Exploring Models of Internet Traffic

Ashley Luo[1], Patrick Devlin[2], Jeremy Kepner[1], Erin Meger[3]

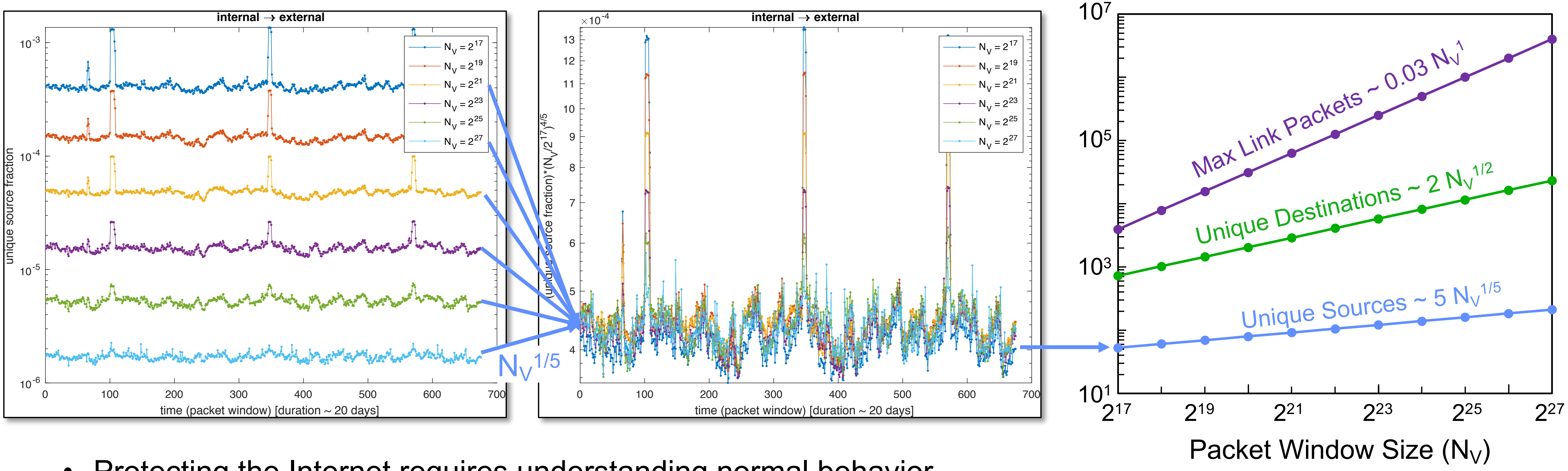
[1]MIT, [2]Yale University, [3]Université du Québec à Montréal

GRAPL Conference
October 9-11, 2020

Massachusetts Institute of Technology

Motivation and Previous Work:
Internet Background Scaling in 100,000,000,000 Packets

