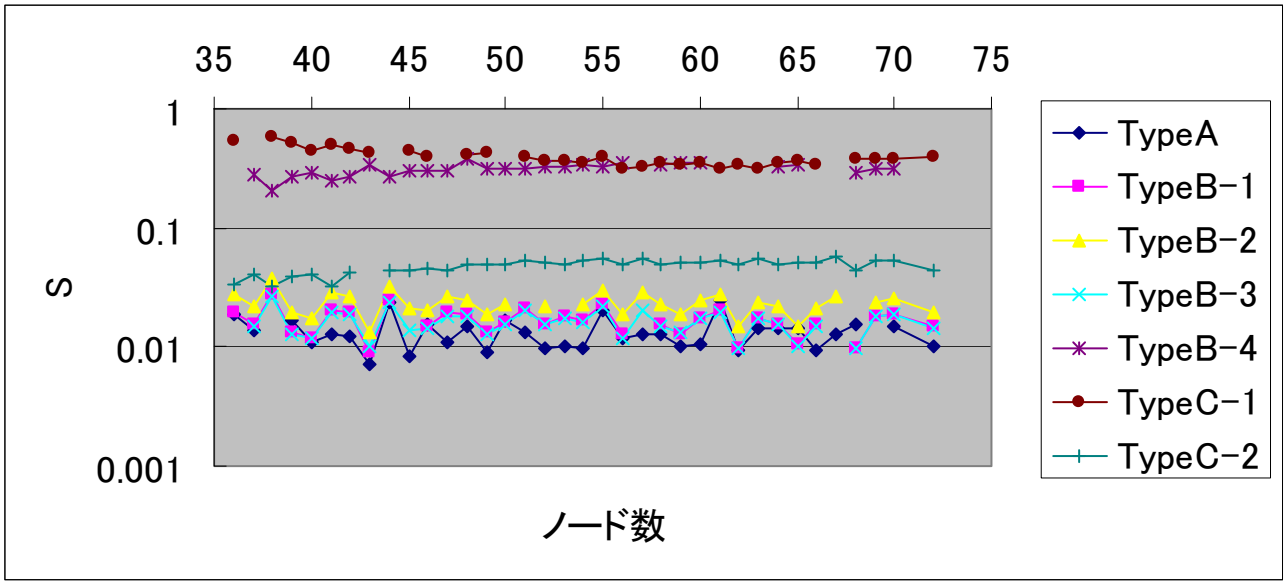
$$S = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i \hat{y}_i^{\,2}}$$

近似実験の流れ

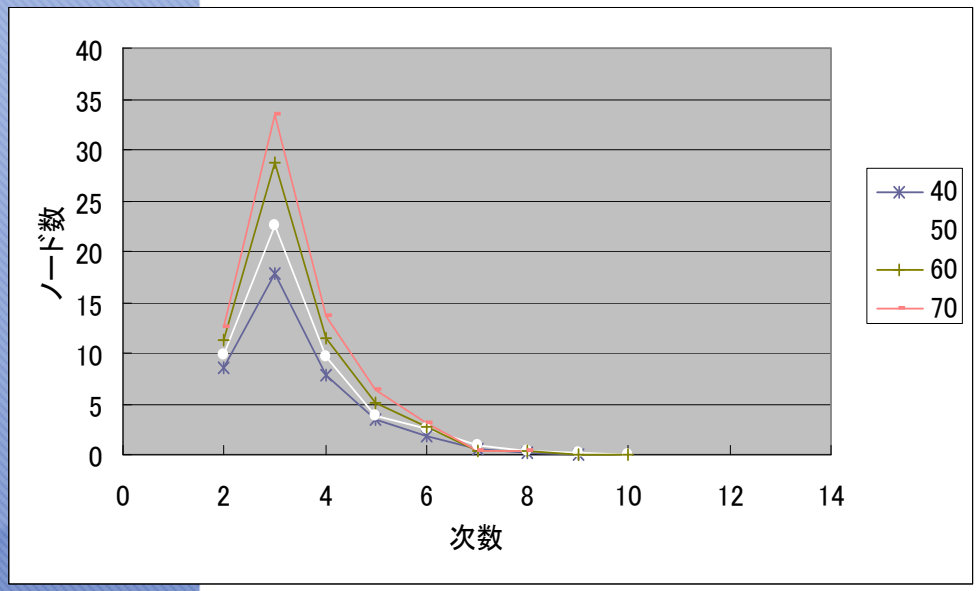
- データセットをノード数ごとに分割
  - 「グループ」と呼ぶ
    - ノード数16のグループ、ノード数53のグループ、など
- グループごとにGA
  - グループごとの近似パラメータを得る
    - ノード数ごとに特徴が異なっている場合を想定して
  - 近似精度より、ターゲット関数を絞る（7→4）
- データセット全体用のパラメータ計算
  - グループに関わらない近似パラメータを得る
  - ノード数ごとの差異を調整するスケール係数の計算

