Longitudinal Analysis of Major Video Streaming Services in the US

Donna Hooshmand, Reza Rejaie PhD

Department of Computer and Information Sciences, University of Oregon,

1.Introduction

- ❖ Video streaming applications (e.g., Netflix, Hulu,…) have become increasingly popular over the Internet in the past decade.
- ❖ It is important to gain insight into their relative popularity and other characteristics.
  - ❖ Ascertaining traits of popular applications is beneficial to determining what aspects contribute to an applications popularity.
- ❖ This study relies on the data from UOnet to perform a longitudinal analysis on the characteristics of popular internet applications.

2.Research Question

In this study we will answer some questions including the following:

- ❖ What percentage of UO traffic is comprised of video streaming applications (e.g., Netflix, Hulu)?
- ❖ What is the duration of the time users use that application?
- ❖ How has the bandwidth (or popularity) of video streaming applications changes over the past few years?
- ❖ Using machine learning and clustering algorithms, could we determine whether a connection is associated with a specific video streaming application based on its network level signature?

3.Methods

- ❖ First, raw data must be turned into the proper format AKA cooked data. Here are some main features of a selected Daily snapshot of a UOnet netFlow data showing the time, size, autonomous system (AS), and internet protocol (IP).
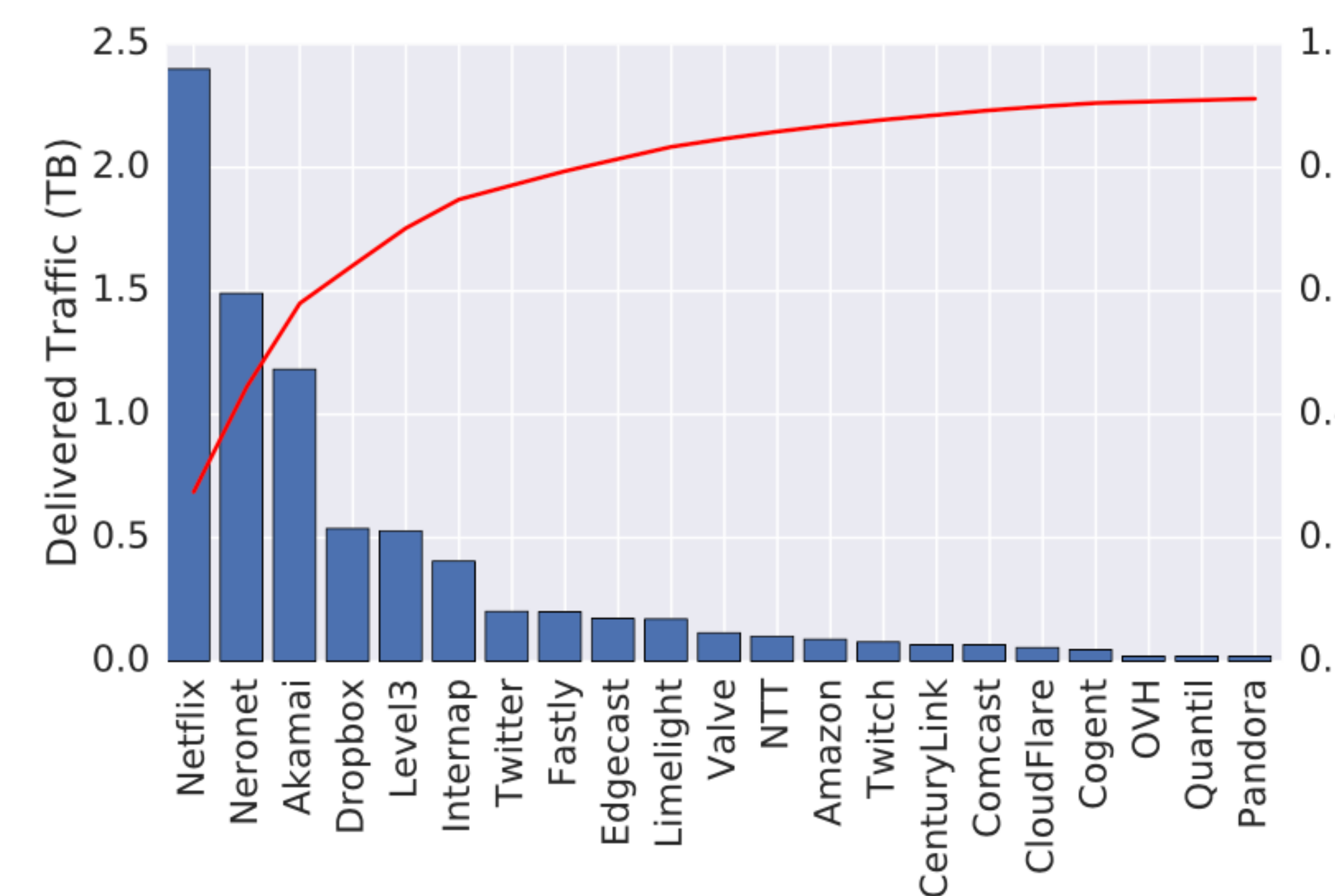
Snapshot	Flows (M)	TBytes	ASes (K)	IPs (M)
10/04/16	196	8.7	39	3.3
10/05/16	193	8.5	37	3.0
10/11/16	199	9.0	41	4.1
10/12/16	198	9.1	41	4.7
10/18/16	202	8.8	40	3.7
10/19/16	200	9.1	38	3.3
10/25/16	205	8.7	37	2.9
10/26/16	209	9.1	40	4.1
11/01/16	212	8.6	39	3.5
11/02/16	210	8.7	40	4.3

- ❖ Running a parser on this data can isolate and filter the information that is relevant to our question.
  - ❖ The created parser consists of different modules that can be specified by the user to get relevant data needed to answer the proposed questions.
- ❖ For example, if one module is for finding the organization for a flow, then this module maps the IP to it's AS, which can then be mapped to the organization.
  - ❖ Previous studies at UO that have used this technique. Here is the result of one focused on traffic localization:
- ❖ This figure depicts the volume of delivered traffic from individual content providers (CPs) to UOnet along with the CDF of aggregate fraction of traffic by top 21 CPs in the snapshot from 10/04/16.

- ❖ Using these same methods, it is possible to determine how various features affect popularity.

4.Project Status

- ❖ This is an ongoing project that was started 1 month ago. Currently, a parser is being developed and is estimated to be fully operational soon.
- ❖ For now, the working parser parses the data in per seconds, per minute, and per hour intervals.
- ❖ Currently a few methods of finding the best way to map IPs to ASes are being tested.
- ❖ From the support for the summer through (NSF-REU), the research will be focusing on fully developing the parser and doing a statistical analysis to determine what makes applications popular.

References

- ❖ **A View From the Edge: A Stub-AS Perspective of Traffic Localization and its Implications**
  Bahador Yeganeh, Reza Rejaie, Walter Willinger
  IEEE/IFIP Network Traffic Measurement and Analysis Conference (TMA), Dublin, Ireland, June 2017
  [acceptance rate 35%]
  One of the top three papers selected for the best paper award.
- ❖ (The provided graphs are from this research).

Acknowledgments