Analysis of On-Off Patterns in VoIP and Their Effect on Voice Traffic Aggregation

Wenyu Jiang, Henning Schulzrinne
{wenyu,schulzrinne}@cs.columbia.edu
Columbia University
Department of Computer Science

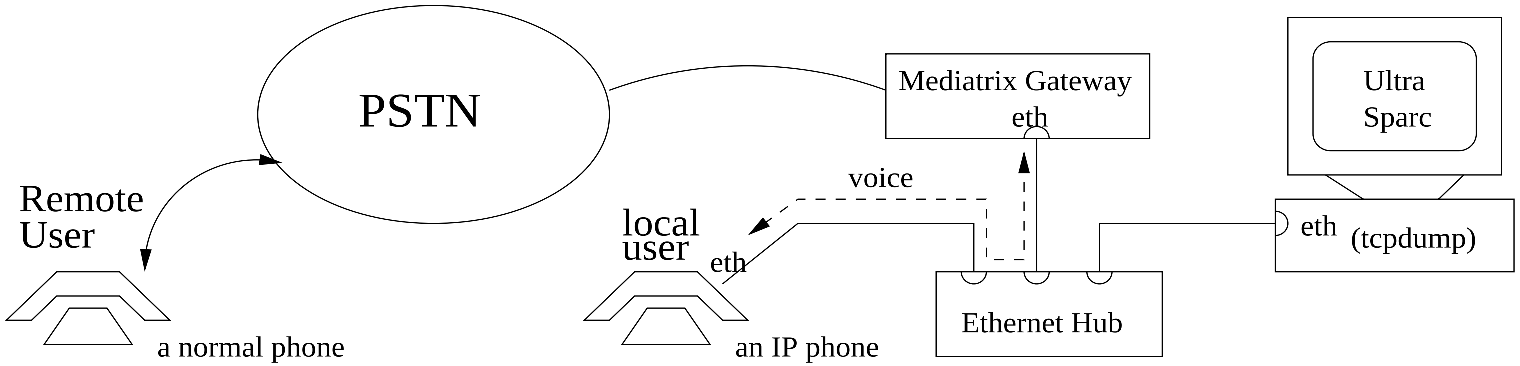
October 10, 2000

Introduction

- Human speech consists of talk-spurts (*on* period) and silence gaps (*off* period)
- Exploiting on-off patterns in VoIP allows
  - higher bandwidth utilization through multiplexing
  - per-spurt playout delay adjustment
  - enabling of echo suppression based on silence detector output
- Three major types of silence detectors
  - traditional: analog or digital ones with fixed energy thresholds
  - adaptive: digital, with adaptive energy thresholds, e.g., NeVoT SD
  - spectral: digital, decision based on spectral analysis, e.g., G.729B
- On-off patterns (spurt/gap distributions) are worth re-examination because
  - previous studies show spurt/gap is exponentially distributed, which are based on traditional silence detectors;
  - modern silence detector may behave differently $\longrightarrow$ may affect performance of traffic multiplexers (e.g., induced packet loss rate).

Recording Setup

- The telephone conversation goes through a PSTN/IP gateway, and gets recorded by *tcpdump* as Sun μ-law *.au* files.
- The gateway performs a 2-wire to 4-wire conversion.
- 6 pairs of recorded conversations, total duration 8743 sec.



The NeVoT Silence Detector

- Frame energy based
- Uses *hangover* rather than *fill-in*
- Dynamic energy threshold with min/max thresholds
- Pre and post spurt hangover

Parameter	Meaning	Default
min thresh	frame energy below which any signal is considered silence.	-45 dB
max thresh	highest allowed silence threshold	-20 dB
pre	pre-spurt hangover time	1 packet
post	post-spurt hangover time	6 packets

The G.729B Silence Detector

- Also frame energy based
- Fully automatic, i.e., no customized parameters
- The rough equivalent of min thresh in G.729B is -55 dB
  $\longrightarrow$ NeVoT SD by default (-45 dB) is less sensitive
- dynamic hangover time

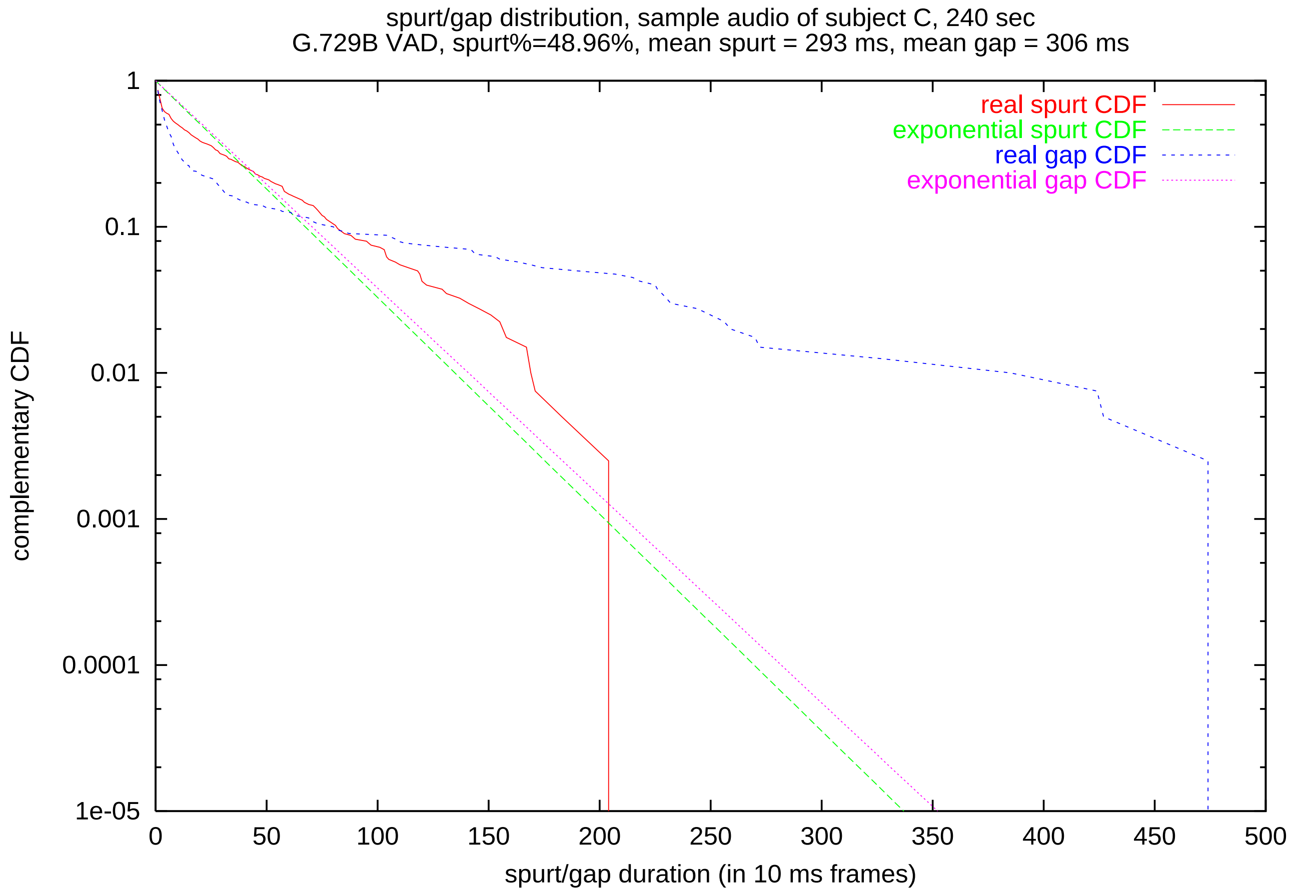
(Comparisons with) Traditional Silence Detectors

