Charging from Sampled Network Usage

Nick Duffield
Carsten Lund
Mikkel Thorup

AT&T Labs-Research, Florham Park, NJ

Do Charging and Sampling Mix?

- Usage sensitive charging
  - charge based on sampled network usage
- Is sampling necessary?
  - just count all packets/bytes in network?
  - measure and export all traffic flows stats?
- Is sampled usage reliable enough?
  - risk of overcharging or undercharging

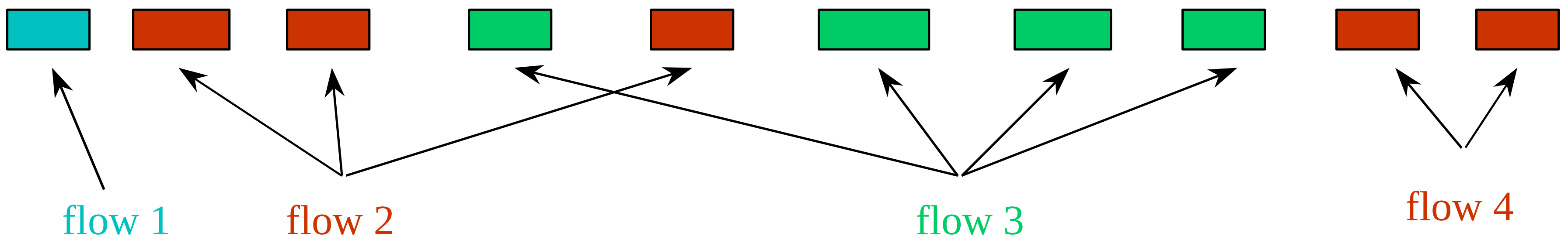
Why usage-sensitive charging?

- Compare charging on port-size
  - coarse granularity OC3⇒OC12⇒OC48⇒0C192
- Implicit resource management
  - price disincentive to greedy use
- Differentiated services
  - will require differentiated charges

Fine count **all** packets/bytes in network?

- Mirror pricing policy in router configuration?
  - separate counter for each billable packet stream
- Scaling/dimensionality issues
  - potentially many determinants to pricing
    - ToS, application type, source/dest IP address, ...
  - routers must support large number of counters
- Configuration issues
  - change pricing policy $\Rightarrow$ reconfigure counters
    - administrative cost

IP Flow Abstraction



- IP flow abstraction
  - set of packets identified with “same” address, ports, etc.
  - packets that are “close” together in time
  - possible protocol-based flow demarcation
    - e.g. terminate on TCP FIN
- IP flow summaries
  - reports of measured flows from routers
    - flow identifiers, total packets/bytes, router state
- Several flow definitions in commercial use

Measure/Export All Traffic Flows?

- Measure traffic flows as they occur
  - export flow summaries to billing system
- Flow volumes
  - one OC48⇒several GB flow summaries per hour
- Cost
  - network resources for transmission
  - storage/processing at billing system

Flow Sampling?

- Sampling
  - statisticians reflex action to large datasets
- Export selected flows
  - reduce transmission/storage/processing costs
- Sufficiently accurate for pricing?
  - risk of overcharging ($\Rightarrow$ irate customers)
  - risk of undercharging ($\Rightarrow$ irate shareholders)

Packet Sampling and Flow Sampling

- Packet Sampling
  - when router can't form flows at line rate
    - – scaling at a single router
- Flow sampling
  - managing volume of flow statistics
    - – scaling across downstream measurement infrastructure
- Complementary
  - could combine
    - – e.g. 1 in N packet sampling + flow sampling

Usage Estimation

- Each flow i has
  - "size" x_i
    - – bytes or packets
  - "color" c_i
    - – combination of IP address, port, ToS etc that maps to billable stream (= customer + billing class)
- Goal
  - to estimate total usage X(c) in each color c