グループごとのGA

- 各ターゲット関数でS値の比較
  - 各S値で明らかに大きいもの（近似失敗）を削除したもので平均S値を計算
  - その中から選択し、4つのターゲット関数に絞り込む
    - Type A、B-1、B-2、B-3



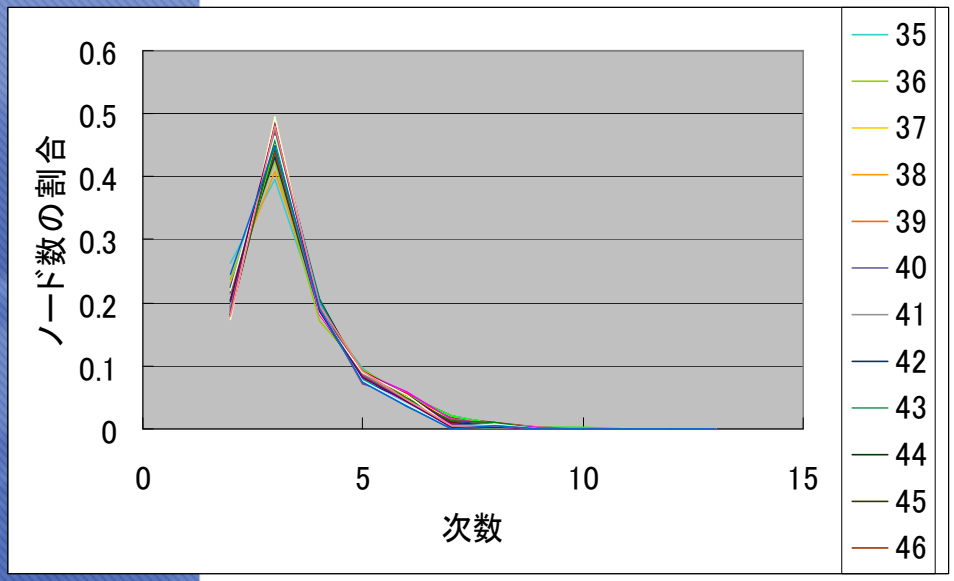
正規化した次数分布



正規化前の次数分布
（ノード数 40、50、60、70）

ほぼ同じ形で、縦（スケール）だけ異なるように見える

正規化（各次数/ノード数）



正規化後の次数分布
（ノード数 35-72）

ほぼ同じ形

スケール係数 Θ以外のパラメータは共通

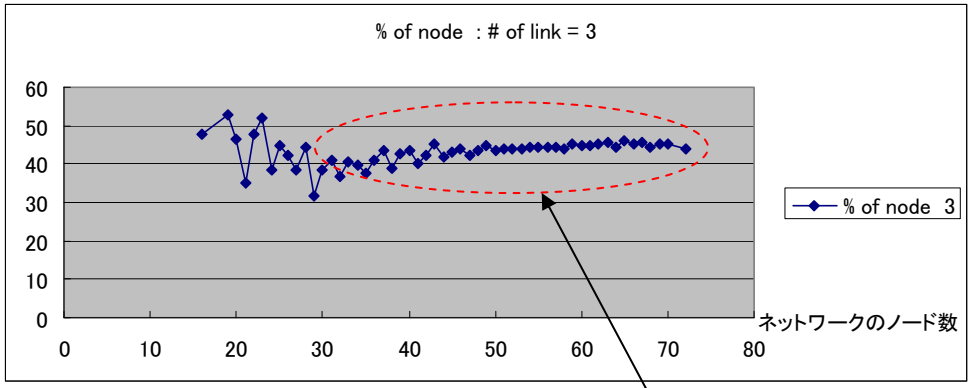
$$f(x) = \Theta x^{\alpha} \exp(\beta x)$$

データセット全体用のパラメータ計算

- 各パラメータの平均値を取り、その適合度を見る
