# Cascaded On Off Processes

Discussed 11/10/00 & 12/8/00

Vidyadhar Kulkarni, Steve Marron & Juhyun Park

# Goal

Find a model for individual TCP Session traces, that:

1. “Looks right”
2. Gives “correct” statistical properties (dependence, …)
3. Aggregates “correctly” (scaling, multifractal, …)
4. Fits easily into queueing analysis.
5. Makes “physical” sense

# Cascaded On-Off Process

Ideas:

- each packet is a "rapid burst" (on times)

- waiting times (off times) in between are very diverse
  (orders of magnitude different)

# Mathematical Formulation

I. Independent On – Off Processes, $X_1(t), X_2(t), \ldots$

where $X_n(t)$ is

"on" for exponential times, with rate $2^{n-1}\lambda$

"off" for exponential times, with rate $2^{n-1}\mu$

Show 1st row of CascOnOffDemo1.ps

# Mathematical Formulation (cont.)

II. Vary the "gap distribution" by multiplying:

$$Y_n(t) = \prod_{i=1}^{n} X_i(t)$$

III. Normalize to keep overall expected value the same:

$$Z_n(t) = meanrate\left(\frac{\lambda + \mu}{\mu}\right)^n Y_n(t)$$

Show lower left of CascOnOffDemo1.ps

# Mathematical Formulation (Cont.)

IV. Cumulative $Z_n(t)$ has “physical interpretations”:

Show lower right of CascOnOffDemo1.ps

- constant “on rates” (reflects “link capacity”)
- session is “off” unless “all nodes are on”
- wide range of “off times”

# Fit Model to Data

Idea: use

- “peak rate” $= r_{peak} = 155 * 10^6 \ (bits/\mathrm{sec})\big/ 8 \ (bits/byte)$

- $N$ number of packets in trace

- $T_i(t)$ time stamp (secs) of $i$-th packet

- $S_i(t)$ size (bytes) of $i$-th packet

to estimate parameters: $\lambda, \ \mu, \ n$

# Microscopic View of Model

Show ToyJumpCDF.ps

- "packet transmission time" give peak rate, $r_{peak}$
- $r_{peak}$ drives "on" and "off" distribution relationship
- does peak rate affect "macroscopic shape"???

# Parameter Estimation 1

For a given value of the level $n$

1. "Get total size right", i.e: est. the "mean rate", $r_{mean}$, by

$$\hat{r}_{mean} = \frac{\sum_{i=1}^{n} S_i}{T_N} = \frac{\text{"}Total\ Size\text{"}}{\text{"}Total Time\text{"}}$$

2. "Make jumps right", i.e: est. the "mean on time", $\tau_{on}$, by

$$\hat{\tau}_{on} = \frac{\hat{r}_{mean}}{r_{peak}} \cdot \frac{T_N}{N} = \text{"}prop'n\ on\text{"} \cdot \text{"}time / packet\text{"}$$

# Parameter Estimation, 2

Still for a given value of the level $n$

3. "Time conservation" gives the "mean off time", $\tau_{off}$, as:

$$\hat{\tau}_{off} = \frac{T_N}{N} - \hat{\tau}_{on} = "time/packet" - "mean\ on\ time"$$

4. Solve rate equations to get:

$$\hat{\lambda}_n = \frac{1}{\hat{\tau}_{on}\left(2^n - 1\right)}$$

$$\hat{\mu}_n = \frac{\hat{\lambda}_n}{\left(\hat{\tau}_{off}/\hat{\tau}_{on} + 1\right)^{1/n} - 1}$$

# Theoretical Variance

Idea: for given $n, \lambda, \mu$ calculate:

$\sigma^2_{Off}$ “variance of off-times dist’n”

View: for several $\lambda$ and $\mu$, study $\sigma^2_{Off}(n)$

Show CascOnOff2ndMT2.ps and CascOnOff2ndMT3.ps

- $\sigma^2_{Off}$ can increase or decrease with $n$ (inc. for us?)
- exponential relationship, log scale is natural?

# Estimation of Cascade Level $n$

Idea: "variance matching"

1. Consider range of $n$ values.
2. Estimate $\hat{\lambda}_n, \hat{\mu}_n$ as before.
3. Compute $\sigma^2_{Off}(n, \hat{\lambda}_n, \hat{\mu}_n)$
4. Choose $\hat{n}$ to "match" $\sigma^2_{Off}(n, \hat{\lambda}_n, \hat{\mu}_n)$ with sample var $s^2_{Off}$

# Application to usual 10 traces

Show CombineCascOnOffData2p1t1.pdf

- Looks great?
- Estimated $n$ is 6 – 12
- Computation time: few minutes – 4 hours
- Peak Session 1 is worst. How bad?
- Still miss “shape”, e.g. “TCP Slow Start”

# How “close” are simulated traces to raw data?

1. Autocorrelations (recall Cons’tive Cascades failed here)

Show CombineCascOnOffData2p3t1.pdf

- Autocorr. for trace shown in red

- Different from 0? Overlay 100 sim’d versions (independent, ie. Autocorr $\equiv 0$, by construction)

- Many follow model

- Off Peak 5: significant at lag 4???

