Fractional Sum-Difference Models for Open-Loop Generation of Internet Packet Traffic

Jin Cao, Bill Cleveland, and Yuan Gao
{cao, wsc, yuangao}@bell-labs.com

Don Sun
dxsun@optonline.com

FSD (Fractional Sum-Difference) Time Series Models

Statistical models for packet traffic arrivals and sizes

Extensive validation

Open-loop generation of arrivals and sizes

- only need to specify
  - traffic load (packets/sec, bits/sec, or number of active connections)
  - parameters of packet size marginal distribution (we give plenty of guidance)
- C code to be available soon
- implemented in NS-2

Useful in general for engineering studies

FSD (Fractional Sum-Difference) Time Series Models

Let x_u be a time series for u = 1, 2, . . .

Let $F(x; \phi)$ be the marginal cumulative distribution function of x_u where ϕ is a vector of unknown parameters

Let $G(z)$ be c.d.f of normal distribution with mean 0 and variance 1

Let $z_u = T(x_u; \phi) = G^{-1}(F(x_u; \phi))$ be the transformation of x_u such that the marginal distribution of z_u is normal with mean 0 and variance 1

We suppose

- z_u is a Gaussian time series
- z_u has a specific parametric form
  - linear combination of a simple fractional ARIMA and white noise
- 2 parameters
  - fractional difference exponent: $0 < d < 0.5$
  - variance of noise term: $0 \leq \theta \leq 1$

To generate x_u, generate z_u and use inverse transform $T^{-1}(z_u)$

FSD Time Series Models: The Gaussian Linear Combination

$$z_u = \sqrt{1-\theta}\; s_u + \sqrt{\theta}\; n_u$$

$$0 \le \theta \le 1$$

s_u and n_u are independent of one another and have mean 0 and variance 1

n_u = Gaussian white noise

s_u is a Gaussian fractional ARIMA

$$(I-B)^d s_u = \epsilon_u + \epsilon_{u-1}$$

$$B s_u = s_{u-1}$$

$$0 < d < 0.5$$

ϵ_u is Gaussian white noise with mean 0 and variance
$\sigma_\epsilon^2 = [(1-d)\Gamma^2(1-d)]/[2\Gamma(1-2d)]$

We coined the term “FSD” because

$$(I-B)^d z_u = \sqrt{1-\theta}\; (I+B)\epsilon_u + \sqrt{\theta}(I-B)^d n_u$$

Modeling Inter-Arrivals, t_u: Marginal Distribution

Weibull variable w

- shape parameter λ and scale α
- $(w/\alpha)^{\lambda}$ is exponential

Empirical distribution of t_u very well approximated by Weibull

Let $W(w; \lambda, \alpha)$ be Weibull c.d.f and let $G(z)$ be c.d.f of normal distribution with mean 0 and variance 1

Transformation to t_u after z_u generation is $t_u = W^{-1}(G(z))$

Why do we need this transformation?

- inter-arrivals grossly non-gaussian and non-linear
- inter-arrival marginal very well approximated by Weibull
- replaces scores of wavelet coefficients with a few parameters