- Usually fixed energy threshold, fixed hangovers or fill-ins.
- If T (hangover) is small, mean spurt is around 200 to 400 ms, and the mean gap is around 500 to 700 ms.
- If T is around 200 ms, both the mean spurt and gap will be on the order of 1 to 2 sec.

	mean spurt	mean gap	talk% time
P.59 w/o hangover	227 ms	596 ms	27.6%
P.59 w hangover	1.004 sec	1.587 sec	38.7%
Brady w hangover	1.2 sec	1.8 sec	40.0%
Sriram (w/o hangover)	352 ms	650 ms	35.1%
G.729B (dynamic hangover)	362 ms	488 ms	42.6%
NeVoT SD (short hangover)	326 ms	442 ms	42.5%
NeVoT SD (default hangover)	903 ms	1216 ms	42.6%

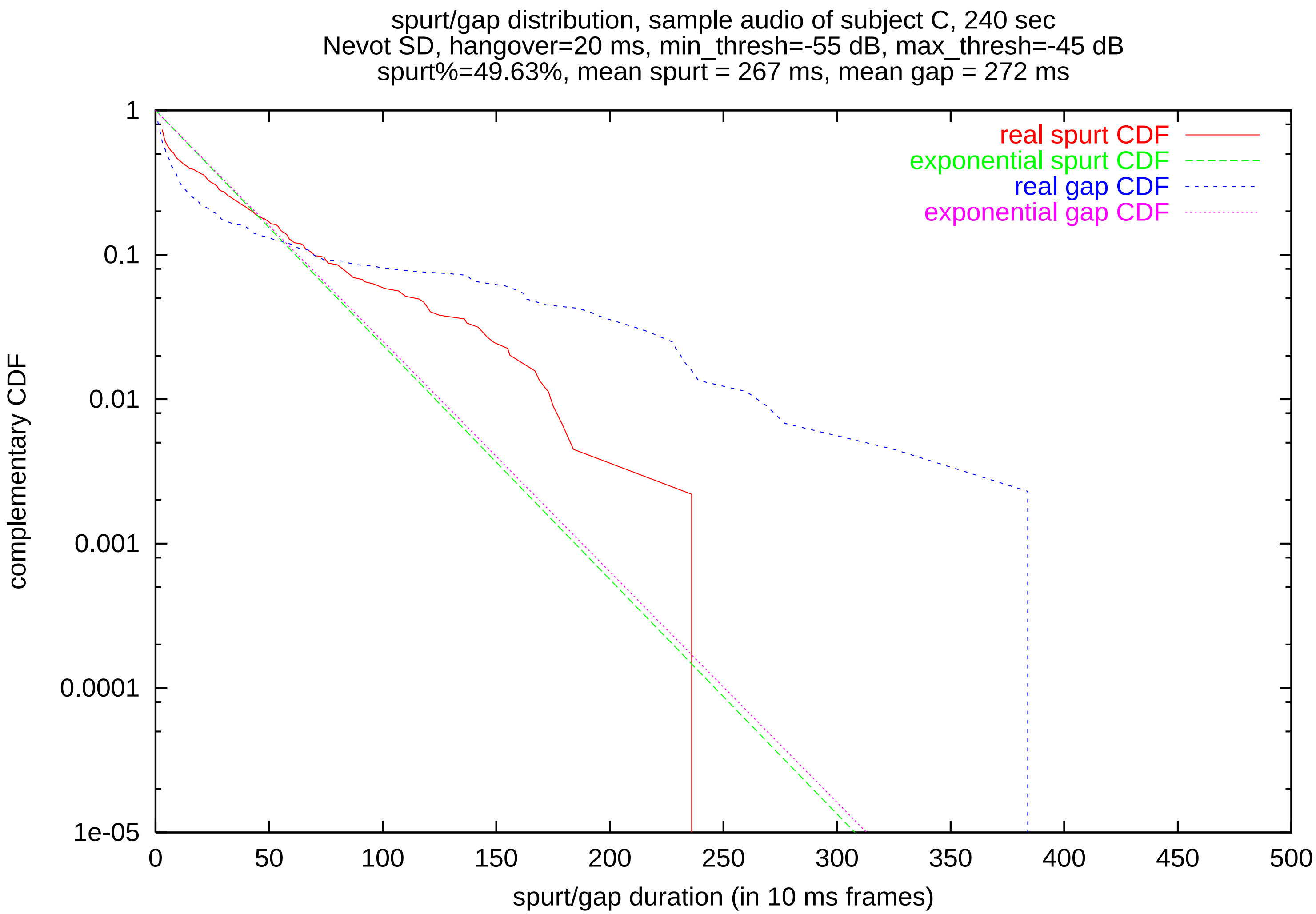
CDF Plot 1 – G.729B

- 6 conversations (5 in Chinese, 1 in English), totally 8743 sec, but no visible impact of language on distribution



