

A study on the Difference in Marginal Distributions of Network Traffic

ネットワークトラフィックのゆらぎが示す
周辺分布の際に関する一考察

森 達哉† 川原 亮一‡

† NTT 情報流通プラットフォーム研究所

‡ NTT サービスインテグレーション基盤研究所

2001年 11月16日

信学会IN 11月研究会

背景 (1/2)

- トラフィックモデルが必要な理由
 - より正確な品質評価
 - サービスレベルマネジメント
 - 正しい制御、予測
 - 効率的なネットワーク設計
プロビジョニング

では実トラフィックの特性は？

背景: (2/2)

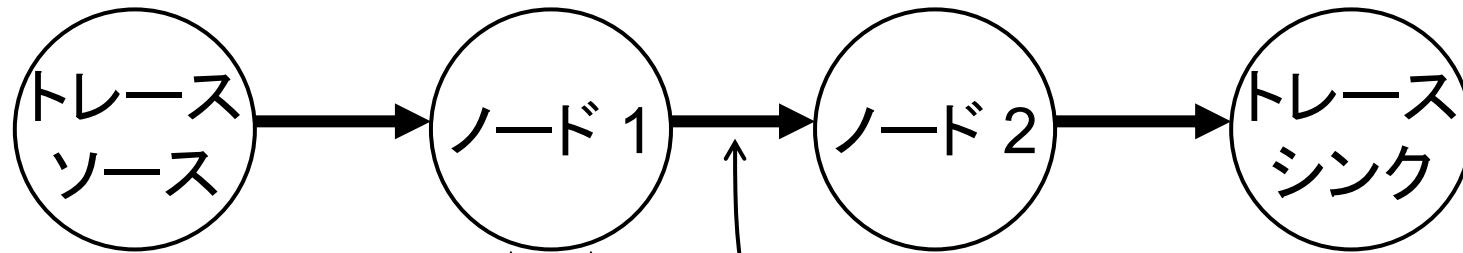
- データ通信トラフィックの変動特性
 - 自己相似性
 - スケール不変のバースト性
 - 長期自己相関
 - 自己相関関数のパワー則

(1990年代初頭~)

これらの特性はハースト指数によって評価できる。
が、品質評価の観点から十分なパラメタといえるか？

ハースト指数とネットワーク品質: (1/3)

- トレース駆動型シミュレーション



パケットスケジューリング = FIFO
バッファサイズ = 50 pkts

使用率 = 0.6 を満たすように
トレース毎に帯域を設定

パケットロス率のモニター

ハースト指数とネットワーク品質: (2/3)

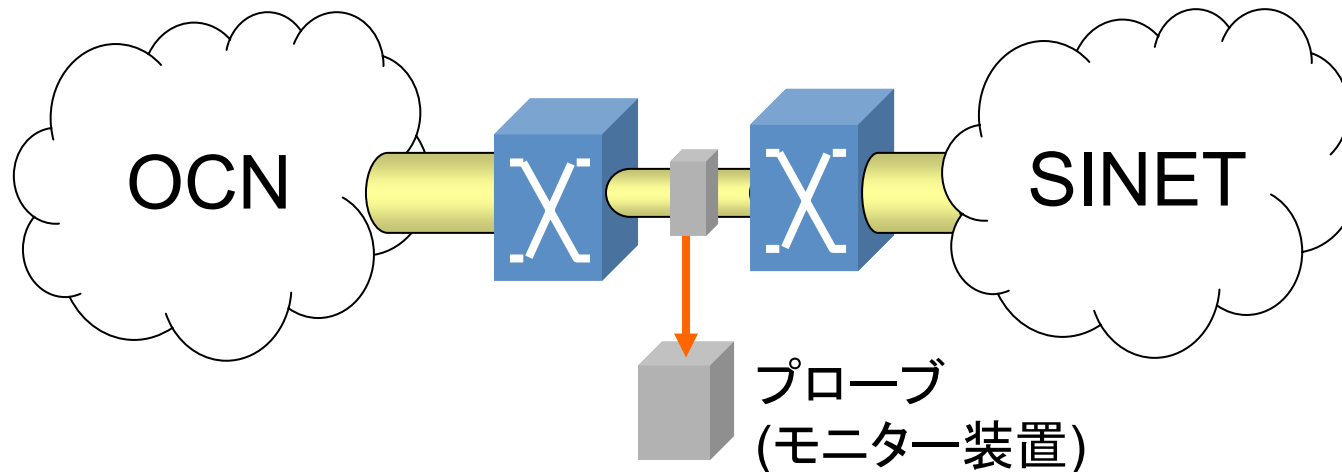
- トレースデータ

OCN-SINET間バックボーントラフィック

135-Mbps ATMリンク (IP over ATM)

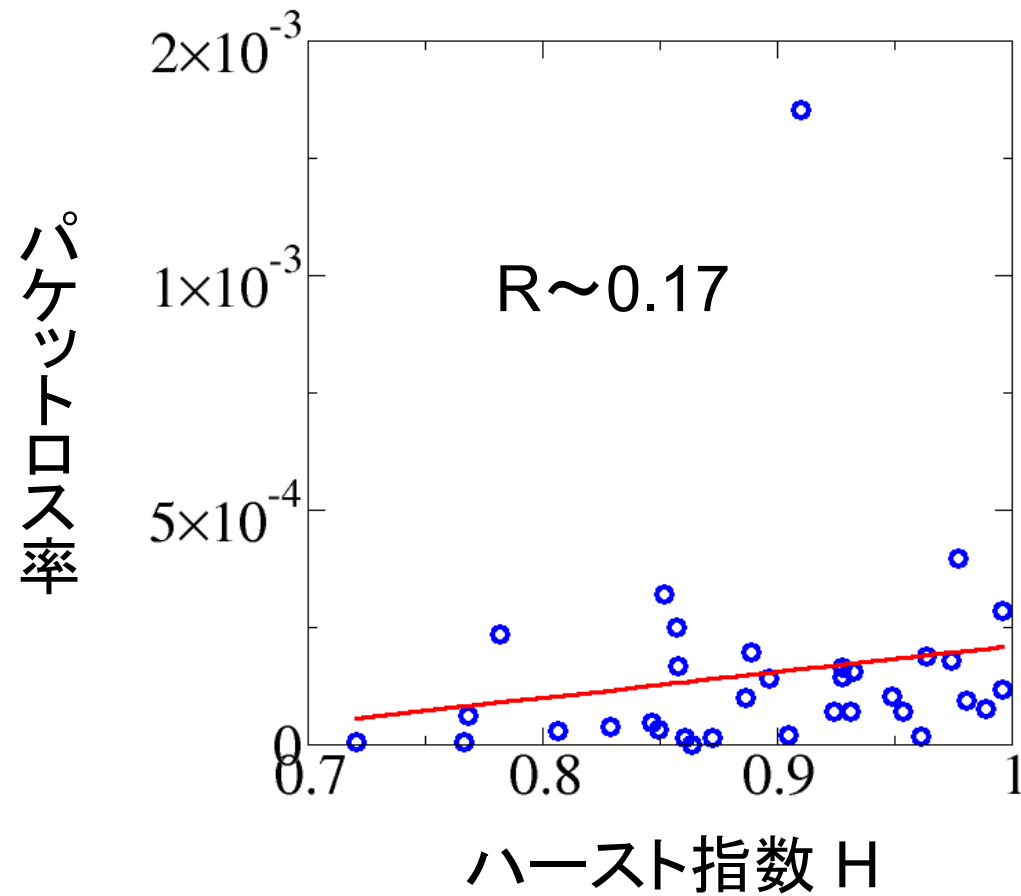
平日最繁期 300秒 x 34 トレース

[Kawahara 2001]



ハースト指数とネットワーク品質: (3/3)