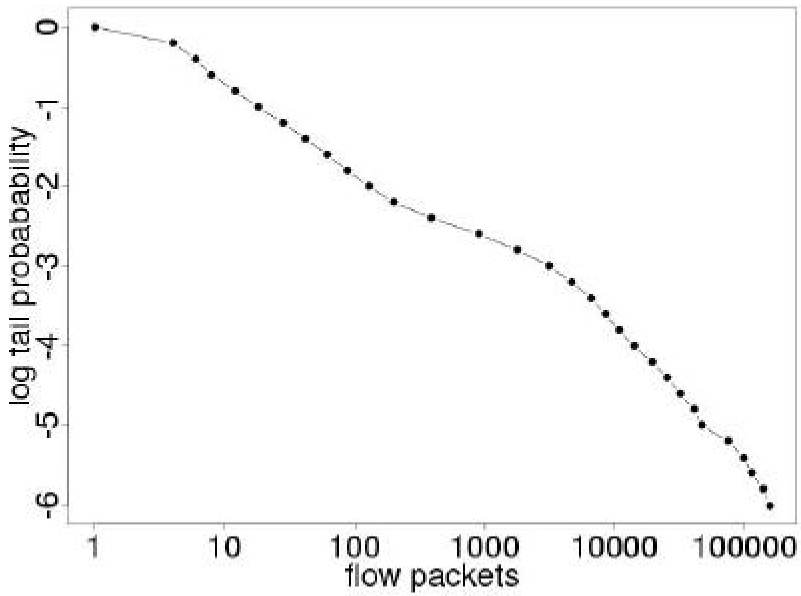
$$X(c) = \sum_{i:\, c_i = c} x_i$$

Basic Ideas

- Match sampling method to flow characteristics
  - high fraction of traffic found in small fraction of long flows
    - sample long flows more frequently than short flows
      - large contributions to usage more reliably estimated
- Manage sampling error through charging scheme
  - make charging **insensitive** to small usage
    - sampling error for small usage not reflected in charge to user
- Trade-off
  - allow small consistent undercount to reduce risk of overcharge
- Show how to relate sampling and charging parameters
  - simple rules to achieve desired accuracy

Size independent flow sampling bad

- Sample 1 in N flows
  - estimate total bytes by N times sampled bytes
- Problem:
  - long flow lengths
    - – estimate sensitive to inclusion or omission of a single large flow



Size dependent flow sampling

- Sample flow summary of size x with prob. p(x)
- Estimate usage X by

$$X' = \sum_{\substack{\text{sampled} \\ \text{flows}}} x \big/ p(x)$$

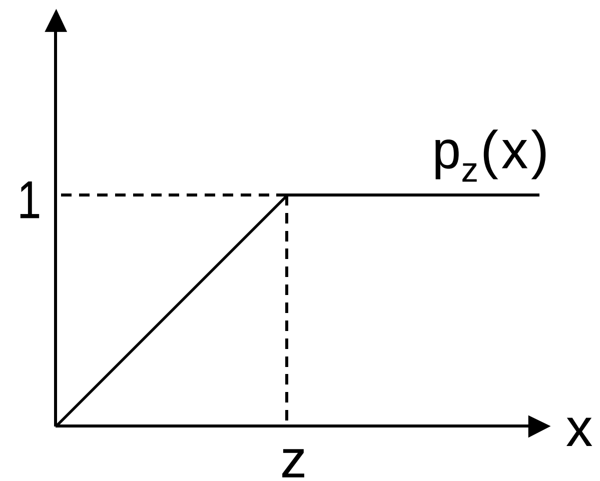
  - boost up size x by factor 1/p(x) in estimate X'
    - compensate against chance of being sampled
- Chose p(x) to be increasing in x
  - longer flows more likely to be sampled
  - compare size independent sampling: p(x) =1/N

Statistical Properties

- Fixed set of flow sizes $\{x_1, x_2, \ldots, x_n\}$
  - we only consider randomness of sampling
- X' is unbiased estimator of actual usage $X = S_i\ x_i$
  - ÄX' = X: averaging over all possible samplings
  - holds for all probability functions p(x)
- Proof:
  - $X' = S_i w_i \quad /p(x_i)$
    - w_i random variable
      - w_i =1 with prob. $p(x_i)$, 0 otherwise
    - Ä$w_i = p(x_i)$ hence ÄX' = Ä $S_i w_i\ x_i\ /p(x_i) = S_i\ x_i = X$

What is best choice of p(x)?