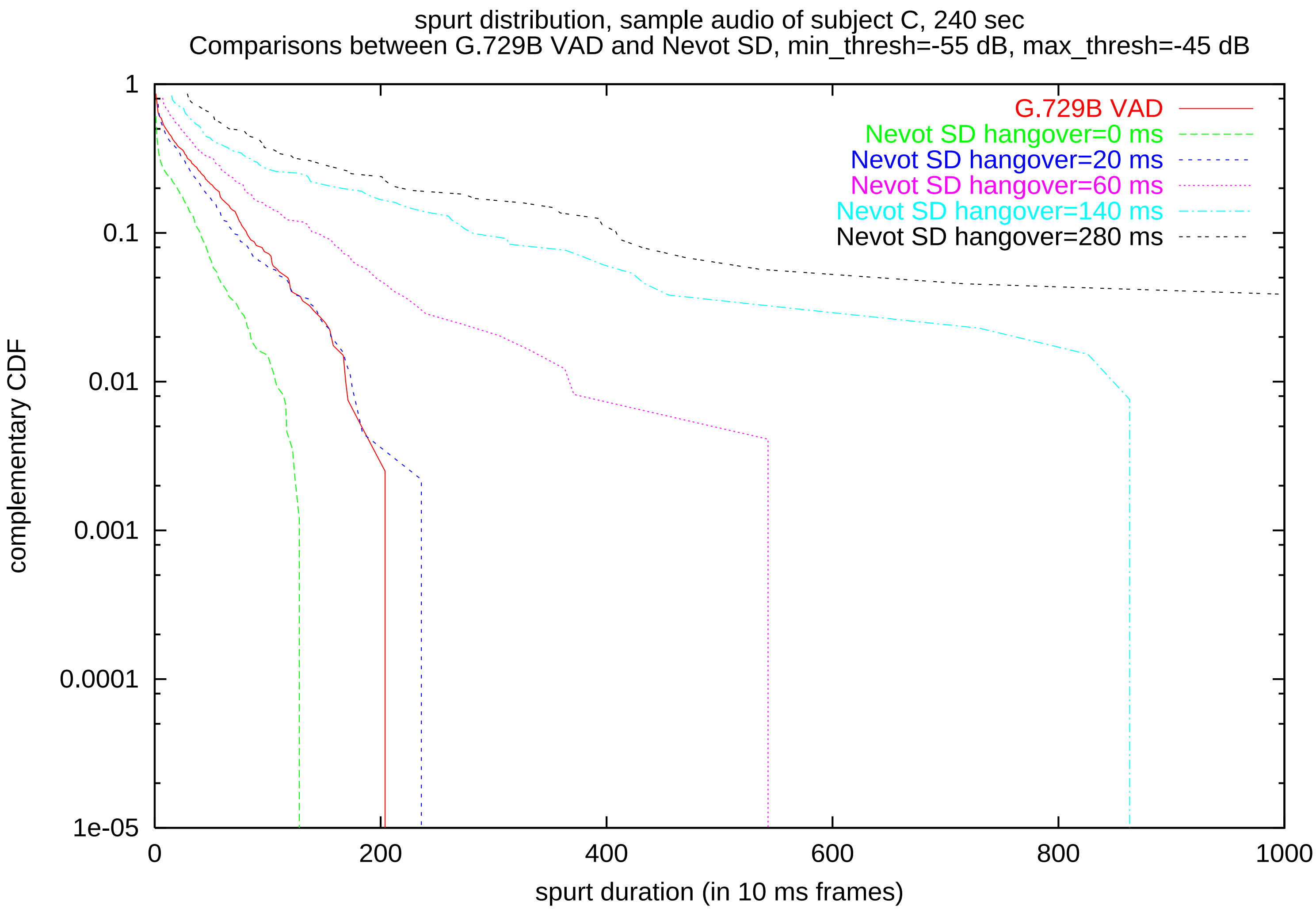
- CDF of one side of a conversation under G.729B;
- green curve is the exponential equivalent of red one (i.e., same mean spurt);
- the sudden drop of curve is due to the inexistence of long spurts and gaps.

CDF Plot 2 – NeVoT SD



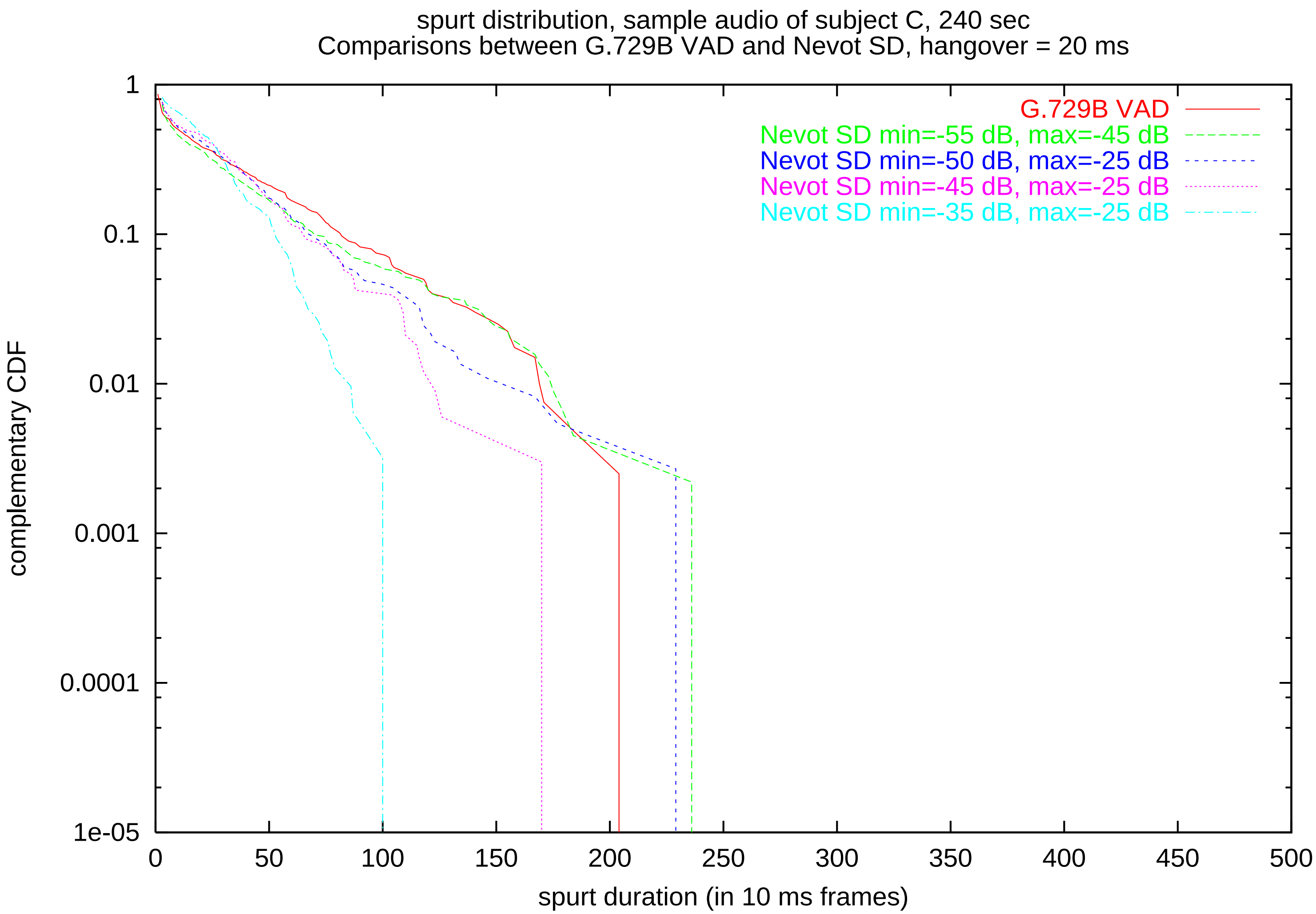
- CDF of the same conversation under NeVoT SD;
- Uses a short hangover of 20 ms and min thresh of -55 dB.