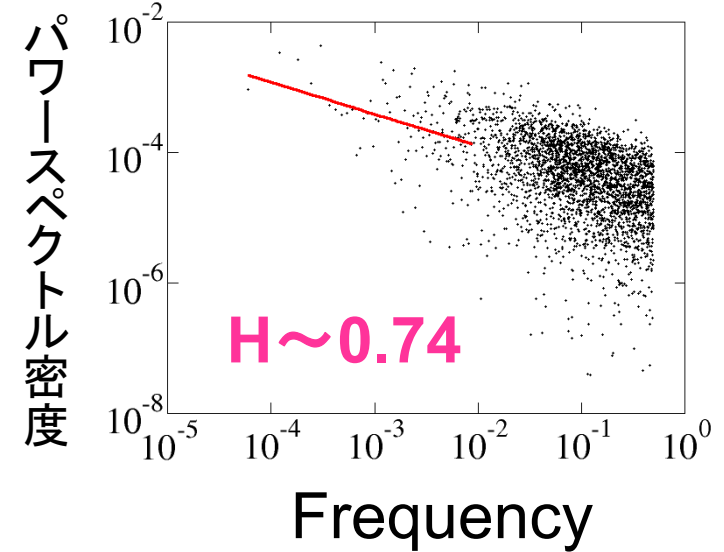
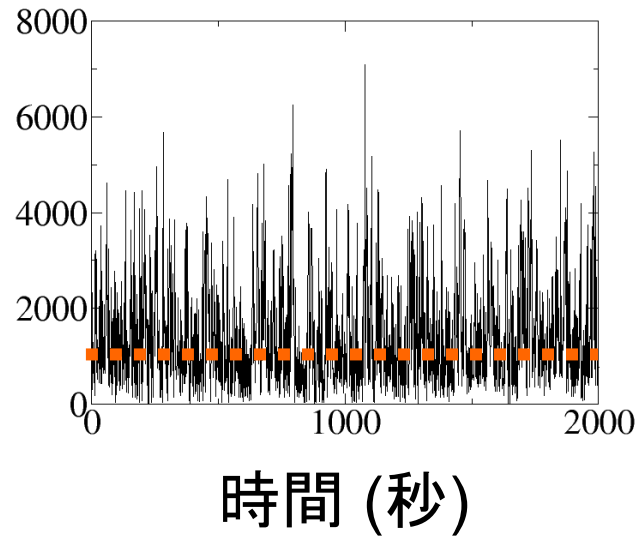
シミュレーション結果



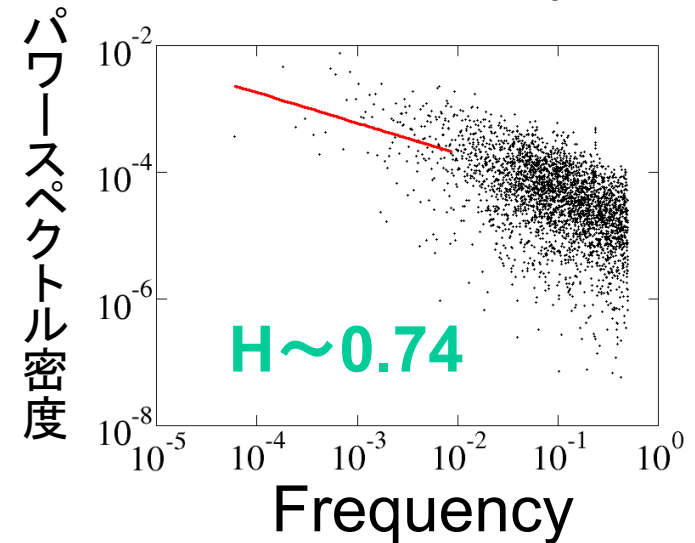
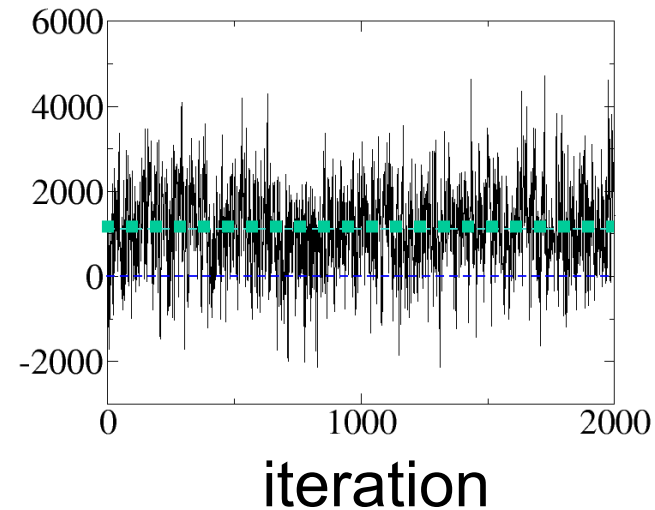
同様なハースト指数を持つ 時系列データの例:

Bellcore LAN
[Leland et.al 1993]

スループット
(kbps)