- Trade-off accuracy vs. number of samples
- Express trade-off through **cost function**
  - cost = variance(X') + z^2 average number of samples
    - parameter z: relative importance of variance vs. # samples
- Which choice of p(x) minimizes cost?
- $p_z(x) = \min \{ 1 , x/z \}$
  - flows with size ≥ z: always selected
  - flows with size < z: selected with prob. proportional to their size



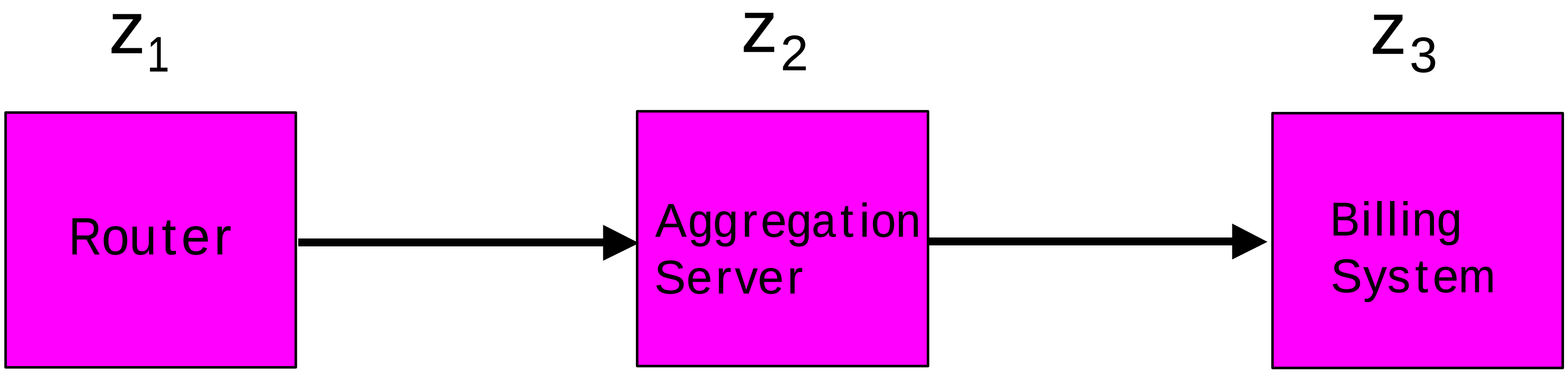
- Trade-off
  - smaller z
    - more samples, lower variance
  - larger z
    - fewer samples, higher variance
- Will call sampling with $p_z(x)$ "optimal"

Implementation

- Nearly as simple as 1 in N sampling
  - use flow size variability as source of randomness
    - no random number generators

```
sample(x) {
  static count = 0
  if (x  > z) {
   select_flow
  }
  else {
   count += x
   if ( count  >  z) {
         count = count - z
         select_flow
   }
  }
}
```

