Lightweight Link Dimensioning using sFlow Sampling

1. Motivation

- Currently, operators use coarse measurements to dimension the capacity of their links at large timescales (up to 15 minutes) that overlooks short-term traffic bursts.
- Alternative approaches for link dimensioning are accurate but do not scale to high-speed networks because they require continuous packet capturing.
- There is a need for a scalable and easy-to-deploy link dimensioning approach to substitute current practices.
- **Research question: can we estimate traffic variance and required capacity using sampled packet captures?**

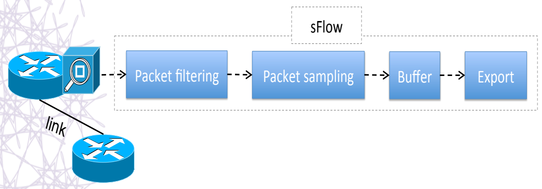
2. Link Dimensioning

- Dimensioning formula proposed in [1]:

$$C(T,\varepsilon) = \rho + \frac{1}{T}\sqrt{-2\cdot\log(\varepsilon)\cdot\upsilon(T)}$$

- Mean traffic ρ is added with a safety margin that depends on the variance υ of traffic at timescale T.
- ε defines the number of time bins in which traffic rate can exceed the estimated capacity $C(T,\varepsilon)$.
- **Main drawback:** needs continuous packet measurements to compute traffic variance at small timescales.

3. sFlow



sFlow sampling algorithm

```
in: sampling rate N
in: input stream of packets observed at the monitoring point
out: output stream of sampled packets

function GetSkip()
    return (rand()%((2*N)-1))+1
end function

procedure SamplePackets(input, output)
    seed random number generator
    s := GetSkip()
    while packet p available from input do
        s := s – 1
        if s = 0 then
            sample packet p to output
            s := GetSkip()
        end if
    end while
end procedure
```

Exporting process:

- Fill in a datagram with as many sampled data as possible and send to the collector.
- At the collector, all the sampled packets within a datagram share the same timestamp.

Challenges of using sFlow for link dimensioning:

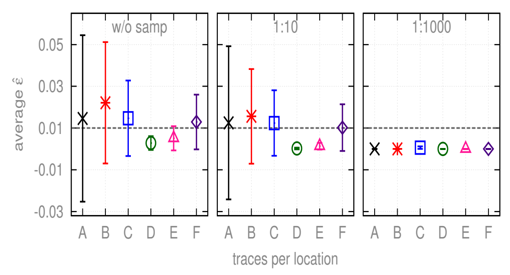
- How to overcome the problem of missing timestamps of individual packets?
- Becomes a problem when unreasonably high sampling rates are used while monitoring lightly-loaded links.

4. Results

- We use a simple procedure to scale traffic mean and variance, where r is the chosen sampling rate:

$$\rho_{est} = \frac{r}{nT}\sum_{i=1}^{n} L_i(T) \qquad \upsilon_{est}(T) = \frac{r^2}{n-1}\sum_{i=1}^{n}(L_i(T)-\rho_{samp})^2$$

- Additional variance from the sampling process is not taken into account and might result in overshoot of the estimated traffic variance.
- $\hat{\varepsilon}$ defines how many time bins of size T the traffic mean was higher than the estimation $C(T,\varepsilon)$. It is desirable that $\hat{\varepsilon} \leq \varepsilon$. In our experiments we set $\varepsilon = 0.01$ (i.e., 1%).



- Results using sampled data as good as, or better than those without sampling.
- Although successful, at higher sampling rates the overestimation happens at undesired levels; and that's why always $\hat{\varepsilon} = 0$.

5. Final Considerations

- We have showed the feasibility of using sFlow sampled data for link dimensioning.
- Our current efforts aim at implementing equations to better estimate traffic variance out of sampled data, not only from sFlow, but also using different sampling strategies.

[1] A. Pras, L. J. M. Nieuwenhuis, R. van de Meent and M. R. H. Mandjes, "*Dimensioning Network Links: A New Look at Equivalent Bandwidth*", IEEE Network, 23(2), 5-10, 2009.

R. de O. Schmidt[1], R. Sadre[2], A. Sperotto[1] and A. Pras[1]
[1]University of Twente, The Netherlands
[2]Aalborg University, Denmark
{r.schmidt,a.sperotto,a.pras}@utwente.nl, rsadre@cs.aau.dk

FLAMINGO UniverSelf UNIVERSITY OF TWENTE. AALBORG UNIVERSITY DENMARK

- Title: Lightweight Link Dimensioning using sFlow Sampling

- Authors: R.de O.Schmidt, R.Sadre, A.Sperotto and A.Pras

- Location: CNSM 2013, Zurich, Switzerland

- Keywords: Link dimensioning and sFlow

The image part with relationship ID rId7 was not found in the file.

- Title: Reports on Internet Traffic Statistics

- Authors: Martijn Hoogesteger, R.de O.Schmidt, A.Sperotto and A.Pras

- Location: TNC 2013, Maastricht, Netherlands

- Keywords: Internet traffic, Statistics