Raw Data | Scaled Data | Scaling Relationships



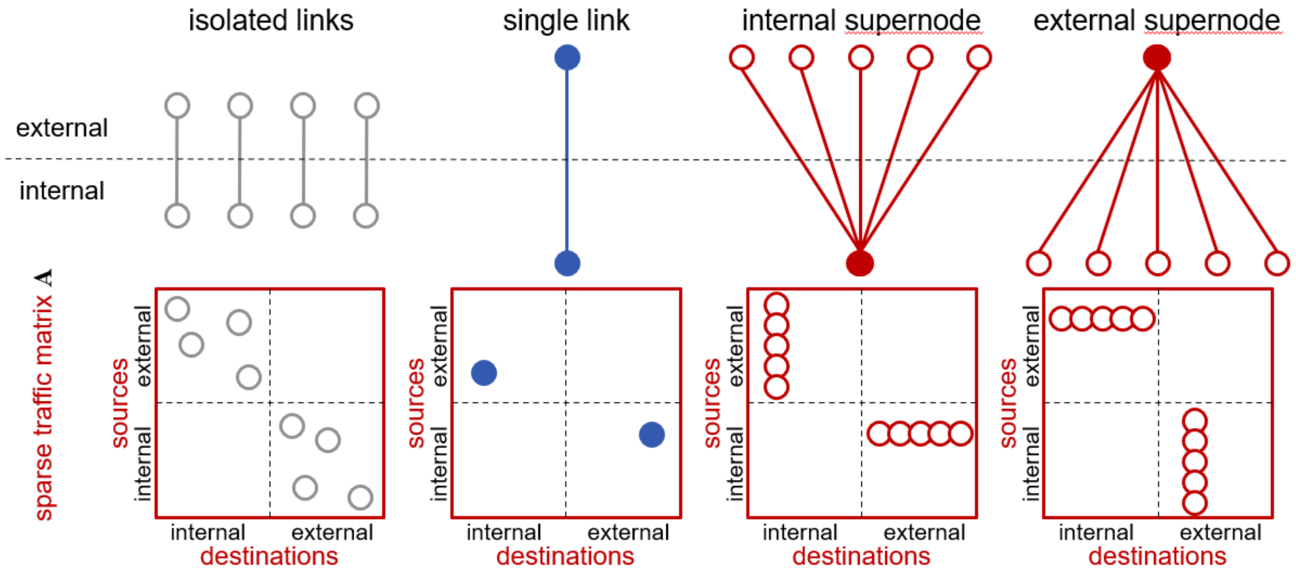
- Protecting the Internet requires understanding normal behavior
- Analysis of 100,000,000,000 packets reveals many new relationships
  - Background traffic strongly scales with packet window size N_V (proxy for time window)

Need new theoretical models to understand and explain these relationships

Multi-Temporal Analysis and Scaling Relations of 100,000,000,000 Network Packets, Kepner et al, IEEE HPEC 2020

Topological Discoveries

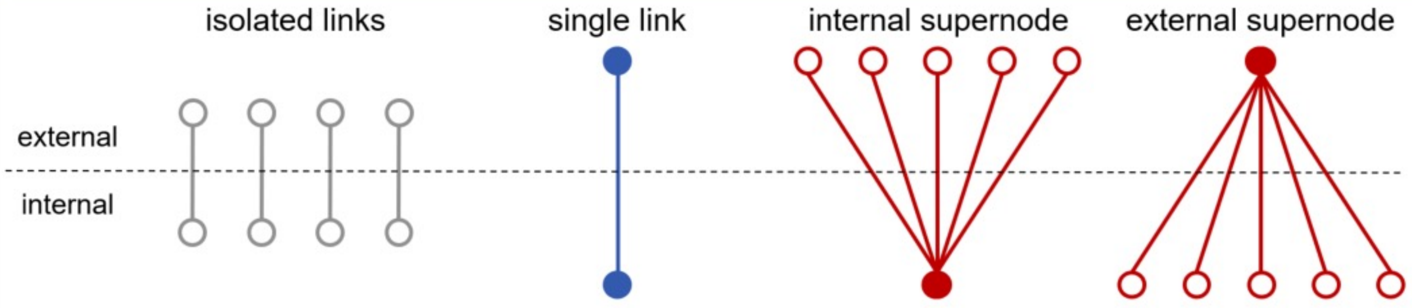
- Identified 4 simple traffic topologies: isolated links, single link, and an internal and external supernode
- We construct a model that consists of a matrix to represent all traffic over a network
  - Each entry in the matrix = # of packets sent from a source to a destination
  - Rows = sources
  - Columns = destinations



Two representations of our network

Implementation Details

- The 4 simple traffic topologies had either linear scaling or constant
- We now map these to new observations with new scaling relations
- Looking at the data available, we fit the 4 simplified traffic topologies to these observations by considering extremal network models and new parameters



	isolated links	single link	internal supernode	external supernode
Primary Characteristic	Number of Links	Number of Packets	Number of Leaf Nodes	Number of Leaf Nodes
Model	IsolatedLinks(N_V)	SingleLinkPackets(N_V)	ExternalLeaves(N_V)	InternalLeaves(N_V)