Optimal Resampling

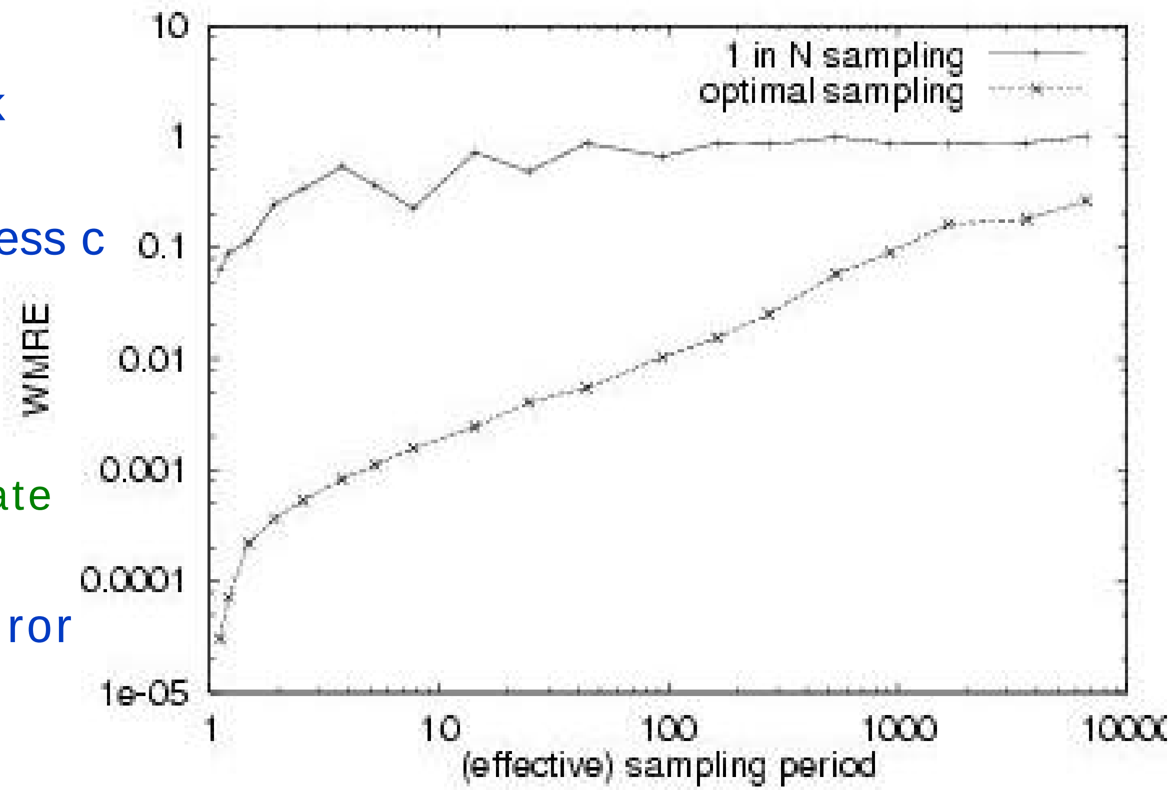


- Resampling to progressively thin flow summaries
- Finer resampling ($z_1 \le z_2 \le z_3$) preserves statistics
  - final flow stream at billing system has same statistical properties as would original stream sampled once with z_3

Optimal vs. size independent sampling

- NetFlow traces
  - 1000's cable users, 1 week
- Color flows
  - by customer-side IP address c
- Compare
  - 1 in N sampling
  - optimal sampling
    - same average sampling rate
- Measure of accuracy
  - weighted mean relative error

$$\frac{\sum_c |X'(c) - X(c)|}{\sum_c X(c)}$$



- Heavy tailed flow size distribution is our friend!
  - allows more accurate encoding of usage information

Charging and Sampling Error

- Optimal sampling
  - **no** sampling error for flows larger than z

- Exploit in charging scheme
  - fixed charge for small usage
  - usage sensitive charge only for usage above **insensitivity level** L

- Charge according to estimated usage

$$f(X'(c)) = a + b \max\{ L , X'(c) \}$$

  - coefficients a, b and level L could depend on color c

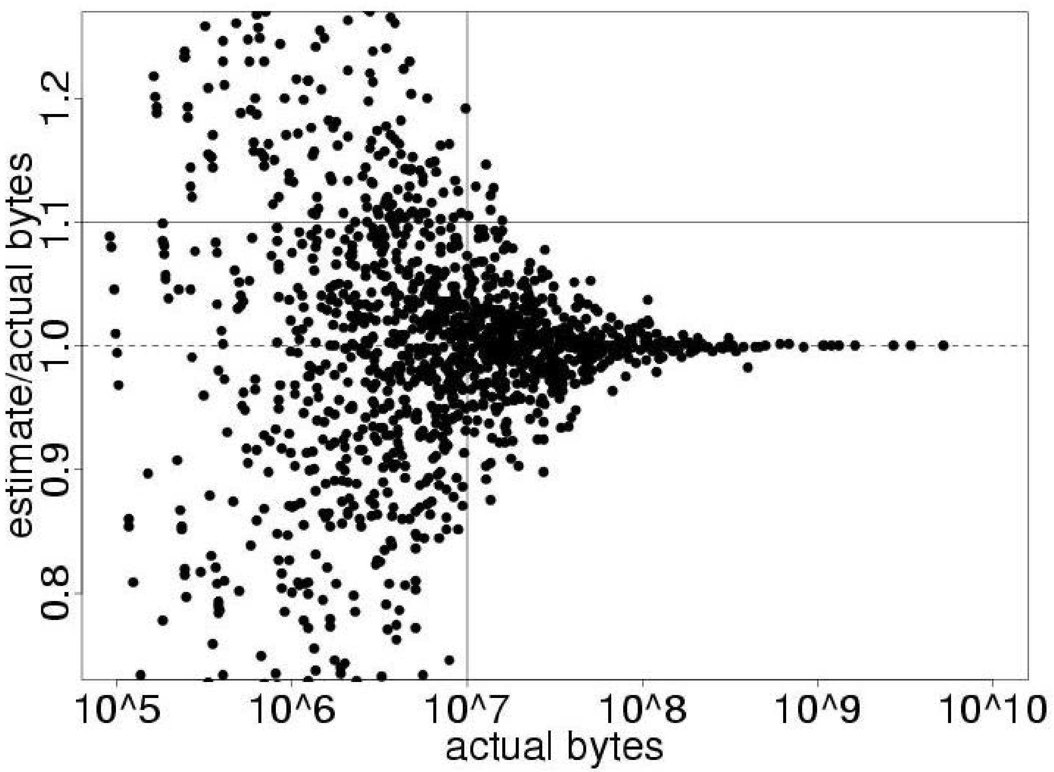
- Only usage above L needs reliable estimation