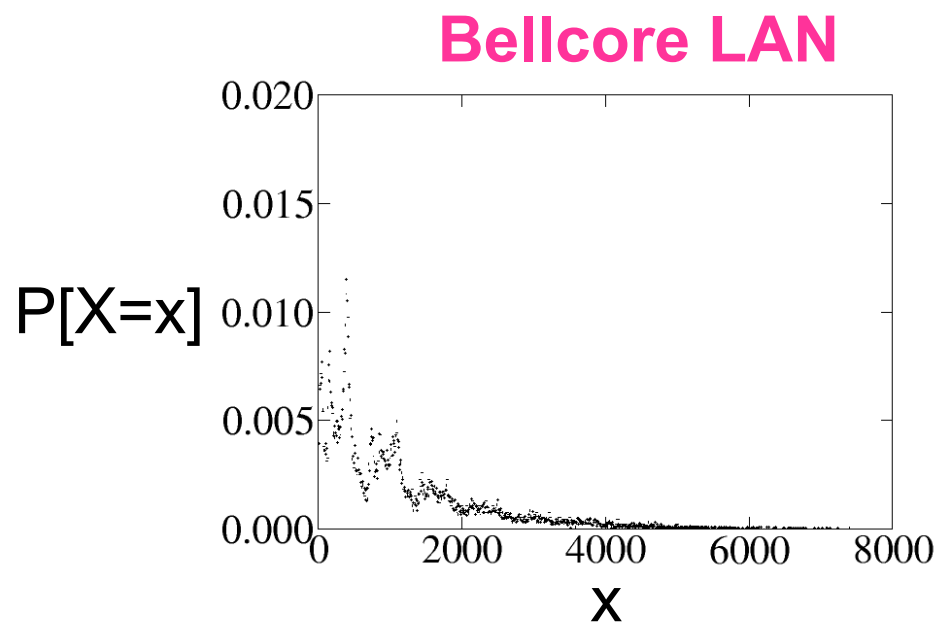


fGN

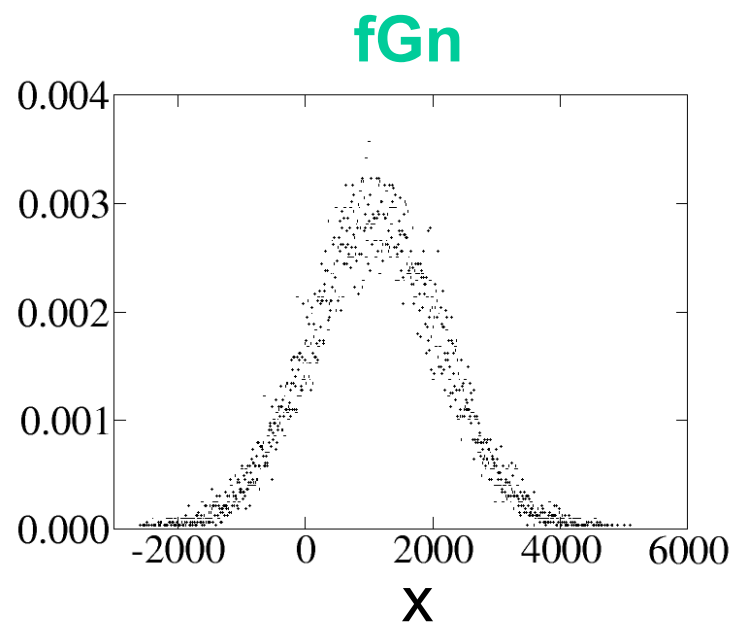


差異は何か？

- 平均値周りの変動
– i.e., 周辺分布



非ガウス分布



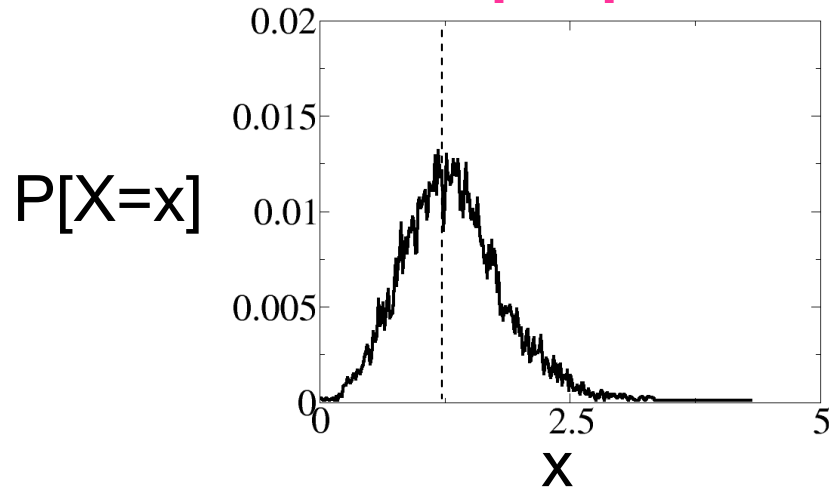
ガウス分布

周辺分布

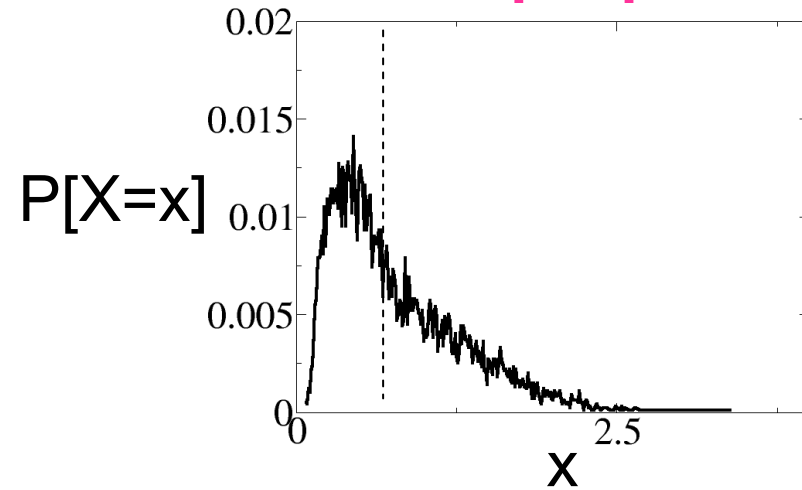
- ガウス分布
 - (一般に)フローの集約 → 中心極限定理
 - (LRDの場合)
 - ex) Pareto On/Offソースの集約
 - **fBm** (fractional Brownian motion)
 - ブラウン過程の拡張
 - i.e., 周辺分布はガウス分布
- 非ガウス分布
 - 実トラフィックの場合、このようなケースが多い
 - 適当な物理的解釈はいまのところ無い
 - (候補: Levy 分布?)

トラフィック変動の示す非ガウス分布の例:

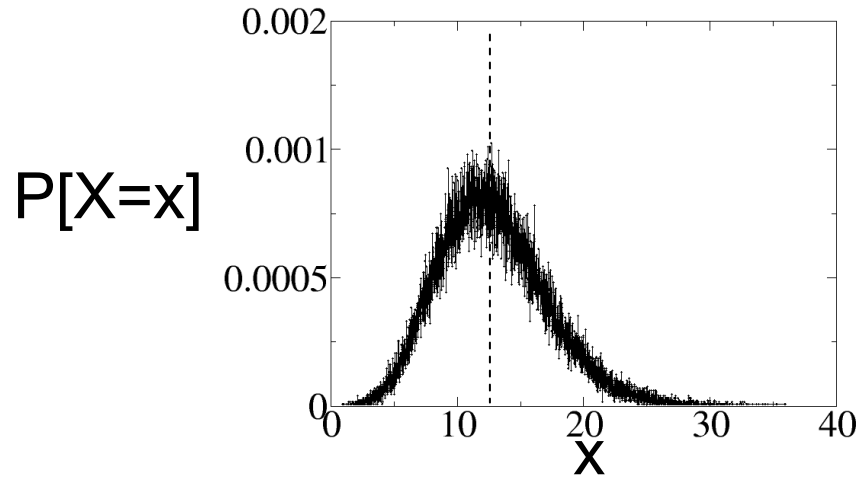
WIDE sample point A-a



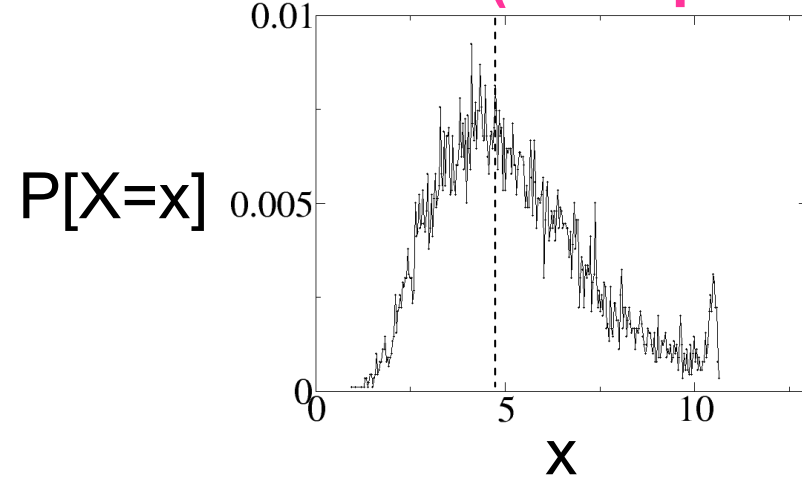
WIDE sample point A-b



OCN-SINET (OC-3)



ECL-OCN (12Mbps ATM)

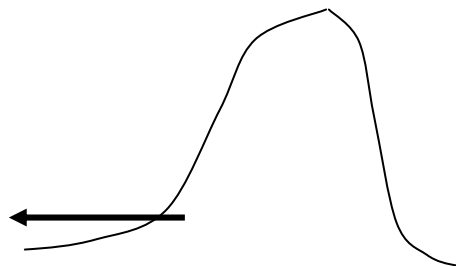


周辺分布の差異の尺度(1/2):

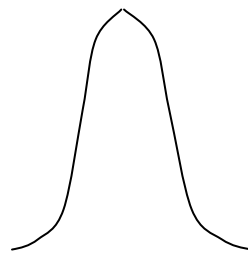
- 3次モーメント → skewness (歪度)

$$\text{skewness} = \frac{\langle (X - \langle X \rangle)^3 \rangle}{\sigma^3}$$

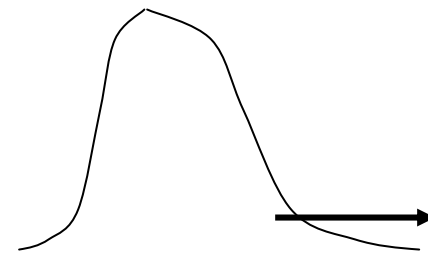
σ : 分散



skewness < 0



skewness = 0
(symmetric)



