Flow size vs flow duration

There is no relation between flow volume and its duration!!!

A flow can generate lots of packets in a short duration, few packets in a long duration and all in between.
Tons of flows with duration equal to zero

- Main reasons for that:
  - The existence of applications (e.g., NTP) that periodically send control messages -> If this periodicity is bigger than the NetFlow inactive timeout, thousands of flows in cache may expire and be exported with duration equal to zero.
  - The usage of sampling -> sampling increases the chance of not inspecting a packet belonging to an existent flow in cache and therefore increases the chance a flow with a single packet is exported due to inactivity.
Conclusions and future work

- **Can we trust (sampled) NetFlow data?**
  - Yes, BUT it depends on the parameters you rely on!!!

  - **Octets** and **packets** have a good reliability specially if only elephant flows are considered (more packets generated, bigger the chances are the packets will be inspected)

  - On the contrary, flow **duration** is considerably affected (specially when outermost packets are missed) and it should be used with precaution (e.g, flow throughput calculation)

- As a future work, a further investigation on how background traffic may influence sampling would be interesting as well.
Acknowledgements & contact information

- Special thanks to Roel Hoek (University of Twente), Hans Trompert (SURFnet), Maurizio Molina (GÉANT) as well as some of DACS students for their valuable contribution to this work.

Tiago Fioreze
t.fioreze@utwente.nl

Lisandro Granville
granville@inf.ufrgs.br

Aiko Pras
a.pras@utwente.nl

Ramin Sadre
r.sadre@utwente.nl

Anna Sperotto
a.sperotto@utwente.nl