Accuracy and Parameter Choice

- Given target accuracy
  - relate sampling threshold z to level L
- Theorem
  - Variance(X') $\leq$ z X (tight bound)
  - now assume: $z \leq \varepsilon^2 L$
    - Std.Dev. X' $\leq \varepsilon$ X if X $\geq$ L
      - bound sampling error of estimated usage > L
    - Std.Dev. f(X') $\leq \varepsilon$ f(X)
      - bound error of charge based on estimated usage
- Bounds hold for any flow sizes $\{x_i\}$
  - no assumption on flow size distribution
    - just choose $z \leq \varepsilon^2 L$

Example

- Target parameters
  - $L = 10^7$, $\varepsilon = 10\% \Rightarrow z = 10^5$
- Scatter plot
  - ratio estimated/actual usage vs. actual usage
    - each color c
  - observe better estimation of higher usage
- Want to avoid
  - ratio > $1+\varepsilon = 1.1$
    and
    usage > $L = 10^7$
- Less than 1 in 1000 "bad" points



Compensating variance for mean

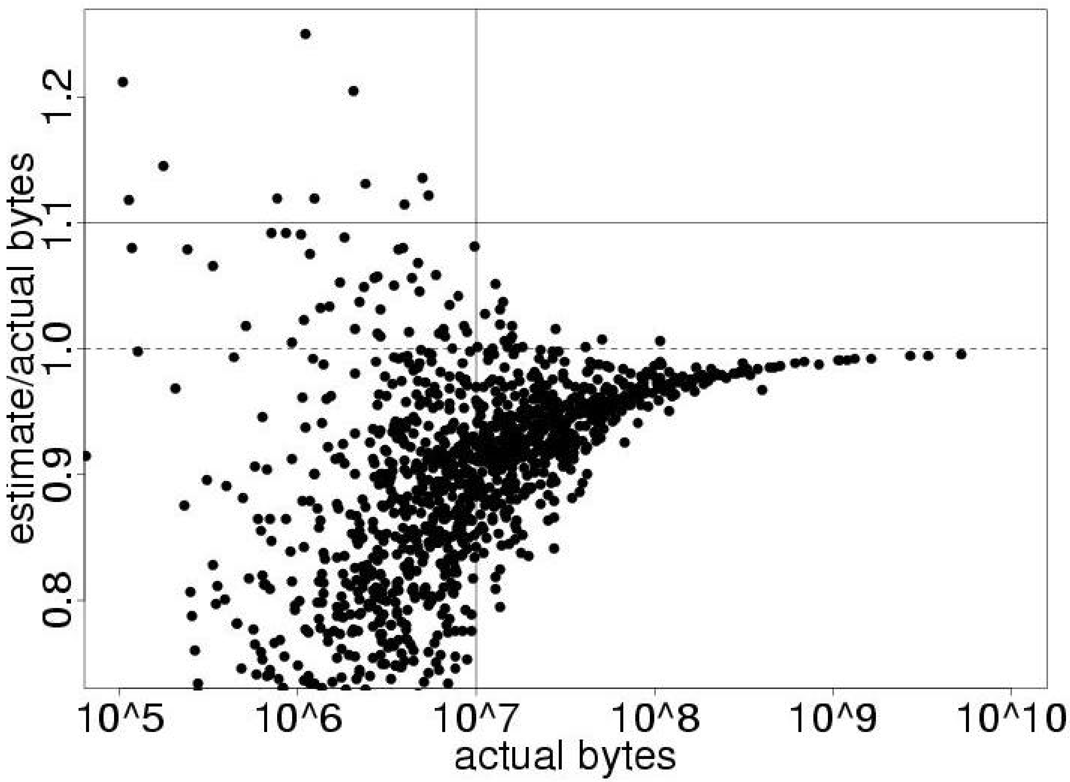
- Aim:
  - reduce chance of overestimating usage
- Method:
  - theorem gave bound: Var(X') $\leq$ z X
  - anticipate upwards variations in X' by subtracting off multiples of std. dev.
    - charge according to