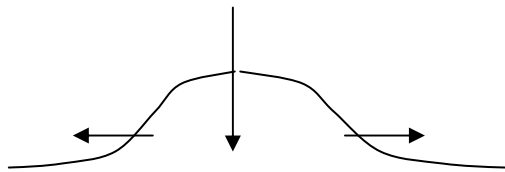
skewness > 0

周辺分布の差異の尺度(2/2):

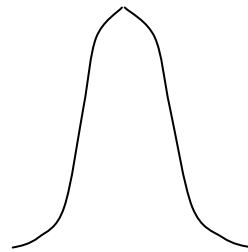
- 4次モーメント → kurtosis (尖度)

$$\text{kurtosis} = \frac{\langle (X - \langle X \rangle)^4 \rangle}{\sigma^4} - 3$$

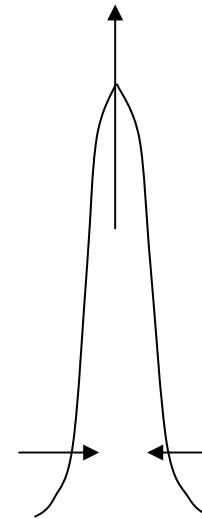
σ : 分散



kurtosis < 0

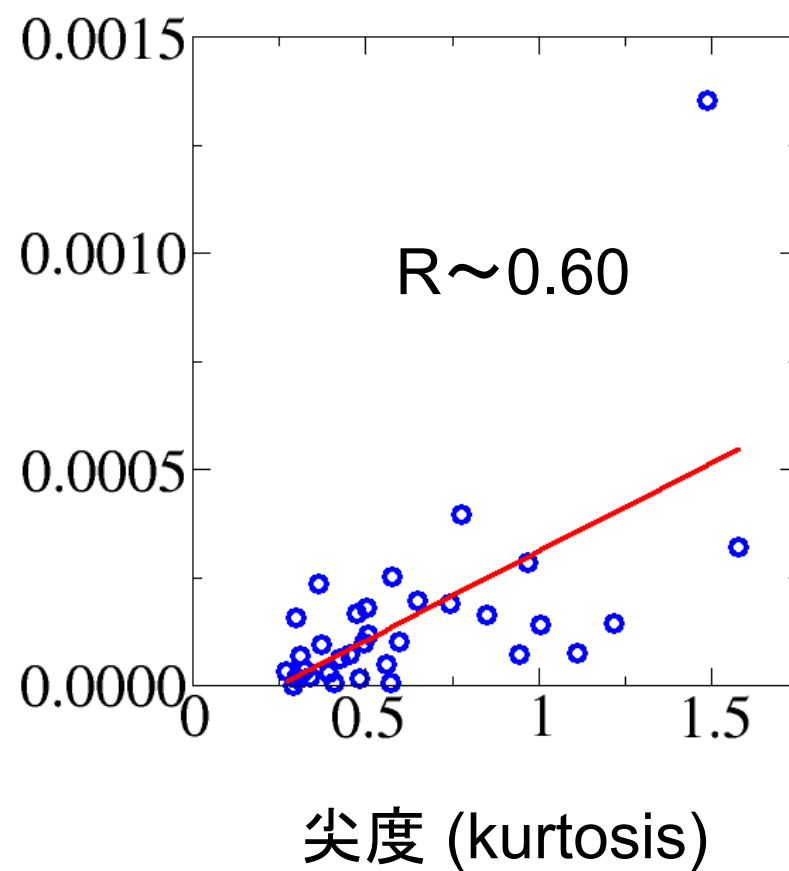
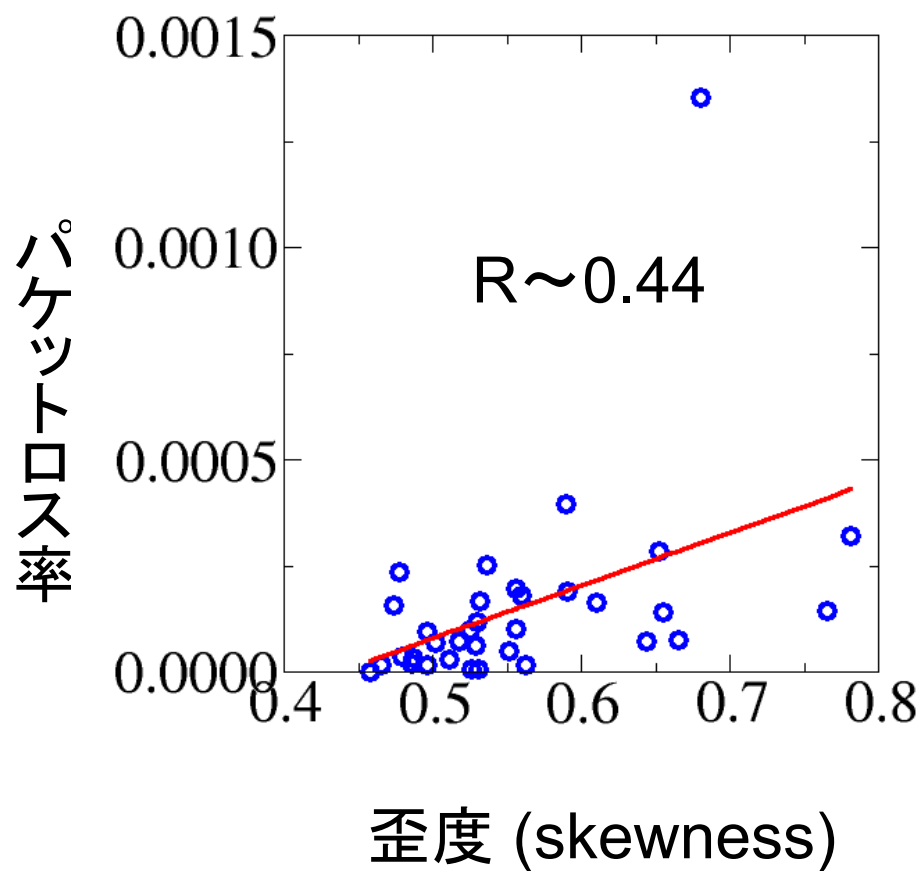