Modeling Sizes, q_u: Marginal Distribution

Discrete-continuous distribution

Atoms at finite number k of values such as 40 bytes, 576 bytes, and 1500 bytes

Uniform otherwise from 40 bytes to 1500 bytes

Parameters are probabilities $\phi_1, \ldots, \phi_k$ at atoms

Transformation to a q_u with this marginal after z_u generation

The Parameters of the FSD Models for Sizes, q_u and Inter-Arrivals, u

Parameters

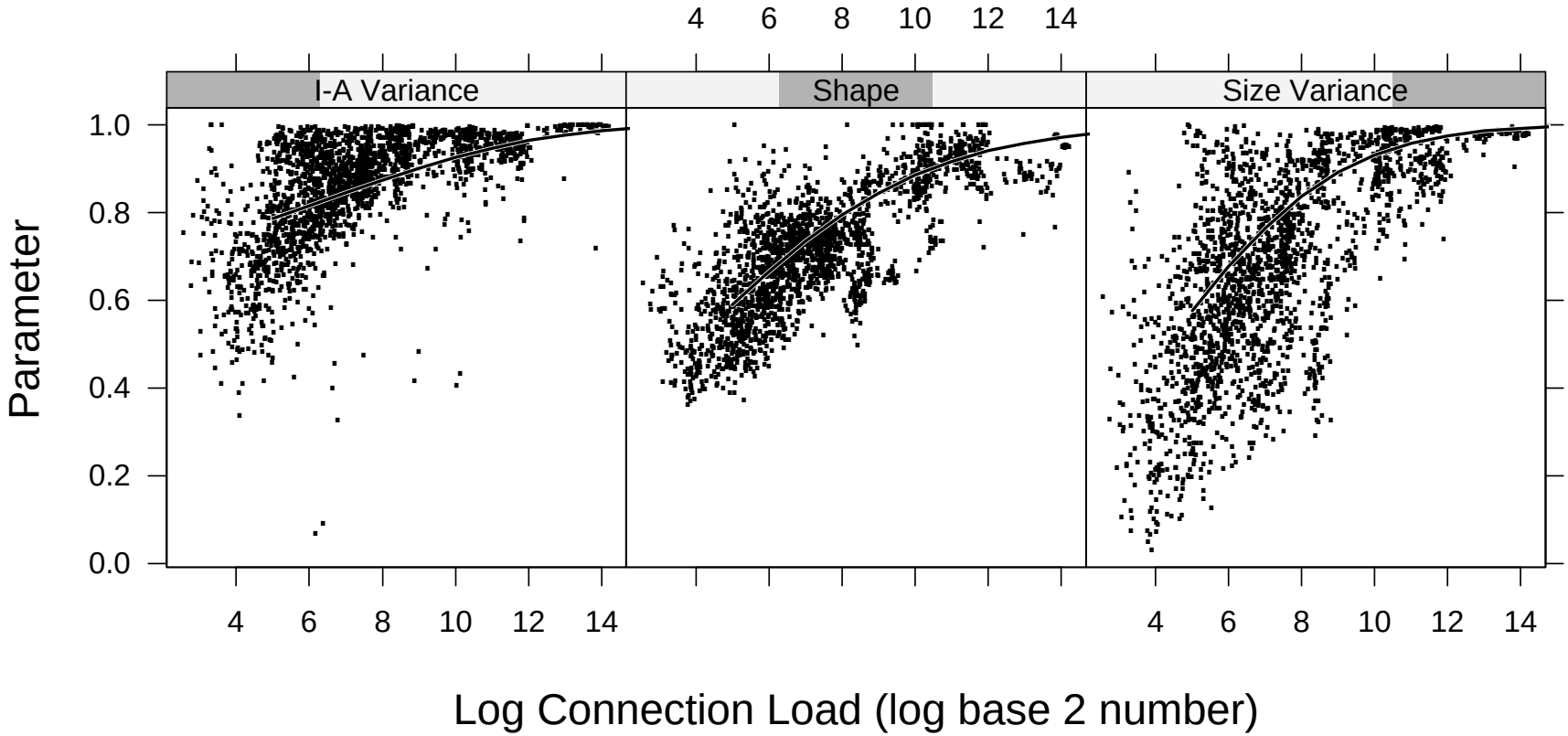
- size marginal
  - atom probabilities $\phi_1, \ldots, \phi_k$
- inter-arrival marginal
  - Weibull shape λ
  - Weibull scale α
- size dependence
  - fractional difference exponent $d_{(s)}$
  - noise variance $\theta_{(s)}$
- inter-arrival dependence
  - fractional difference exponent $d_{(i)}$
  - noise variance $\theta_{(i)}$

$d_{(s)}$ and $d_{(i)}$ are constant across links and loads and are both taken to be 0.41

$\phi_1, \ldots, \phi_k$ change with the link but not appreciably with the load on the link

λ, α, $\theta_{(s)}$, and $\theta_{(i)}$ change with the load

$\theta_{(i)}$, λ, and $\theta_{(s)}$ vs. Log Average Connection Load for 2072 traces



Simple models describe dependence of three parameters on connection load

$$\alpha = E(t_u)/\Gamma(\lambda^{-1} + 1)$$

Arrivals tend TOWARD Poisson and sizes tend TOWARD independent

EXTENSIVE VALIDATION

Tens of thousands of packet traces

- 90 sec to 15 min in duration
- 16 Internet links with speeds ranging from 100 mbps to 2.5 gbps

S-Net system for packet traffic analysis and visualization

- about 20 methods of analysis applied to each trace
- results conveyed by a collection of visualization tools, some VLDs (very large displays) with one page per trace