CDF Plot 3 – varying NeVoT SD parameters



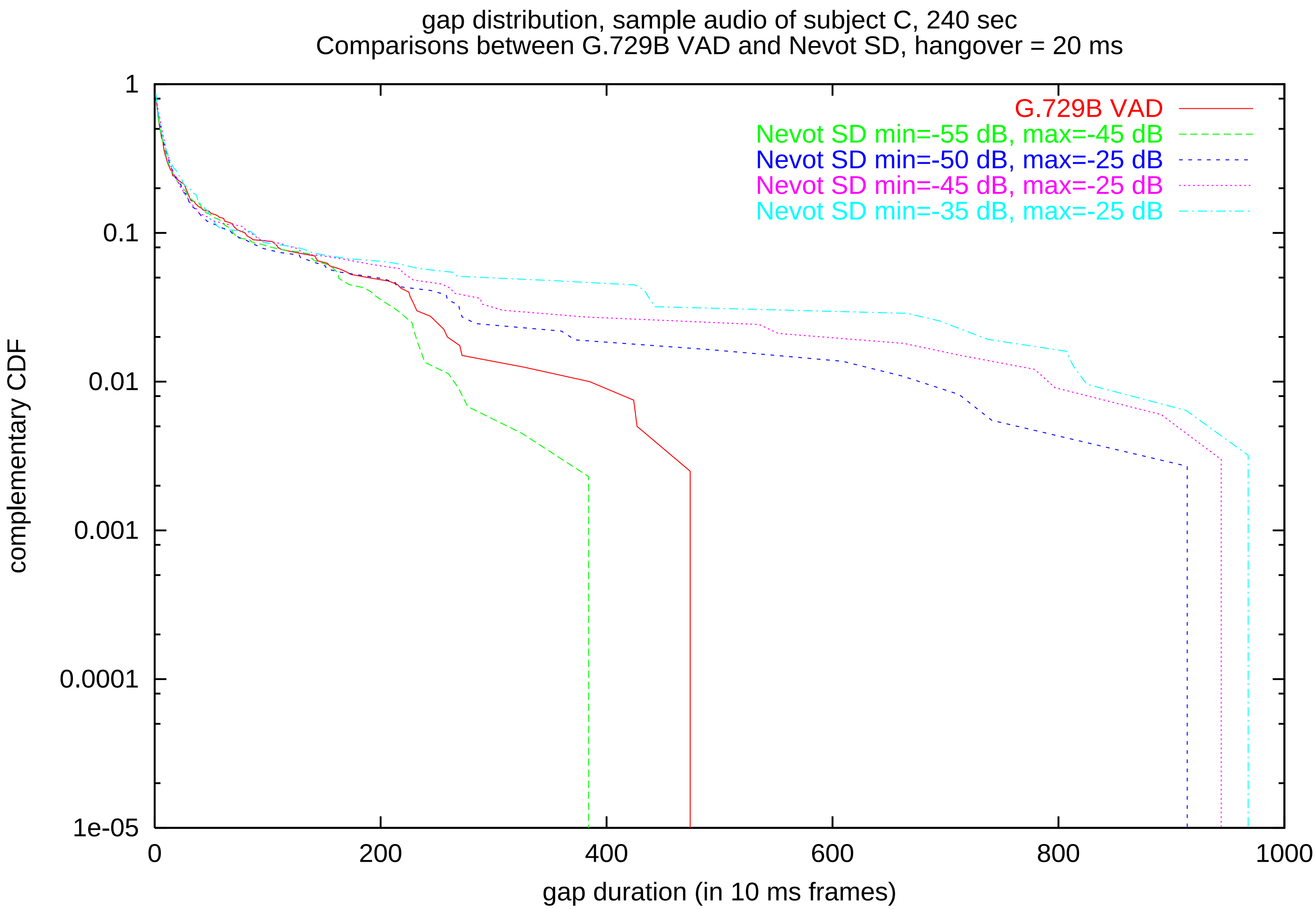
- change of spurt distribution when varying hangover time

CDF Plot 4 – varying NeVoT SD parameters



- change of spurt distribution when varying min and max thresh

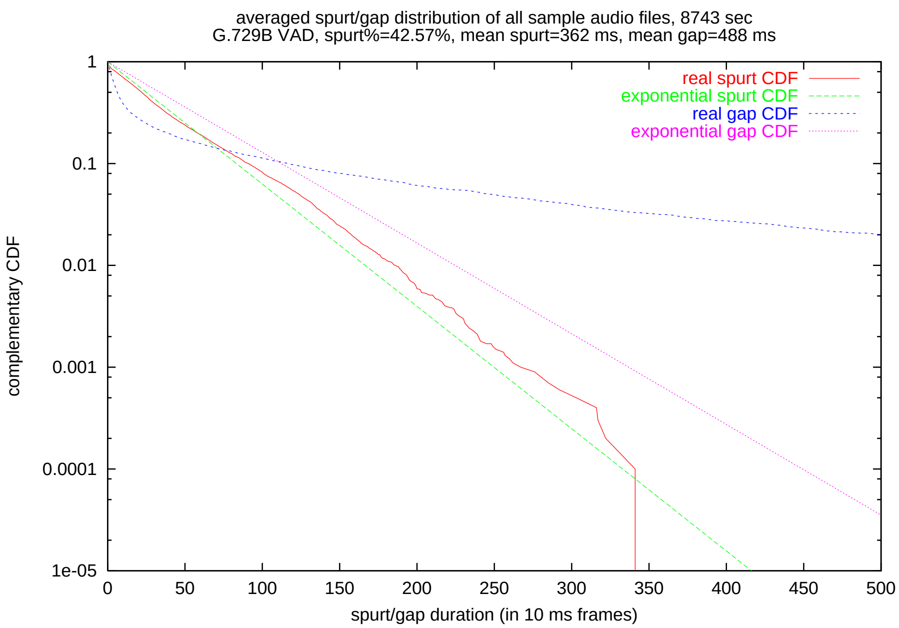
CDF Plot 5 – varying NeVoT SD parameters