kurtosis = 0

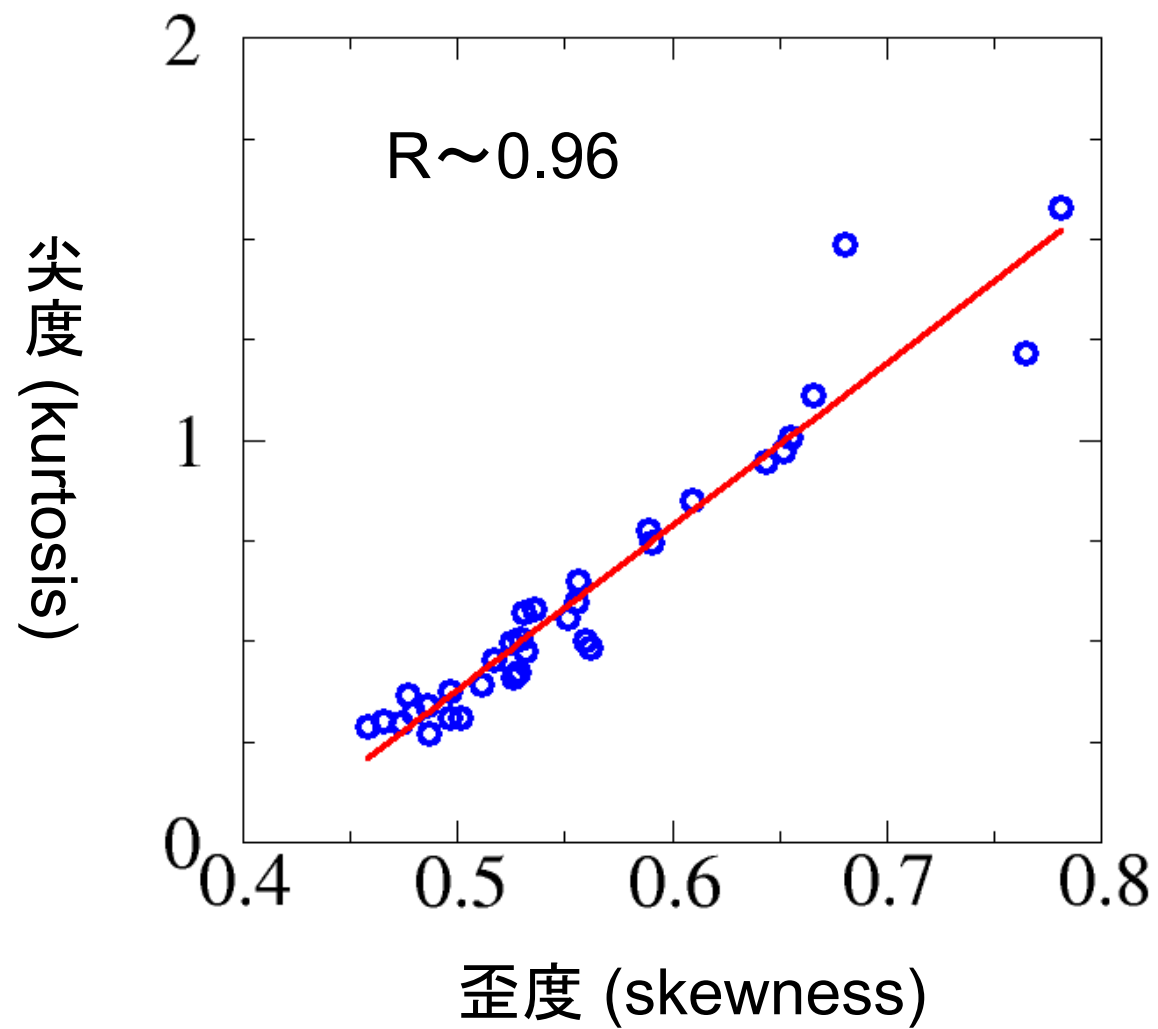


kurtosis > 0

周辺分布とネットワーク品質: (1/3)



周辺分布とネットワーク品質: (2/3)



周辺分布とネットワーク品質: (3/3)

- OCN-SINETトレース
 - 大半は非ガウス分布
i.e., skewness > 0 , kurtosis > 0

- 歪度、尖度はネットワーク品質と正の相関をもつ
(ハースト指数に対する相関より大)
 - 品質評価の観点からはハースト指数だけでは特性を捕らえるのに**十分ではない**
 - 全く同じ周辺分布の場合、ハースト指数が効くと考えられる
(ガウス分布の場合はよく知られている)

なぜ非ガウス分布なのか？

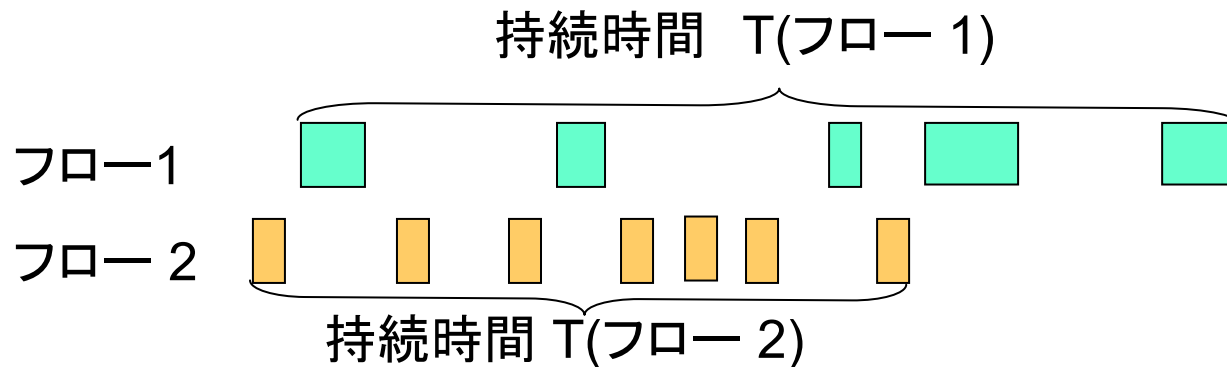
- 集約トラフィックの構造を考える
 - 重ねあわせ(集約)が十分でない？
 - 中心極限定理を満たしていない？
→ 平均アクティブフロー数
 - フロー毎の振る舞い
 - 独立同分布ではない？
→ 単位時間毎のフローサイズ分布

