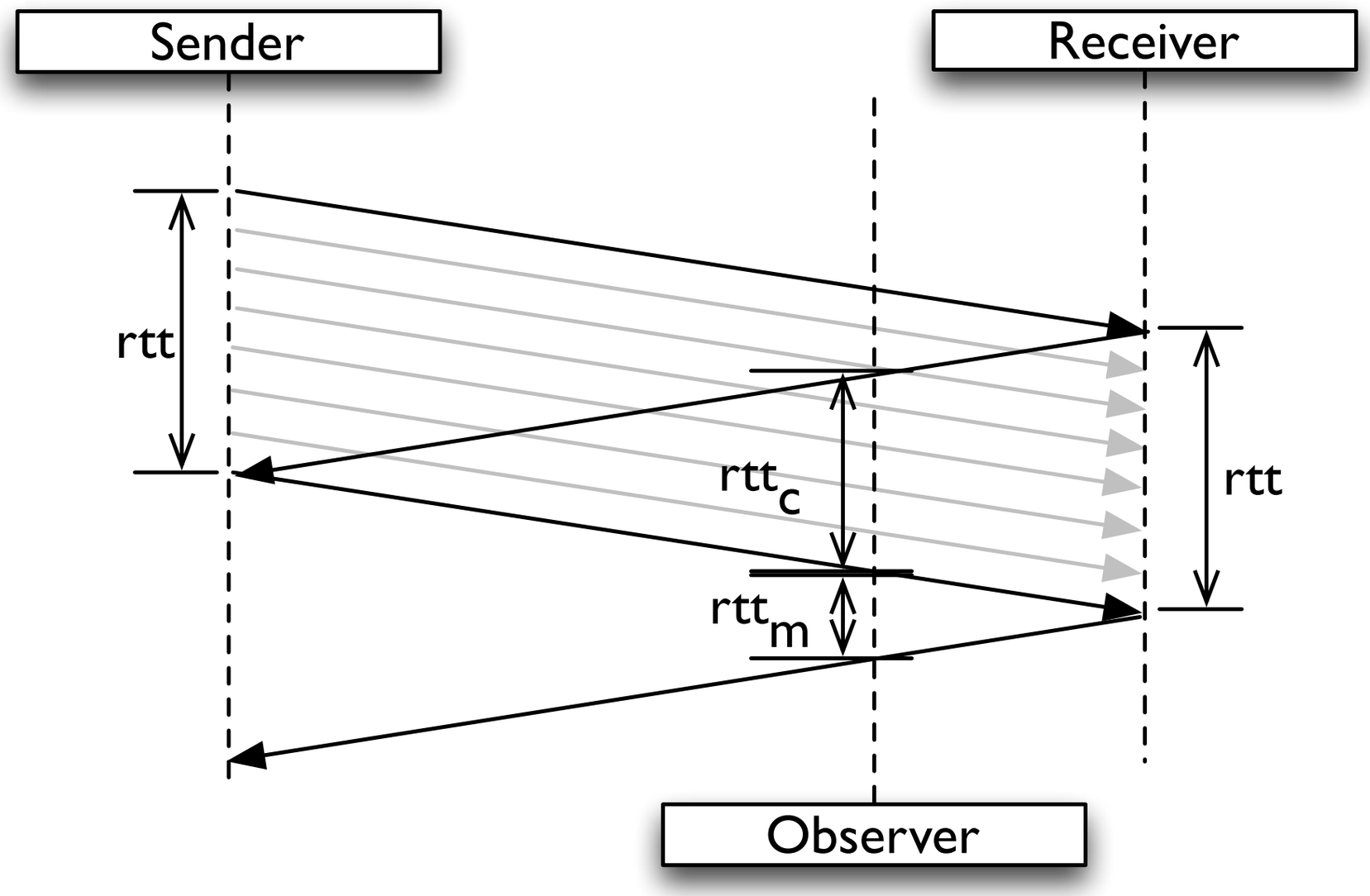
Integrating IPFIX with Pandas
for Exploratory Analysis in Research

Brian Trammell, CSG, ETH Zürich
5th NMRG NetFlow/IPFIX Workshop
Tuesday 30 July 2013 — Berlin