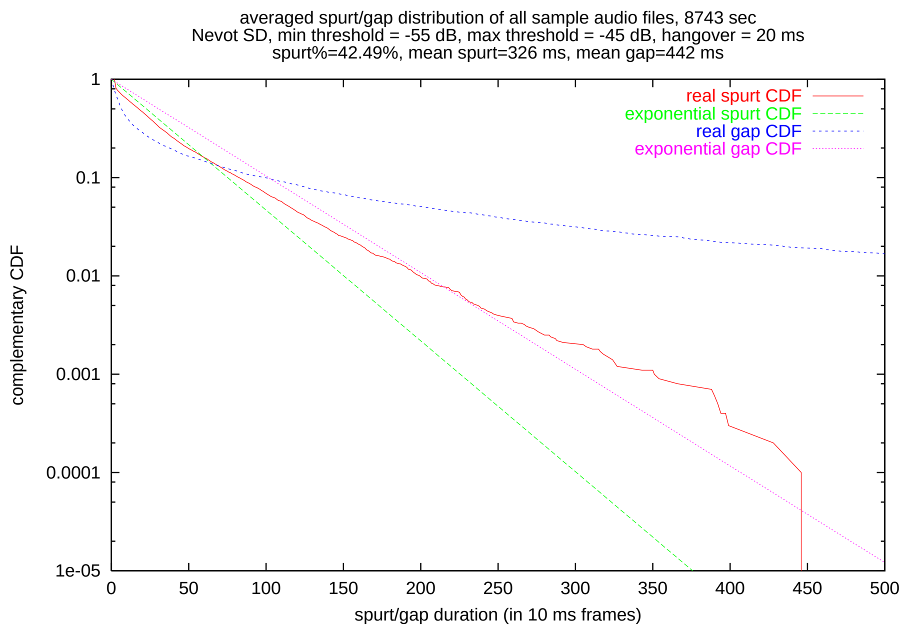


- change of gap distribution when varying min and max thresh

CDF Plot 6 – Averaged CDF

- Remarks on NeVoT SD parameters
  - min thresh is the most important, the next is hangover.
  - a min thresh of -55 dB and hangover of 20 ms yields similar results to G.729B

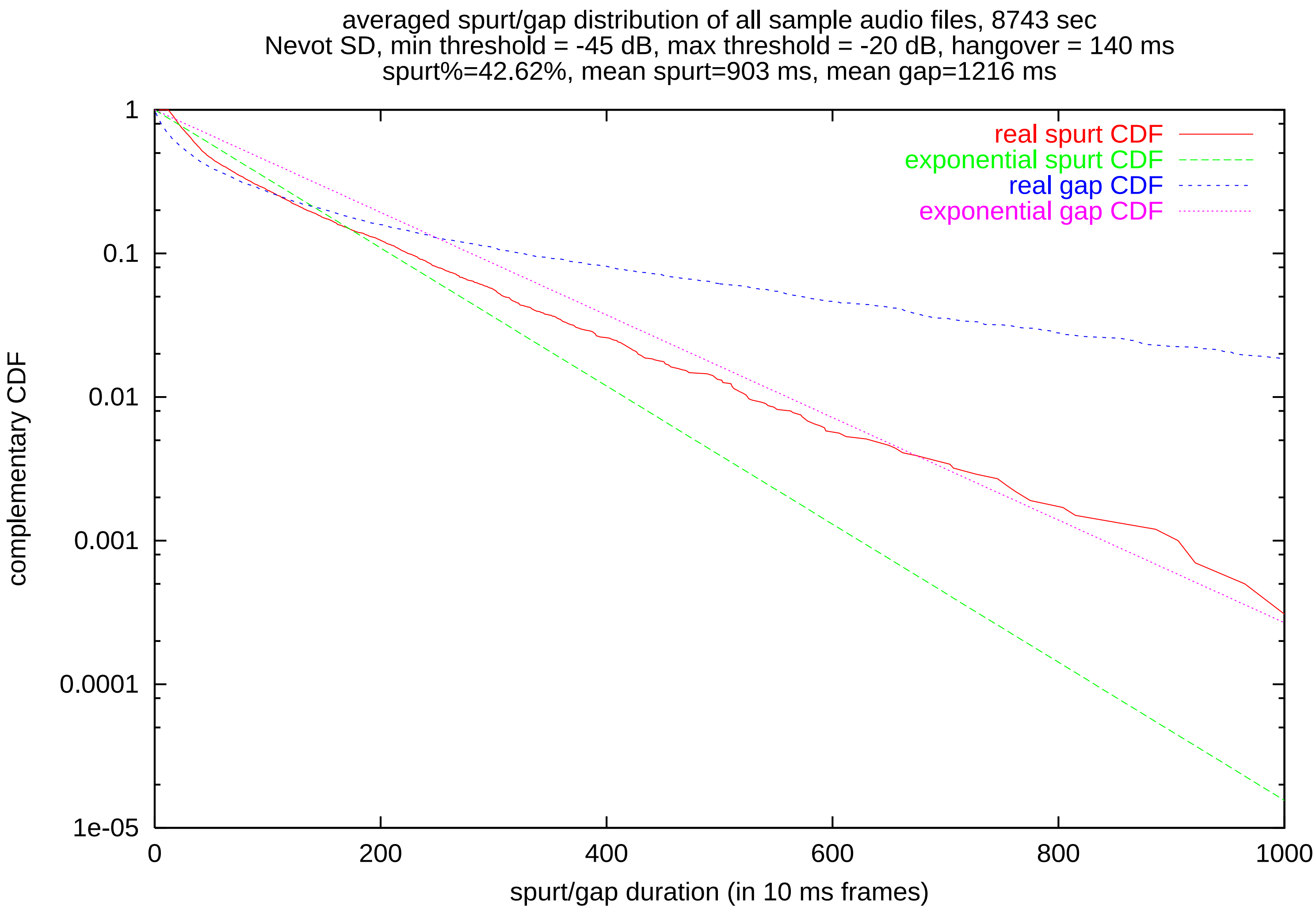




(a) G.729B

(b) NeVoT SD (20 ms hangover)