IP フロー統計

- IPフローの定義

- 以下のユニークな組み合わせを持つパケット群

<送信IPアドレス> <受信IPアドレス>
<送信ポート番号> <受信ポート番号>
<プロトコル>



平均アクティブフロー数

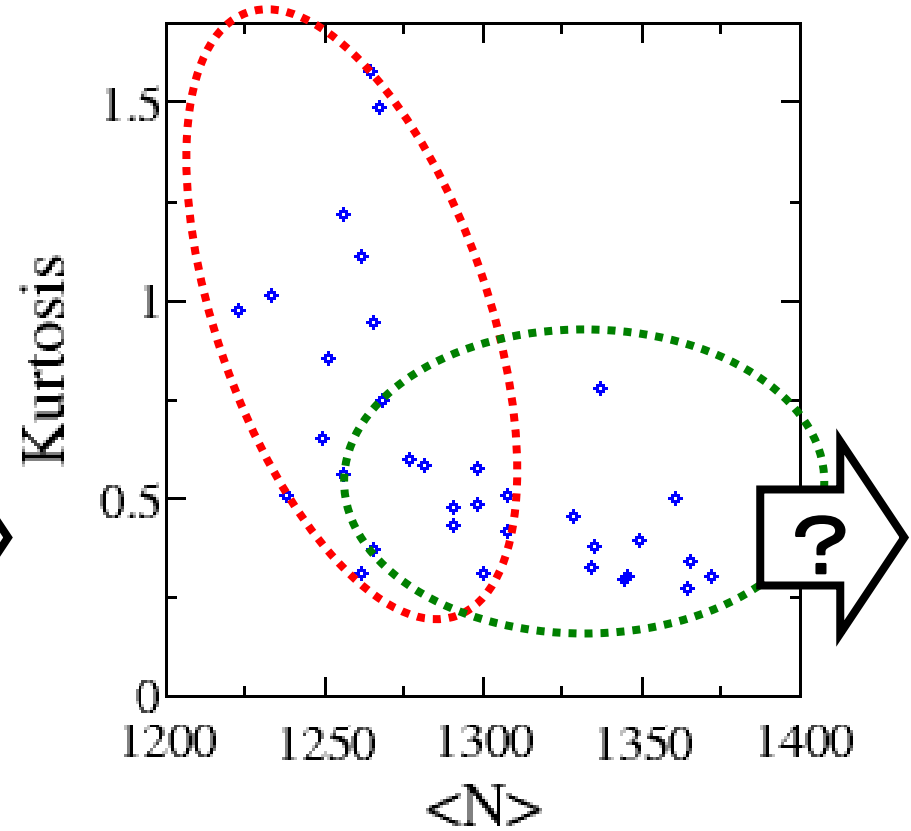
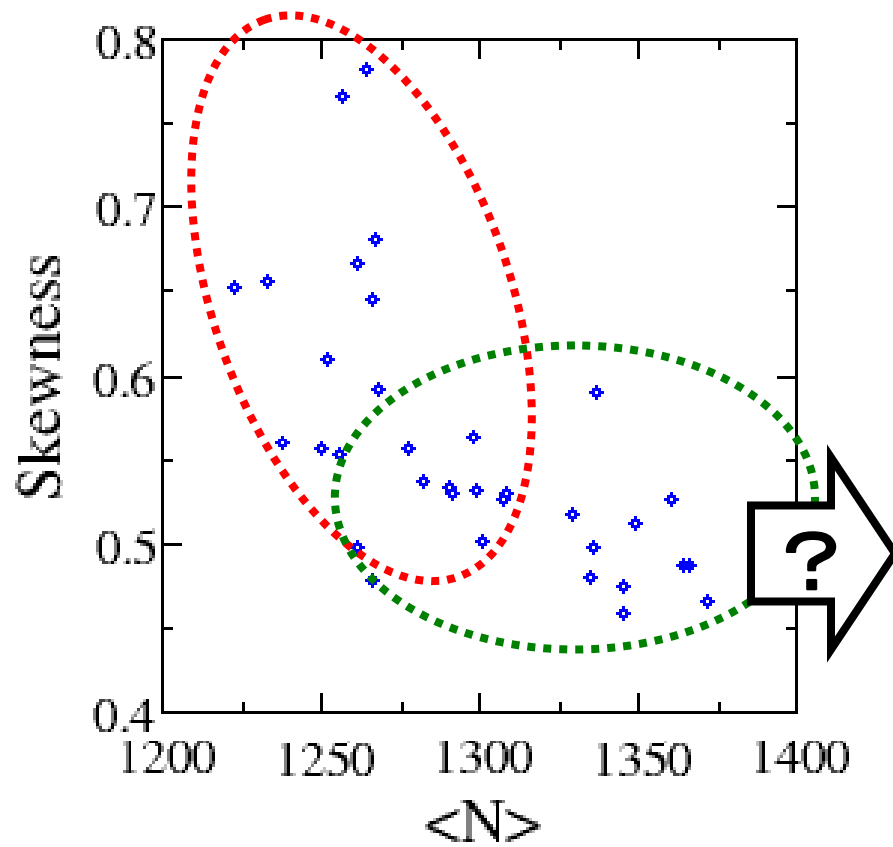
- 平均アクティブフロー数 $\langle N \rangle$
 - 統計多重度を見る
 - リトルの公式