  - ノード数のグループに関わらず、一定の特徴を持つか
- スケール係数 Θ の計算が必要

Θ（スケール係数）の計算

リンク数3の割合



安定してきている

ノード数36以上（ケース数200以上）
で平均を取ると 43.859%

全体のノード数 N × 43.859% ⇒最適なリンク数3のノード数： n

$f(3) = n$ になるように、 Θ の値をそれぞれ計算する

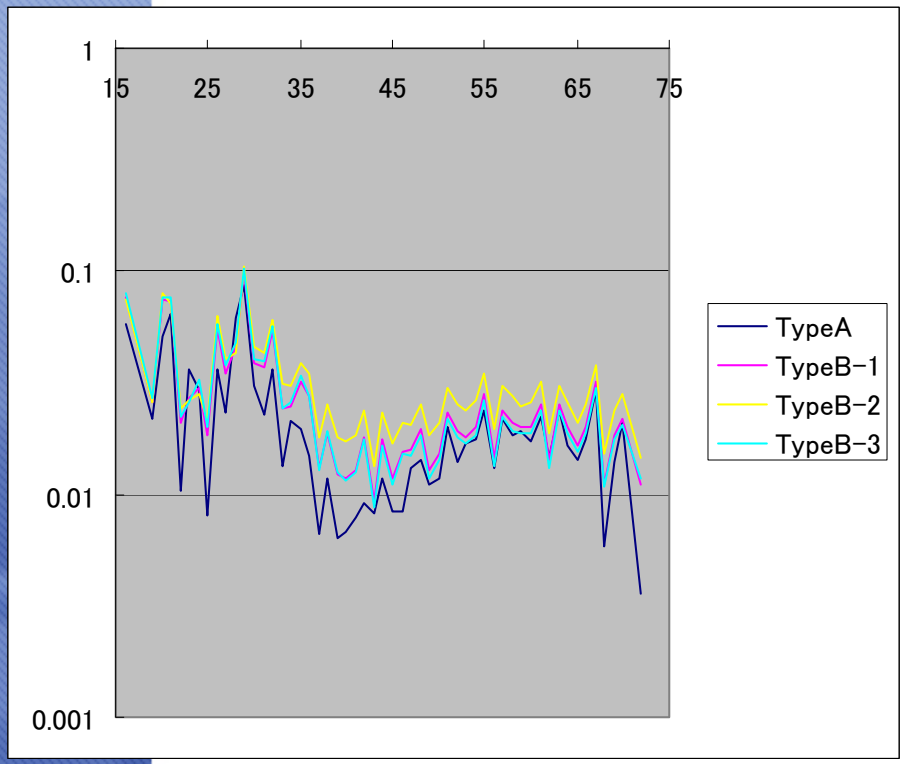
TypeAの場合

$$\Theta \frac{3^{30.11228}}{(1+0.52391649\cdot 3)^{52.29806}} = 0.43859N \Rightarrow \Theta = \frac{(1+0.52391649\cdot 3)^{52.29806}\cdot 0.43859}{3^{30.11228}}\cdot N$$