$$X_s' = X' - s\sqrt{zX'}$$

  - again: no assumptions on flow size distribution

Example: s=1

- Scatter pushed down:
  - no points with ratio>1.1 and usage > 10^7
- Drawback
  - more unbillable usage
    - when $X'_s < X$
- Small unbillable usage for heavy users
  - ratio→1
  - Std.Dev.(X')/X' vanishes as X grows

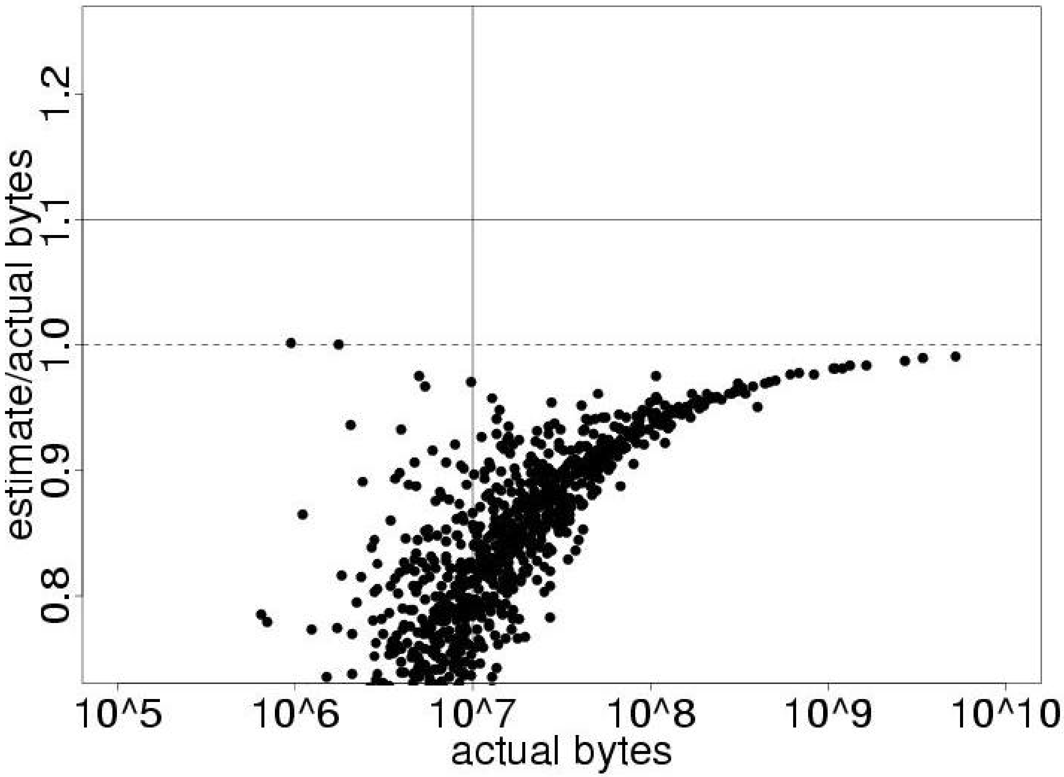


Example: s=2

- Scatter pushed down further:
  - no points with ratio > 1

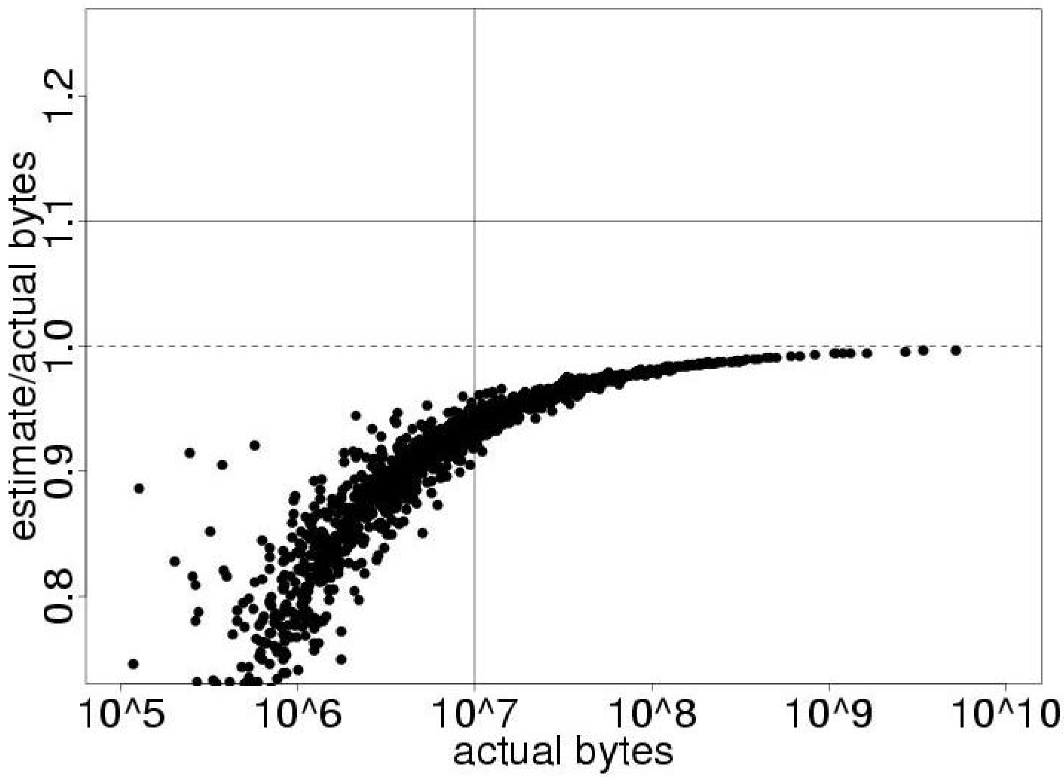
- Trade off
  - unbillable usage vs. overestimation

s	unbill. bytes	$X'_s > X$?
0	-0.1%	50%
1	3.1%	3%
2	6.2%	0%



How to reduce unbillable usage?

- Make sampling more accurate
  - reduce z!
- For unbillable fraction < η
  - chose $s\ z \leq \eta^2 L$
- Example:
  - $s = 2$, $\eta = 10\%$
  - reduce z
    - from 10^5 to 10^4
- Alternative
  - increase coefficent a in charge f(X) to cover costs



Tension between accuracy and volume

- Want to reduce z
  - better accuracy, less unbillable usage
- Drawback
  - increased sample volume
- Solution
  - make billing period longer instead
    - usage roughly proportional to billing period
    - allows increased charge insensitivity level L
  - sample production rate controlled by threshold z
    - rate $r \Sigma_x f(x) p_z(x)$
      - flow arrival rate r, fraction f(x) of flows size x
- Need only $z = \varepsilon^2 L$
  - larger L allows smaller error ε for given z

Summary

- Size dependent optimal sampling
  - preferentially sample large flows
    - more accurate usage estimates for given sample volume
    - sample flow of size x with probability $p_z(x)$
- Charging from measured usage X'
  - charge f(X') = a +b max{L,X'}
    - fixed charge for usage below insensitivity level L
    - only need to reliably estimate usage above L
- Sampling/charging accuracy
  - choose $z = \varepsilon^2 L$ to get standard error ε
- Variance compensation
  - replace X' by $X_s' = X' - s\sqrt{zX'}$
- Longer billing cycle
  - increases L, better accuracy (ε) at given sampling rate (z)

Further Work

- Dynamic control of sample volume
  - aim:
    - bound sample rate **when arrival rate r varies**
  - method:
    - dynamic adjustment of sampling threshold z