$$\langle N \rangle = \lambda \langle T \rangle$$

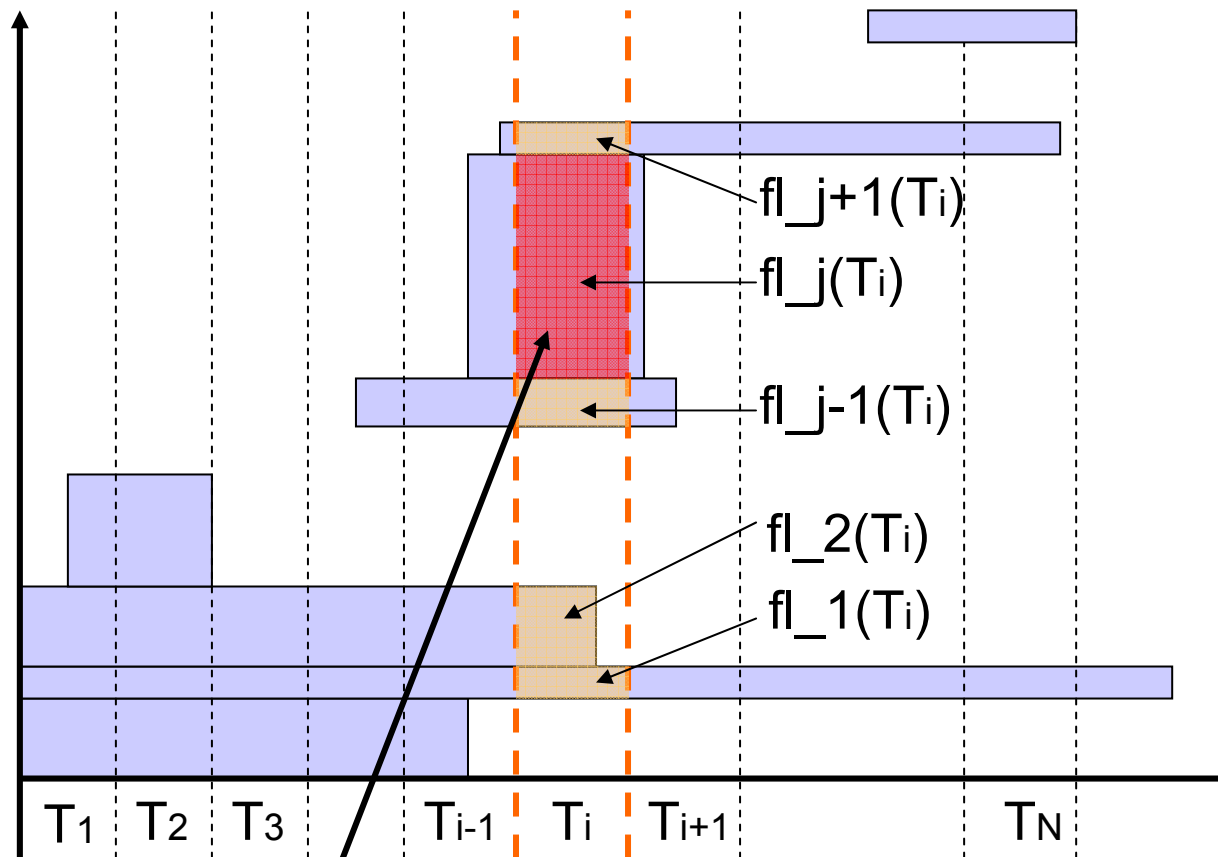
λ : 平均フロー到着率

$\langle T \rangle$: 平均フロー持続時間

平均アクティブフロー数

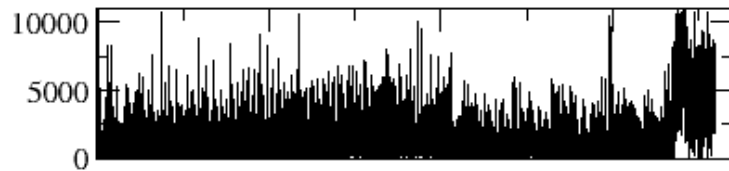


単位時間毎のフローサイズ分布

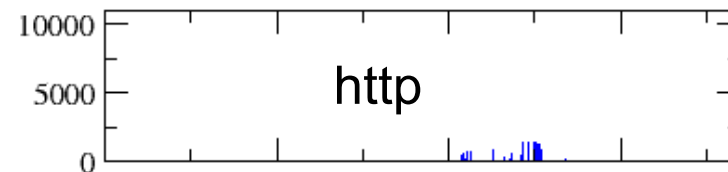
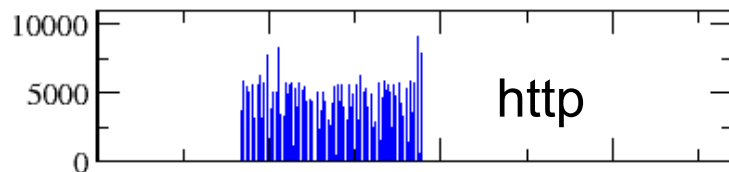
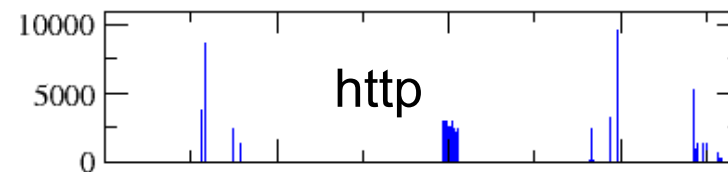
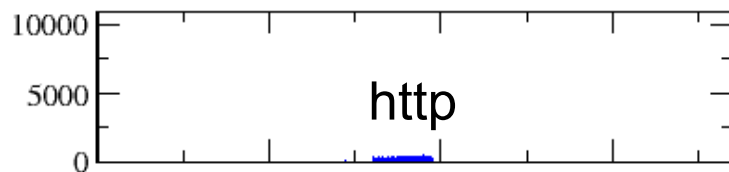
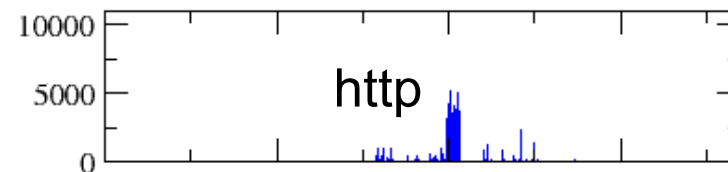
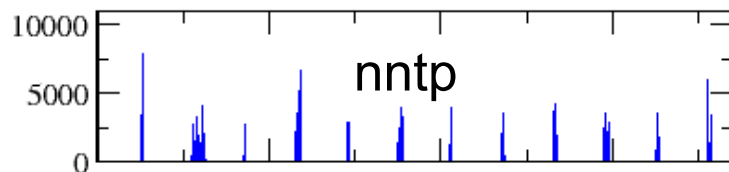
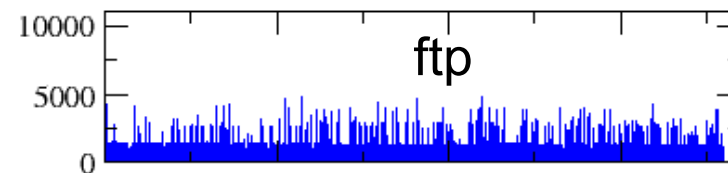
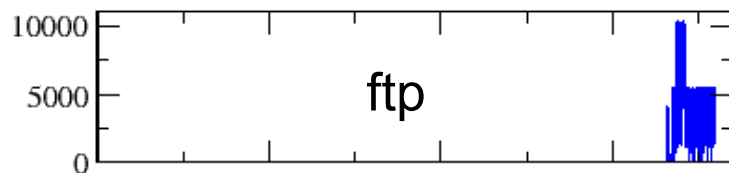


Greedy flow

フローの実際



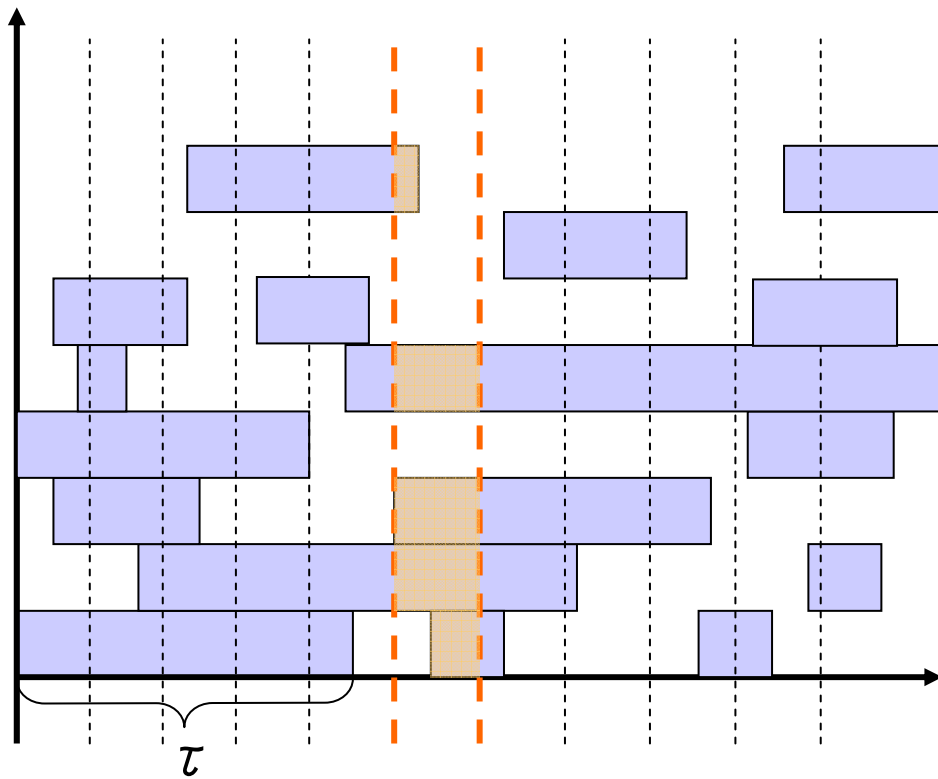
トータル



370秒

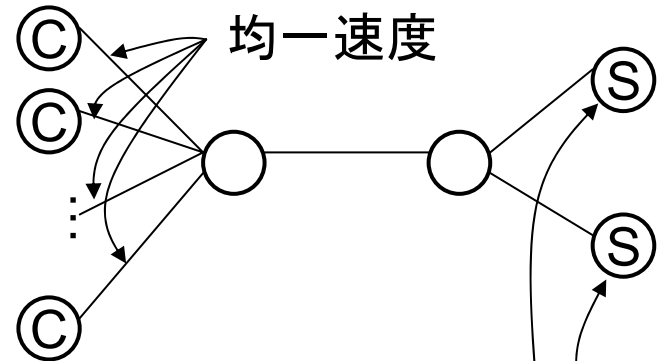
370秒

[cf] traditional Pareto ON/OFFモデル



$$P[\tau > t] \approx t^{-\alpha}$$

[Park 96]



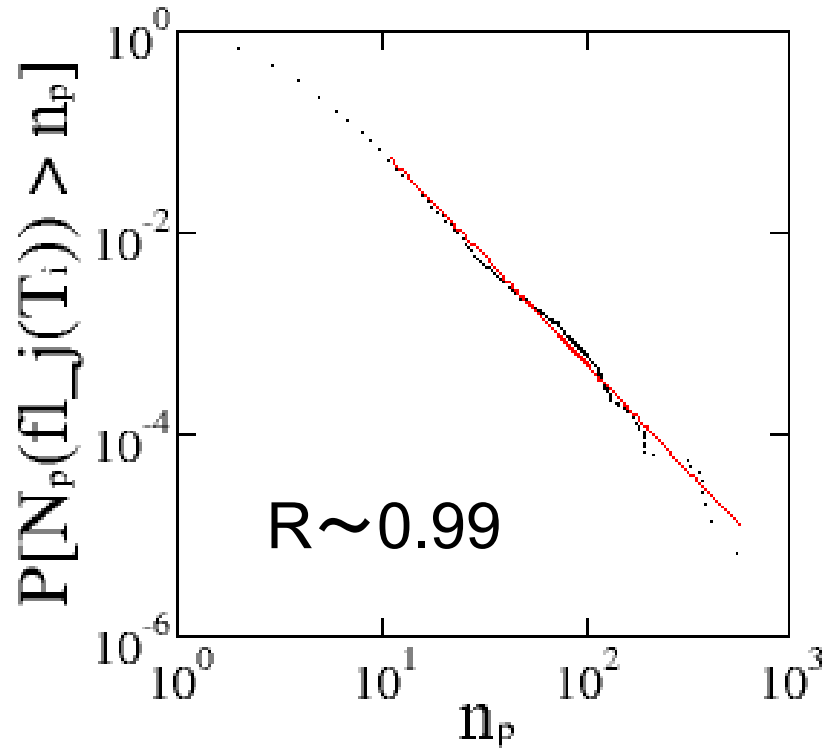
File size dist.

$$P[X > x] \approx x^{-\alpha}$$

No greedy Flow!