- Peak Session 1: lots of negative corr.???

# Autocorrelations (Cont.)

- Peak Session 2: lag 1 positive corr.???

- Peak Sessions 3,4,5: sign't $\rho > 0$, at lags 6, 12,18, due to "TCP Windowing seasonal effect"?

- Explained by looking at "rates"?

Show UncSessionData\CombineSessionData1p11.pdf

Should look at "time series of off times"
Could try "Vper analysis"

Conclusion: model generally holds up
(with room for "fine tuning", esp. w.r.t. TCP Window)

## Aside: careful look at “periodicity”, via GPVPER

Ideas:

i. Extract “best fit seasonal components” of time series
ii. Study residuals
iii. Study decomposition of “sums of squares”

Show CombineCascOnOffData2p4t1.pdf, in order p9,p10, p8, others

Results:

- Peak Sess. 4: strong periodic component, lag 6
- Peak Sess. 5 & 3: moderate per’c comp’t, lags 6 & 17
- Most others: no strong periodic component
- Peak Sess. 1: Could look at “Frequency Modulation”?

# How “close” are simulated traces to raw data? (Cont.)

2. Summary statistics

Ideas:

a. look simultaneously at *many* of these

b. Assess where “data trace” lies in “simulated bundle”

c. Visualize via “parallel coordinate plots”

Show CombineCascOnOffData2p2t1.pdf or maybe individual files CascOnOffData2p2d1t1nn10.ps - CascOnOffData2p2d10t1nn8.ps

# Summary statistics (Cont.)

First round of statistics:

i. Mean of Off Times
ii. Median of Off Times
iii. Standard Deviation of Off Times
iv. MAD (Median Absolute Deviation) of Off Times
v. Mean of {Off Times $<$ Median}
vi. Median of {Off Times $<$ Median}
vii. S. D. of {Off Times $<$ Median}
viii. MAD of {Off Times $<$ Median}
ix. Maximum of Off Times
x. $2^{nd}$ Maximum of Off Times
xi. $3^{rd}$ Maximum of Off Times
xii. $4^{th}$ Maximum of Off Times

# Summary statistics (Cont.)

General ideas:

- {Off Times < Median} is population of "off times while packets are moving rapidly"

Show CombineCascOnOffData1p6t1.pdf

- Mean and Median relationship reflects "skewness"

- S.D. and MAD relationship reflects "kurtosis"

- Stat's are "standardized", i.e. (Stat. – mean) / SD

- Thus comparisons are relative

- Could try "better" ordering of parallel coords?

# Summary statistics (Cont.)

Observations:

Off Peak 1:

- Main Dist’n OK
- Rapid Dist’n has wrong skewness and kurtosis
- 1st two maxima OK, but too big afterwards

Off Peak 2:

- Main Dist’n OK
- Rapid Dist’n too small and tight
- 1st two maxima OK, but too big afterwards

Off Peak 3 (and 4):

- Main Dist’n has some skewness and kurtosis
- Rapid Dist’n has bad skewness and kurtosis
- Maxima well modeled

# Summary statistics (Cont.)

Off Peak 5:

- Main Dist’n has strong skewness
- Rapid Dist’n extremely different
- 1st two maxima OK, next too big

Peak 1:

- Main Dist’n skewed and kurtotic
- Rapid Dist’n too big
- maxima all OK

Peak 2:

- Main Dist’n OK
- Rapid Dist’n too small and too tight
- maxima all OK

# Summary statistics (Cont.)

Peak 3:

- Main Dist'n has some skewness and kurtosis
- Rapid Dist'n too small and too tight
- 1st maxima too small, others too big

Peak 4:

- Main Dist'n OK
- Rapid Dist'n too tight
- 1st two maxima OK, others too big

Peak 5:

- Main Dist'n has skewness and kurtosis
- Rapid Dist'n too big on average
- all maxima too big

# Summary Statistics, Conclusions

1. Useful “global summary method”?
2. Quantifies  and enhances “what we can see”?
3. Main distributions “pretty good”?
4. Rapid distributions less so?   (does this matter?)
5. Maxima “pretty good”?
6. Estimation “feels largest max” too much?