Define new parameters based on a topology’s primary characteristic

Hybrid Model

Two networks in the hybrid model:

- Underlying network of ”true” connections
- Sampled network through probability and data collection

Three main components to the new model:

- Core: preferential attachment, highly connected core
- Unattached nodes: small connected star components with few neighbors
- Leaves: degree 1 nodes adjacent to vertices in the core

Model generation algorithm including probabilistic sampling to simulate data collection methods

Model Parametrization

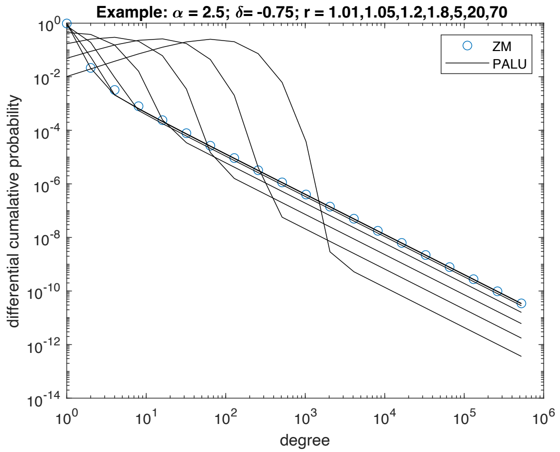
- λ average degree of the unattached nodes in underlying network
- *C,L,U* proportions of nodes in each of the core, the leaves, and the unattached nodes in the underlying network, conforming to the relationship

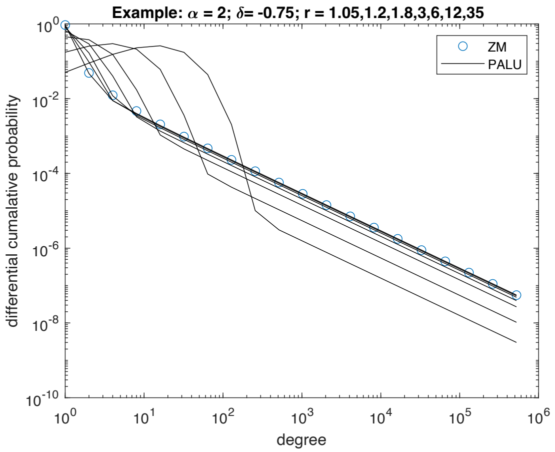
$$C + L + U\left(1 + \lambda - e^{-\lambda}\right) = 1$$

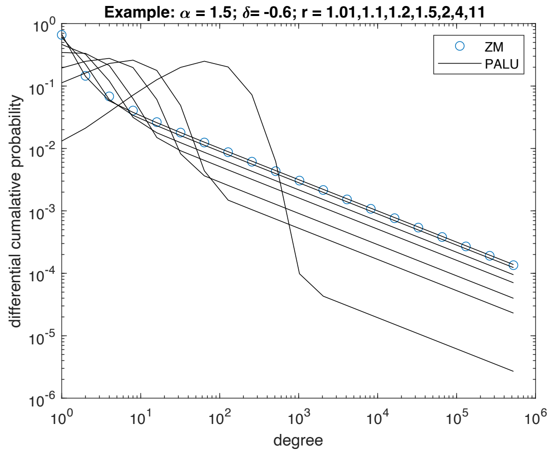
- α exponent of power-law decay of the degree distribution of the underlying core
- *p* proportion of underlying network being observed