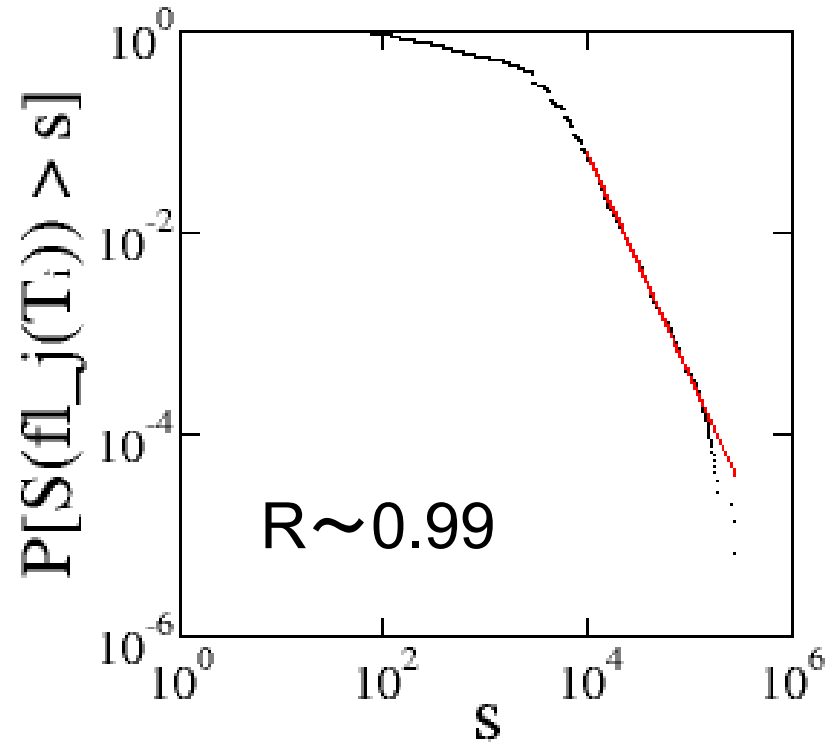
単位時間毎のフローサイズ分布

パケット数



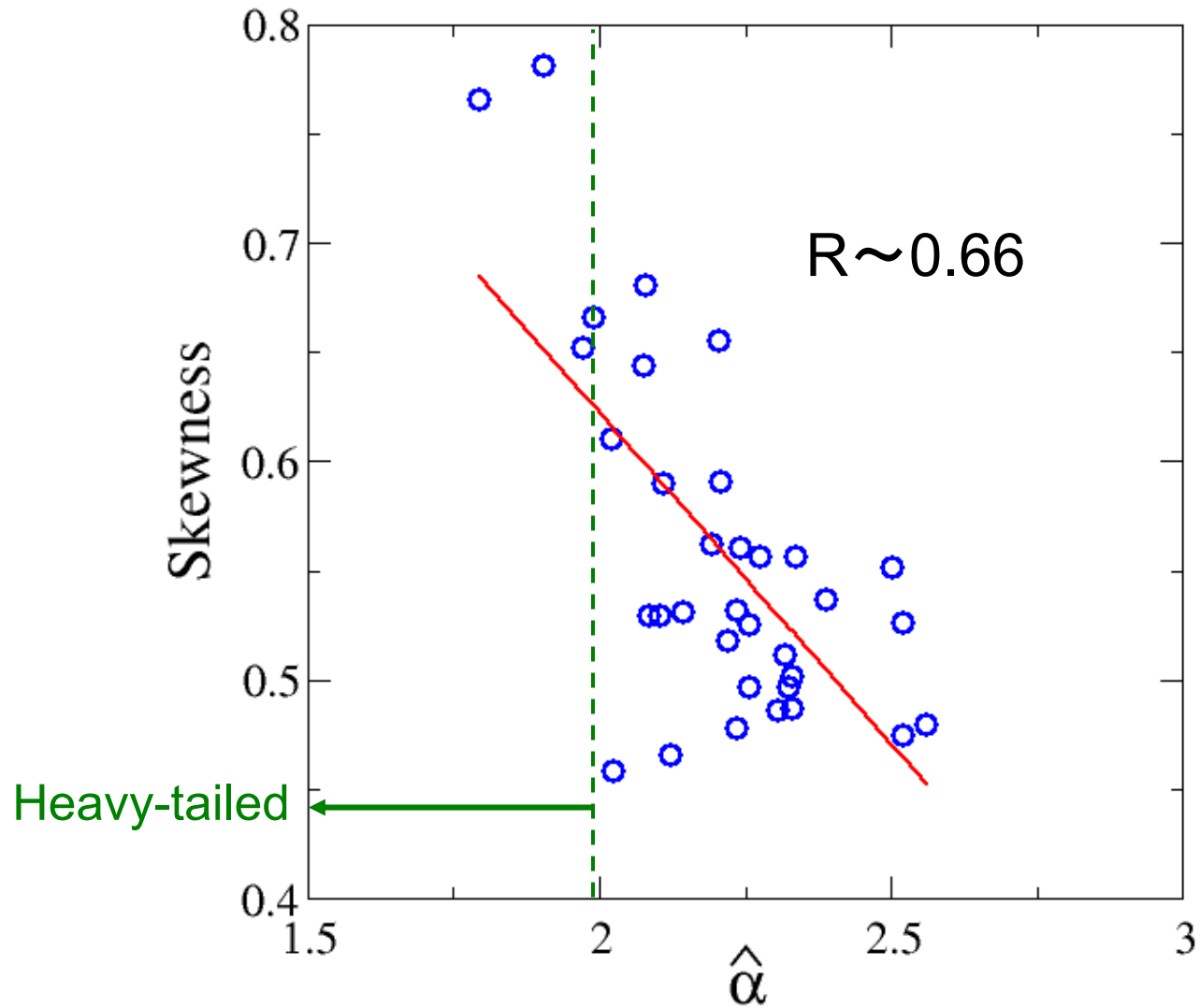
$$P[N_p(fl_j(T_i)) > n_p] \approx n_p^{-2.14}$$

パケットサイズ (byte)



$$P[S(fl_j(T_i)) > s] \approx s^{-2.198}$$

フローサイズ分布と周辺分布の関係



まとめ

- トラフィック変動の周辺分布
 - ネットワーク品質に大きな影響
 - 非ガウス分布
- 平均アクティブフロー数
 - フロー数増加により必ずしもガウス分布に収束しない
ex) アクセス系のブロードバンド化
- 単位時間毎のフローサイズ分布
 - パワー則
 - If heavy-tailed ($\alpha < 2$) → 周辺分布の歪度増加

Future work

- 非ガウス分布の偏在性
 - 更なるデータの収集、解析が必要！
 - WIDE, OCN-SINET, ECL-OCN, Auckland, Bellcore等で確認済み
- 非ガウス分布の原因
 - 統計多重が足りない？
 - だとしてどの程度必要か？
 - ガウス分布以外に収束している？
 - Levy分布？
- 非ガウス分布モデルの提案
 - 任意の周辺分布を持つ自己相似過程？
 - パワー則に従うサイズ分布を持つフローの集約？

Who is Greedy Flow?

- Application-based classification
 - Web?
 - Napster, Gnutella, WinMX,...?
 - Peer-to-peer → \exists strong-to-strong peer
 - DDOS attack?
- Network topology-based classification
 - Hop count (from TTL field) 30,32,60,64,128,255
 - Passive OS detection technique
 - Uplink or Backbone?