## How “close” are simulated traces to raw data? (Cont.)

3. Quantiles:

Idea: for a vector of “probs”: $\alpha = 0.02, 0.04, ..., 0.98$

compare corresponding “real trace quantiles” $q_\alpha$, with “simulated trace quantiles” $\hat{q}_\alpha$

Show CascOnOffData2p5d1t1nn10.ps … CascOnOffData2p5d5t1nn10.ps

Comparison 1: “Normalized quantiles” $\dfrac{q_\alpha - mean(\hat{q}_\alpha)}{sd(\hat{q}_\alpha)}$

Comparison 2: “Q-Q plot” (real vs. simulated)

# Quantile Comparison

Main Lessons:

- Good approximation for large off times
- Poor approximation elsewhere (especially very small)
- Consistent with “good visual effect”

# Simulation of Estimation

Idea: better understand estimation process

1. For real traces, estimate $\hat{n}$, $\hat{\lambda}_{\hat{n}}$ and $\hat{\mu}_{\hat{n}}$ as above.

Show top of CombineCascOnOffData2p1t1.pdf

2. Simulate 100 traces, as above.

Show bottom right of CombineCascOnOffData2p1t1.pdf

3. Get 100 simulated estimates, $\hat{\hat{n}}$, $\hat{\hat{\lambda}} = \hat{\hat{\lambda}}\left(\hat{\hat{n}}\right)$ and $\hat{\hat{\mu}} = \hat{\hat{\mu}}\left(\hat{\hat{n}}\right)$, from simulated traces.

# Simulation of Estimation, (Cont.)

A. For computational speed, restricted $n \leq 10$

B. Compare sim'd $\hat{\hat{\lambda}}$ (red), and $\hat{\hat{\mu}}$ (blue), with "true values" $\hat{\lambda}$ (yellow) and $\hat{\mu}$ (green).

Show CombineCascOnOffData5.pdf, upper left

i. Sometimes "est's too big": when $\hat{n} \geq 10$

Show CombineCascOnOffData2-5Big.pdf

ii. Sometimes "est's about right": when $\hat{n} = 9$

Show CombineCascOnOffData2-5OK.pdf

iii. Sometimes "est's too small": when $\hat{n} \leq 7$

Show CombineCascOnOffData2-5Small.pdf

# Simulation of Estimation, (Cont.)

C. Looks like "clusters" - investigate with k.d.e:

Show CombineCascOnOffData5.pdf, upper right

-"factor of 2 between peaks"?

D. Joint distributions of $\hat{\hat{\lambda}}$ and $\hat{\mu}$

Show CombineCascOnOffData5.pdf, lower left

- usually have strong relationship (not surprising)
- don't lie on a line???

# Simulation of Estimation, (Cont.)

E. Joint distribution of $\hat{\hat{\tau}}_{on}$ and $\hat{\hat{\tau}}_{off}$:

Show CombineCascOnOffData5.pdf, lower right

- more “independent” than $\hat{\hat{\lambda}}$ and $\hat{\hat{\mu}}$.

- So is this a “better parametrization”?

- Simulated $\hat{\hat{\tau}}_{on}$ always << $\hat{\tau}_{on}$

- Simulated $\hat{\hat{\tau}}_{off}$ usually < $\hat{\tau}_{off}$

# Simulation of Estimation, (Cont.)

## F. Investigation of "clustering"

Show CombineCascOnOffData5N.pdf

- Clusters explained by $\hat{\hat{n}}$

- Clustering appears for $\hat{n}$ quite small, $\leq 6$

Show CombineCascOnOffData2-5Small.pdf

- Then have $\hat{\hat{n}}$ mostly >> $\hat{n}$

- Otherwise $n \leq 10$ constraint "takes over"???

- Note: "factor of 2" between peaks

# Explanation of "Factor of 2"

For $n \to \infty$:

$$\hat{\hat{\lambda}} = \frac{1}{\hat{\hat{\tau}}_{on}\left(2^n - 1\right)} = \frac{1}{2^n} \cdot \frac{1}{\hat{\hat{\tau}}_{on}} + o\left(\frac{1}{2^n}\right)$$

$$\hat{\mu} = \frac{\hat{\hat{\lambda}}}{\left(\hat{\hat{\tau}}_{off} / \hat{\hat{\tau}}_{on} + 1\right)^{1/n} - 1} = \frac{1}{2^n} \cdot ??? + o\left(\frac{1}{2^n}\right)$$

Should reparametrize, and work with $\hat{\hat{\tau}}_{on}$ and $\hat{\hat{\tau}}_{off}$???

## Alternative Parameterization:

$$\lambda^* = 2^n \cdot \lambda, \qquad \mu^* = 2^n \cdot \mu$$

{or $\left(2^n - 1\right)$?}

- Will make “cluster” disappear????

- Then can formulate and address “bias” problems?

- Still need to tackle problems with bias in $\hat{n}$???

# New Results

1. Better variance calculation gives higher $n$ range

Show CombineCascOnOffData3p1sim.pdf

2. Bigger $n$ range labelling of clusters

Show CombineCascOnOffData3p3.pdf

- Bigger $\hat{n}$ bias for Peak (2-3), than Off-Peak (1-2)???

3. Search for reason behind $\hat{n}$ bias

Show CombineCascOnOffData3p4sim.pdf