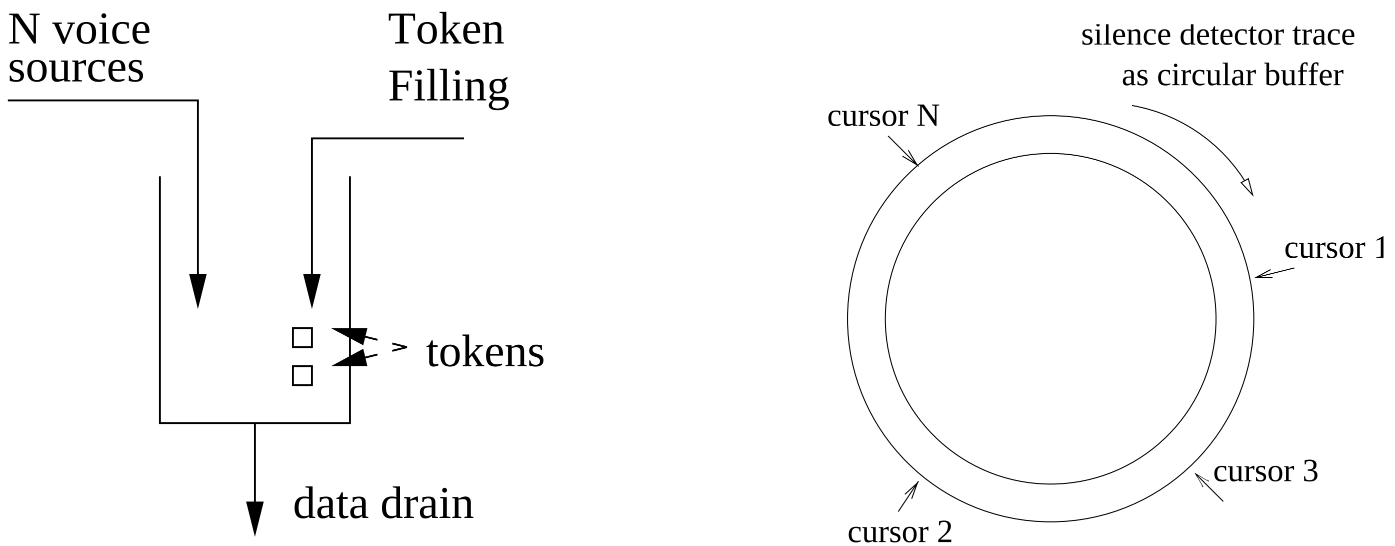
CDF Plot 6 – Averaged CDF, Continued



- NeVoT SD when using a 140 ms hangover

Voice Traffic Aggregation: Token Bucket Simulation Setup

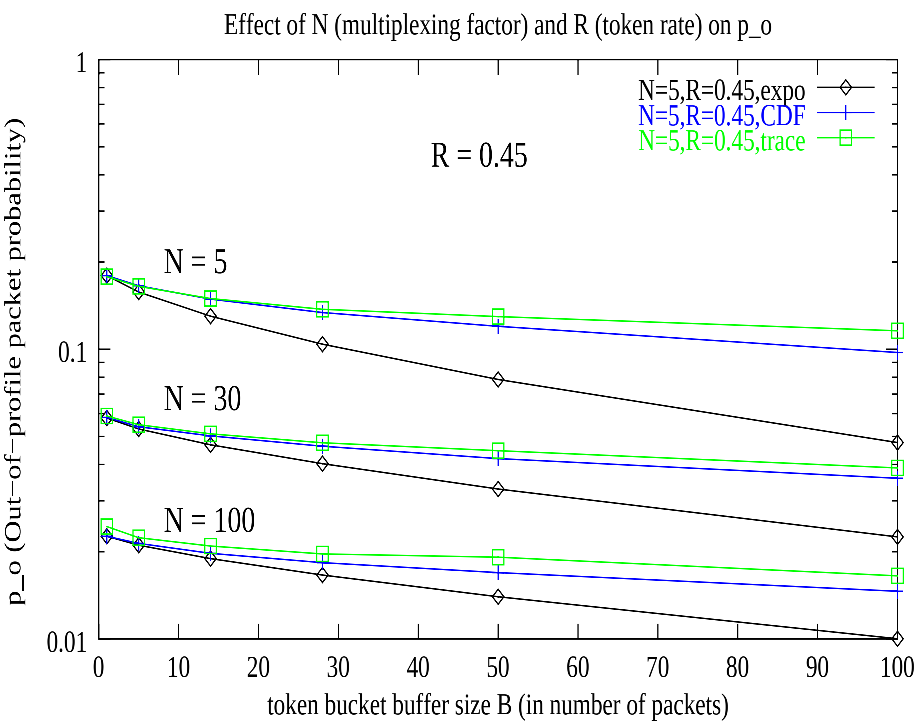
- Voice traffic is usually aggregated with a token bucket (or leaky bucket) multiplexer, especially in a DiffServ environment, to provide QoS control.
- Related work by Bruno *et al*, but they assume exponential distributions. (We achieved similar results for the exponential model)
- Three simulation models: exponential, CDF, and raw silence detector trace.



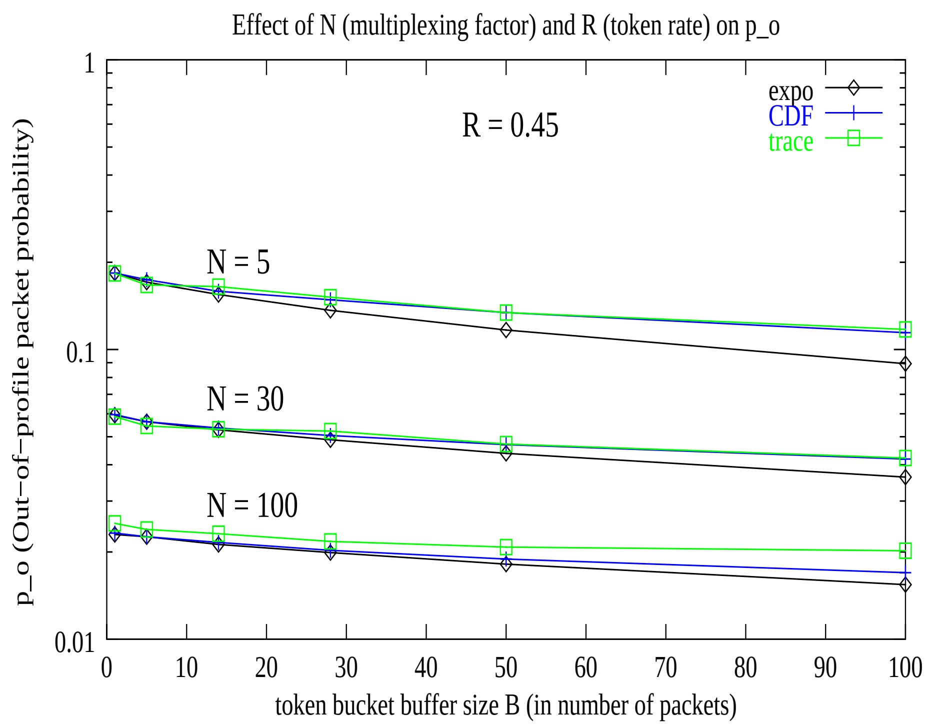
(a) A token bucket filter in action (b) Illustration of trace-based simulation

Simulation Results

- N - number of sources, R - reserved bandwidth, as % of peak bandwidth
- B - token buffer size, as in number of packets
- p_o - probability of a packet is out-of-profile