最適近似タイプの決定

- タイプA

$$f(x) \propto \frac{x^{\alpha}}{(1+\beta x)^{\gamma}}$$



	Avg. of S	Avg. of S (>35)
TypeA	0.021239929	0.014275489
TypeB-1	0.026811348	0.01820695
TypeB-2	0.031670138	0.023775861
TypeB-3	0.026853023	0.017311313

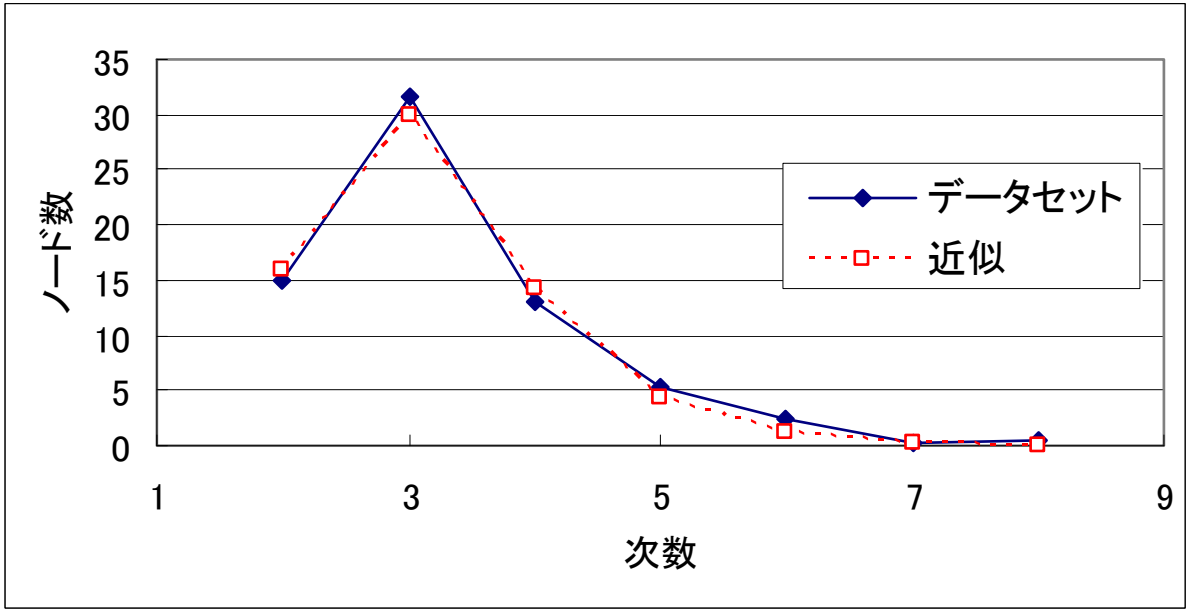
	α	β	γ
固定値	29.526	0.518	50.019

$$\Theta = \frac{(1+0.52391649\cdot 3)^{52.29806}\cdot 0.43859}{3^{30.11228}}\cdot N$$

最適なアクセス制御状態にあるときの次数分布

$$f(x) = \Theta \frac{x^{29.526}}{(1+0.518x)^{50.019}}$$

$$\left(\Theta = \frac{(1+0.52391649 \cdot 3)^{52.29806} \cdot 0.43859}{3^{30.11228}} \cdot N \right)$$



まとめ

- アクセス制御が適切になされたネットワークシステムは、NSQモデルの次数分布において一定の形状を取る
  - 近似をすることで、パラメータで表現することも可能になった
- 一定形状の近似関数を遺伝的アルゴリズムを用いて求めた
- 近似関数は右のとおり

$$f(x) = \Theta \frac{x^{29.526}}{(1+0.518x)^{50.019}}$$

$$\left(\Theta = \frac{(1+0.52391649 \cdot 3)^{52.29806} \cdot 0.43859}{3^{30.11228}} \cdot N \right)$$

- 今後
  - パラメータを用いた、アクセス制御状態の評価尺度
  - 評価尺度に基づいた設計アルゴリズム構築/修正支援アルゴリズム構築