Some Background

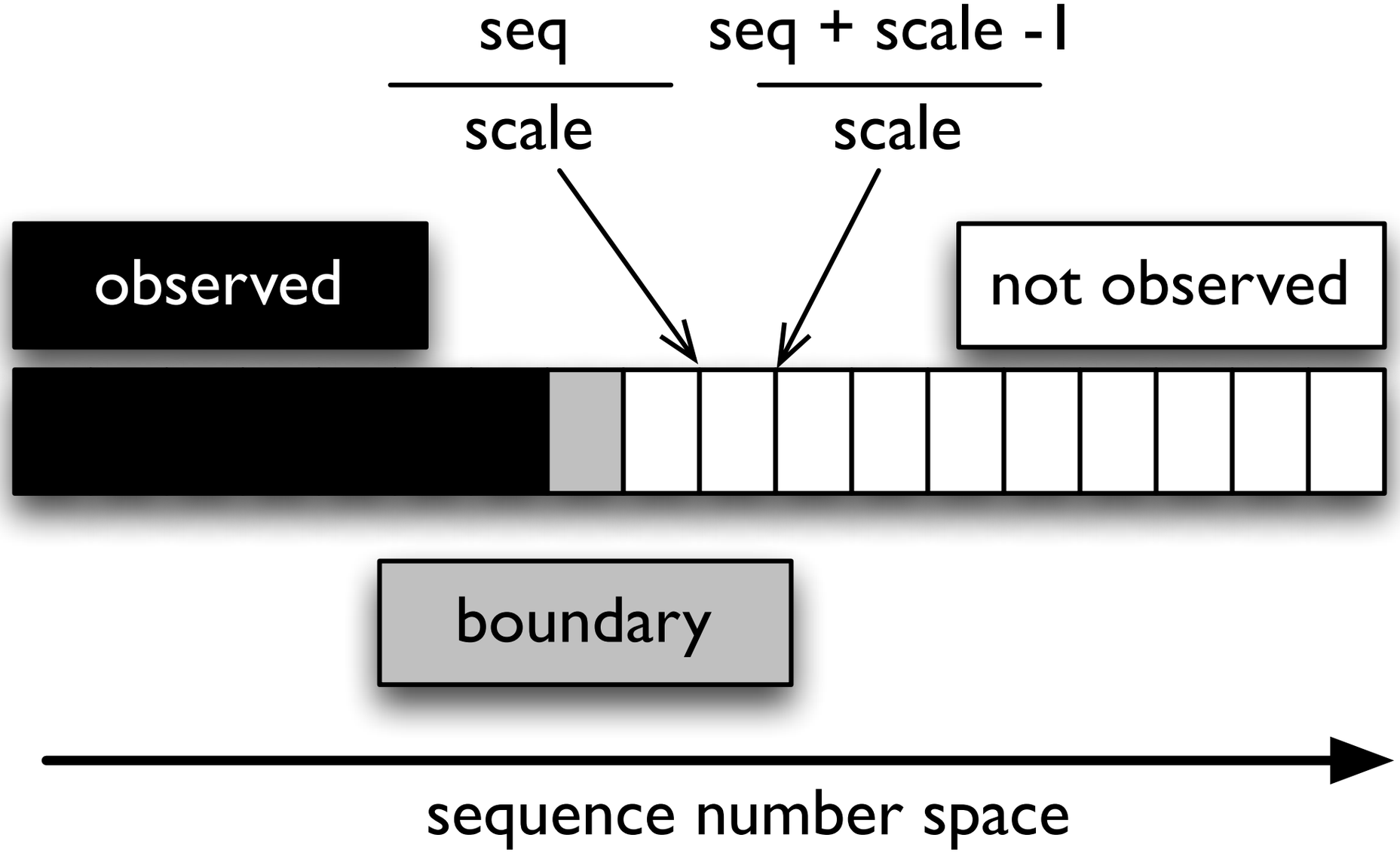
- Goal: augment flow data with TCP performance information
  - Measurement study on effects of network environment on TCP congestion control
- Step 1: QoF: open-source IPFIX MP/EP
  - Fork of CERT/NetSA YAF
    - + TCP-specific Information Elements
    - – DPI features
  - Philosophy: efficiency/scalability over precision
  - Prerelease but available: http://github.com/britram/qof

Efficient Passive RTT Estimation



Veal et al, "New Methods for Passive Estimation of TCP Round Trip Times", PAM 2005

Efficient RTX and Reorder Detection



Analysis

- Great, we have a lot of data!
  - Tools for handling IPFIX don't know about our IEs, because we just made them up.
    - Reinventing the wheel on deadline is a bad idea.
- Need something to enable fast exploratory analysis
  - Understand the shape of the data
  - Direct next questions to ask
  - Find bugs in prerelease code

Pandas to the rescue!

- IPython: interactive execute-explore workflow
  - "Notebooks": annotated interactive code
- Pandas: tools for exploring large datasets, based on numpy/scipy
  - Python interface for simplicity
  - C and Fortran machinery for speed
  - matplotlib for visualization

ipfix for python

- http://pypi.python.org/pypi/ipfix
- pip install ipfix *or* easy_install ipfix
- docs: britram.github.io/python-ipfix
- source: github.com/britram/python-ipfix
- manipulation of IPFIX templates, messages, and message streams
- bridge to python dict and tuple types

Let’s see it in action

http://nbviewer.ipython.org/urls/raw.github.com/britram/qof/nmrg-berlin/pytools/nmrg-talk.ipynb

Acknowledgments

- FP7-mPlane (http://www.ict-mplane.eu)
- Nevil Brownlee & ITS,
  University of Auckland, New Zealand