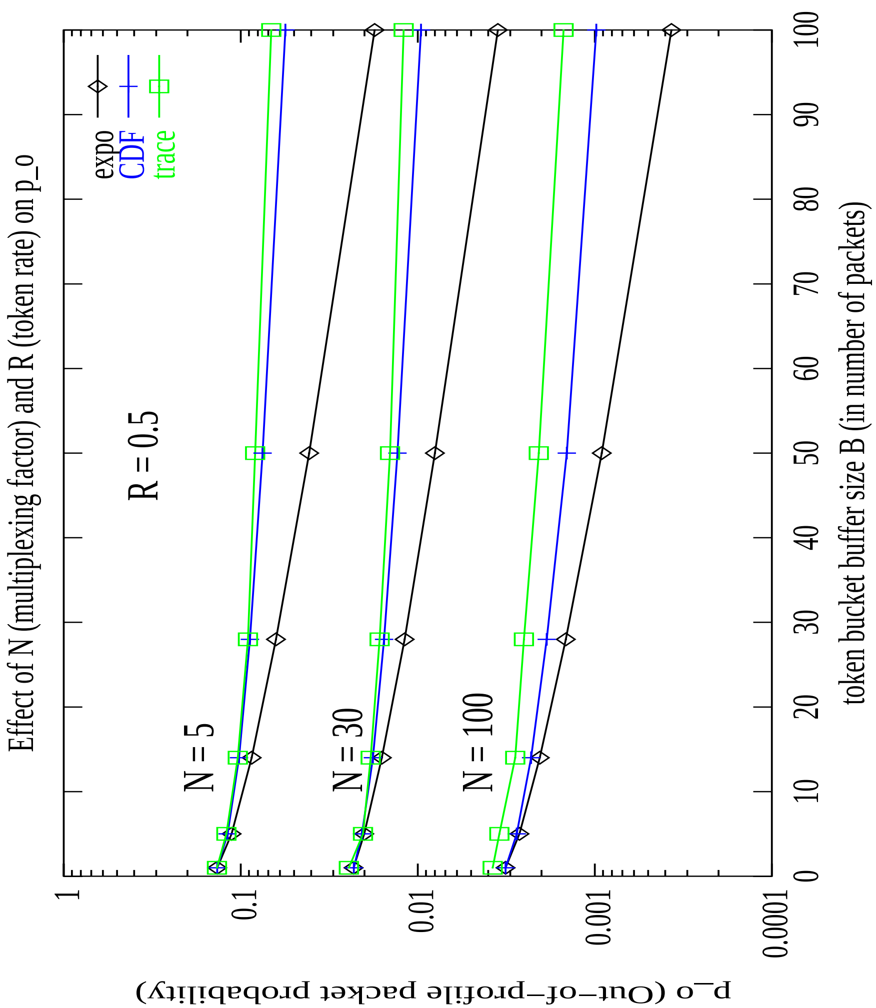
(a) p_o vs. B, G.729B, R = 0.45



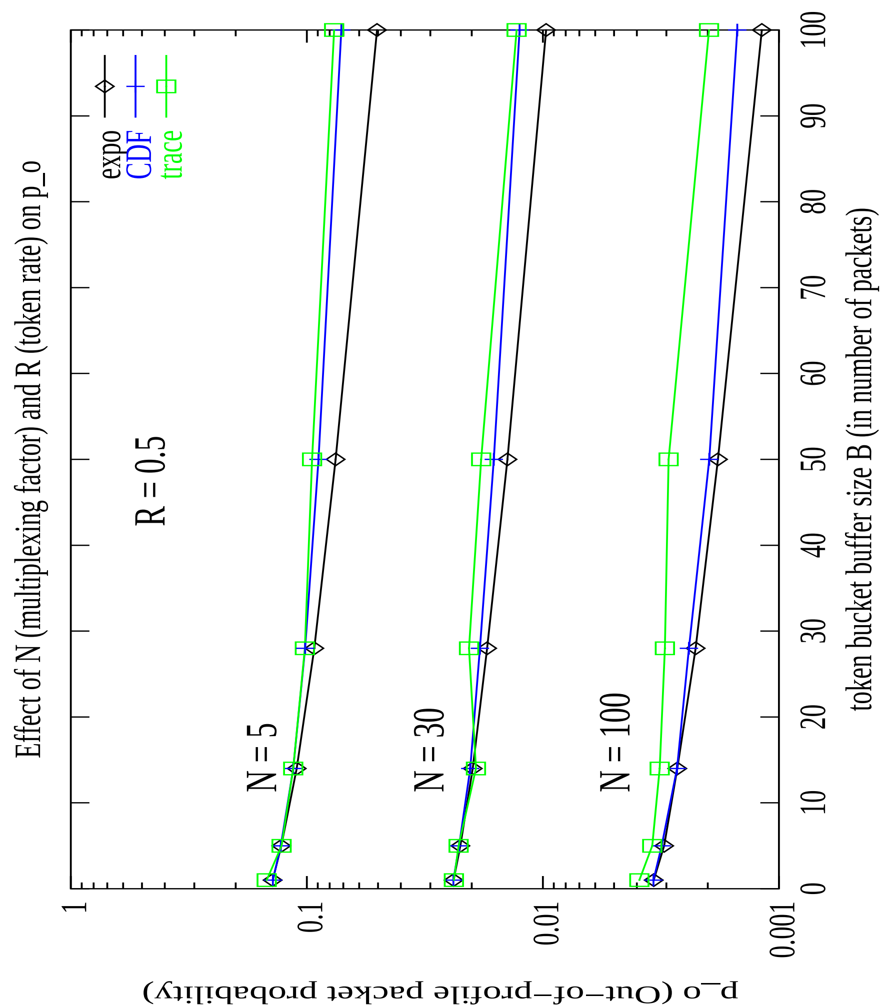
(b) p_o vs. B, NeVoT SD, R = 0.45

Simulation Results, Continued

(a) G.729B, $R = 0.5$

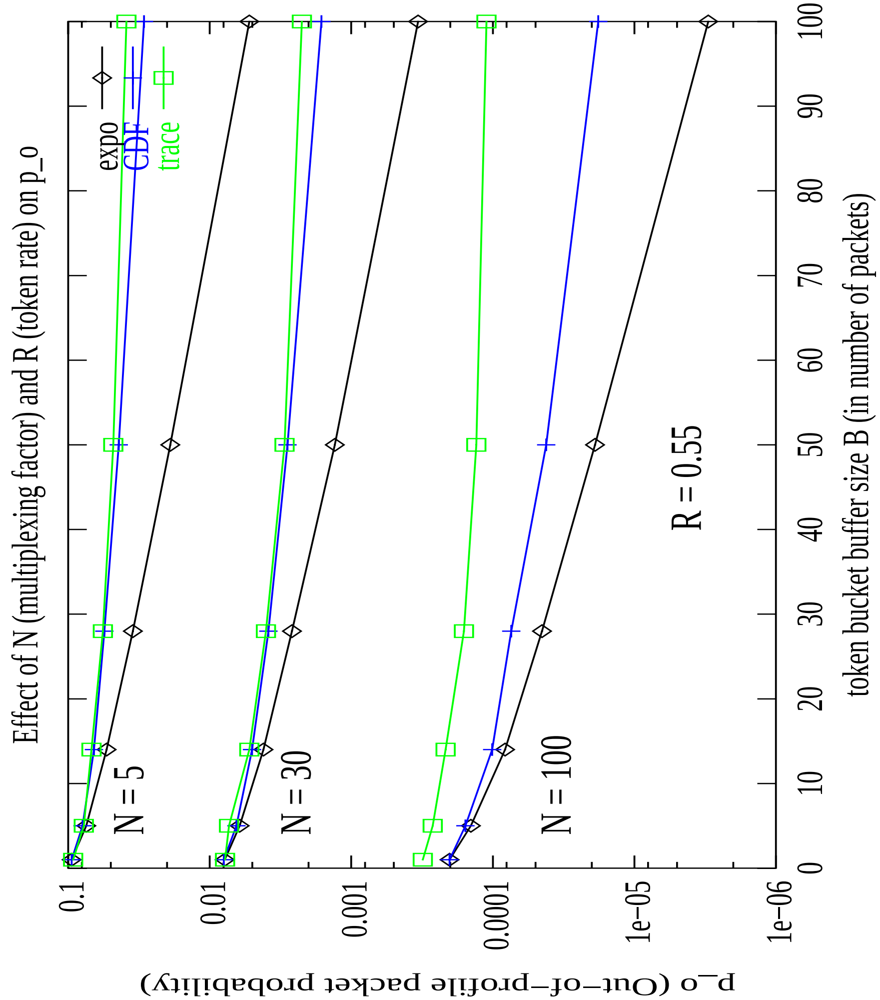


(b) NeVoT SD, $R = 0.5$

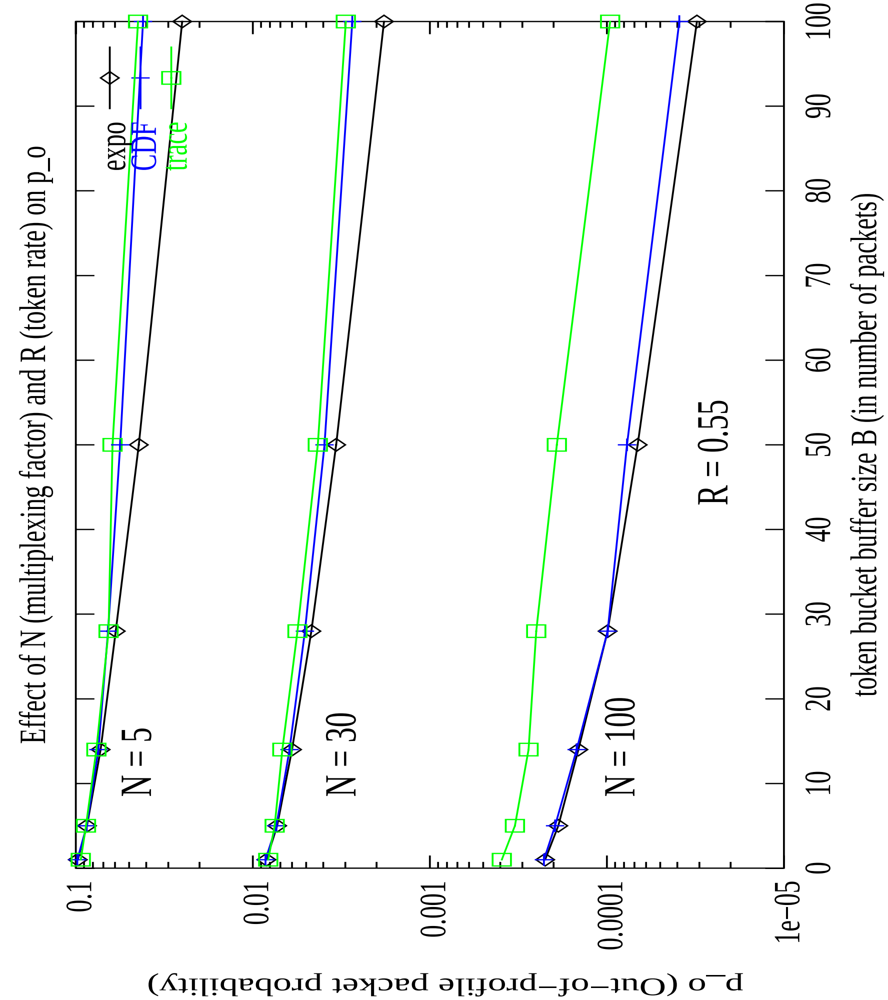


Simulation Results, Continued

(a) G.729B, $R = 0.55$



(b) NeVoT SD, $R = 0.55$



Remarks on Simulation Results

- Absolute difference between three models is small, especially for small R and/or large N.
- Relative difference can be large when B is large and/or N is small.

N	R	B	p_o expo	p_o CDF	p_o trace
5	0.45	14	0.130	0.149	0.150
5	0.55	100	0.005	0.029	0.039
100	0.55	100	3×10^{-6}	1.8×10^{-5}	1.11×10^{-4}

- Trace model yields a slightly but consistently higher p_o than the CDF model $\longrightarrow$ trace model is slightly burstier than CDF model.
- Anomaly data point observed at N=100, R=0.55, B=100: the relative difference even between CDF and trace model is large. Cause: unknown.
- The difference between three models for NeVoT SD with default hangover is smaller, possibly because its distribution is closer to exponential than that of G.729B.

Implications to the user and service provider

- If p_o corresponds exactly to the packet loss rate, the difference between three models will have an effect on VoIP quality, but the degree of such effect also depends on other factors like loss concealment and FEC (Forward Error Correction).
  - an effective loss concealment method can repair most of the signals when loss rate is below a threshold.
  - FEC can recover any lost packet within an FEC block if a minimum percentage of the block is received.
- For some users, a 99% good and a 99.5% good circuit may sound similar, but for a strict SLA, 0.5% and 1.0% loss rate are clearly different.

Conclusions

- The spurt/gap distributions produced by some modern codecs (G.729B and NeVoT SD) are not exactly exponential.
- Such distributions depend highly on the min threshold and the hangover mechanism.
- The effect of deviation from the exponential on-off model is illustrated by a series of token bucket simulations.
  - The absolute difference between the exponential, CDF and trace models are small in general;
  - But the relative difference can be large when B (token buffer size) is large and/or N (number of sources) is small.
  - The implications of relative difference depend on the applications in use and its requirement (e.g., a strict SLA may have a strong requirement).