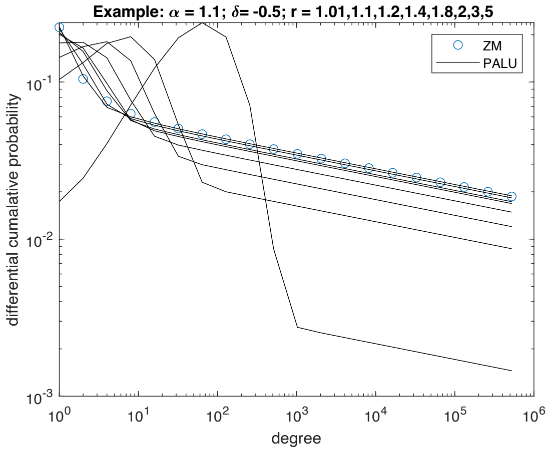
Parameterized sampling

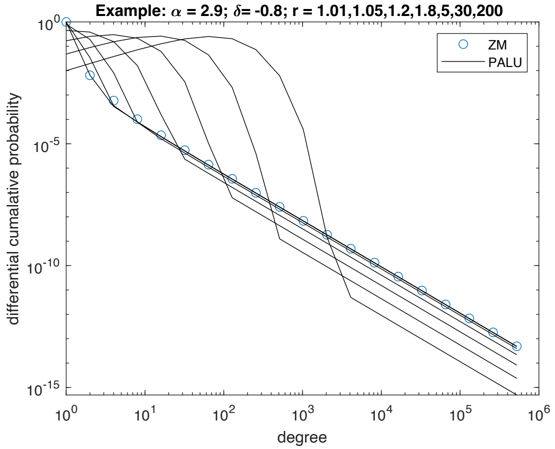
Results









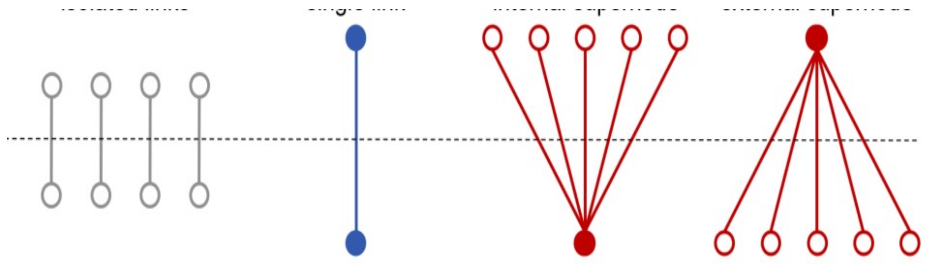


Hybrid model fits Zipf-Mandlebrot distribution

Hybrid Model: Underlying Network and Sampled Network

Three main components to the new model:

- Core: preferential attachment, highly connected core
- Unattached nodes: small connected star components with few neighbors
- Leaves: degree 1 nodes adjacent to vertices in the core

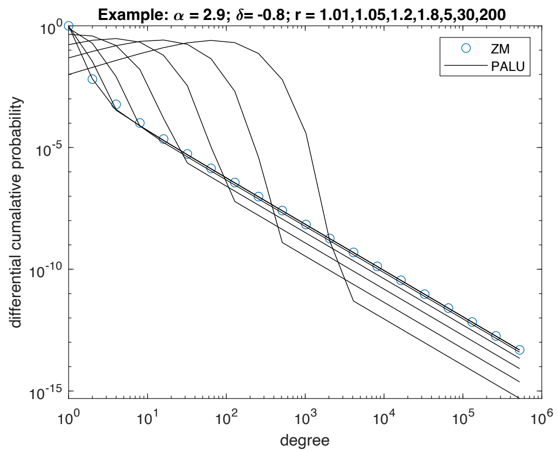


Four parameters:

- λ average degree of the unattached nodes in underlying network
- *C,L,U* proportions of nodes in each of the core, the leaves, and the unattached nodes in the underlying network, conforming to the relationship

$$C + L + U\left(1 + \lambda - e^{-\lambda}\right) = 1$$

- α exponent of power-law decay of the degree distribution of the underlying core
- *p* proportion of underlying network being observed



Model generation algorithm including probabilistic sampling to simulate data collection methods