In-Degree Dynamics of Large-Scale P2P Systems

Zhongmei Yao (University of Dayton)
Daren B.H. Cline (Texas A&M University)
Dmitri Loguinov (Texas A&M University)

ACM HotMetrics 2010
June 18
Agenda

• Motivation and background
  – Peer churn and Palm-Khintchine Theorem
• General Edge-Creation Model
• Edge Arrival Process
• In-Degree
• Wrap-up
Dynamics of Distributed Systems

• System of \( n \) nodes
  – ON (green) and OFF (grey) states

• Each user selects \( k \) out-going neighbors
  – Repair links upon neighbor failure

• Want to know in-coming edges of a node
  – More in-links, smaller isolation probability
  – More in-links, more likely this node will be overloaded
Decompose into Two-State Processes

- Each user $i$ is either ON or OFF [Yao06]:

- Each outlink $c$ is ALIVE/DEAD:

- No complete modeling framework in prior work; no rigorous results on in-degree dynamics [Leonard07]
Edge Arrival Process

- Let $\xi_{n,i}(t)$ be a marked point process
  - Mark processes $Y_i^c(n, t)$ if user $i$ delivers edges to peer $v$
- The edge arrival process to node $v$ is $\sum_{i=1}^{n} \xi_{n,i}(t)$
  - Superposition of $n$ point processes!
The Classic Poisson Result

“Under mild conditions, the superposition of $n$ independent stationary renewal processes approaches Poisson …

- Let $M_{n,i}(t)$ count the number of renewals in interval $[0, t]$ with inter-arrival time distribution $F_{n,i}(t)$
- The Palm-Khintchine theorem [Heyman and Sobel]:
  Process $M_n(t) := \sum_{i=1}^n M_{n,i}(t)$ converges in distribution to a homogeneous Poisson process as $n \to \infty$ if
    - Processes $M_{n,i}(t)$ are stationary and independent
    - Given any $\epsilon > 0$, for each $t > 0$ and $n$ sufficiently large, $F_{n,i}(t) \leq \epsilon$ for all $i$
    - And the aggregate arrival rate converges to a constant: $\lim_{n \to \infty} \sum_{i=1}^n \lambda_{n,i} \to \lambda$
Focus of This Paper

1. A complete modeling framework for understanding peer churn and in-degree dynamics

2. Superposition of a large number of dependent marked point processes
Agenda

• Motivation and background
  – Peer churn and Palm-Khintchine Theorem
• General Edge-Creation Model
• Edge Arrival Process
• In-Degree
• Wrap-up
Modeling Assumptions

• **Assumption 1:** The number of outlinks $k$ a user monitors is a constant for all $n$

• **Assumption 2:**
  1) Given a fixed set of user types, the user ON/OFF durations of type $j$ respectively follow CDFs $F^{(j)}(x)$ and $G^{(j)}(x)$ with finite means
  2) Each user ON/OFF duration CDF is labeled with type $j$ with probability $p_j$, where $\sum_j p_j = 1$
  3) Given that users have chosen their types, $\{Z_i(t)\}_{i=1}^n$ are mutually independent, stationary alternating renewal processes
Dependency

- Edge creation processes are dependent
  - Multiple users may **concurrently** connect to the same neighbor
  - Each out-link may point to a peer $v$ again after $v$ **re-appears** in the system
- User $i$’s current selection **depends on the history** observed by $i$
  - As a result, the model for the number of users available at each $z$-th selection time is intricate

$$Y_i^c(n, t)$$

$\uparrow$

$i = 1, \ldots, n$
$c = 1, \ldots, k$
Main Theorem

- Define $\xi_{n,i}(t)$ to be the edge arrival process from $i$ to $v$:

  $$\xi_{n,i}(t) := \sum_{z=1}^{W_i(n,t)} I_{i,z}^v$$

- Theorem 1: Under Assumptions 1-2 and uniform selection, conditioned on $Z_v$, the superposition $\sum_{i=1}^n \xi_{n,i}(t)$ converges in distribution as $n \to \infty$ to a non-homogeneous Poisson process with local rate $\gamma Z_v(t)$
Proof Overview

- Our main task is to show [Resnick87]:
  - Continuity: the probability that no point occurs exactly at time $t$ is 1
  - Mean convergence:
    \[
    \forall t > 0 : \lim_{n \to \infty} E \left( \sum_{i=1, i \neq v}^{n} \xi_{n,i}(t) | Z_v \right) = \int_0^t Z_v(u) du
    \]
  - Probability convergence:
    \[
    \forall t > 0 : \lim_{n \to \infty} P \left( \left( \sum_{i=1, i \neq v}^{n} \xi_{n,i}(t) \right) = 0 | Z_v \right) = \exp \left( -\gamma \int_0^t Z_v(u) du | Z_v \right)
    \]
Proof Overview 2

- As $n$ increase, the probability that each user $i$ selects *any* other peer *more than once* in $[0, t]$ becomes smaller
  - The edge arrival process from each $i$ to $v$ becomes *sparser*

- To *bound* the above probability, we must first show that *moments* of collection $\{W_i(n, t)\}_{n>1}$ exist for all $n$
  - Lemma 3 in the paper

- The mean number of edges created by each $i$

  $$
  \lim_{n\to\infty} E[W_i(n, t)|i\text{'s type}] = k\lambda_i t E[U(L_i)|i\text{'s type}]
  $$

- The edge arrival rate $\gamma$ to user $v$ when $v$ is alive must converge

  $$
  \gamma = \lim_{n\to\infty} E \left[ \frac{\sum_{i=1}^{n} W_i(n, t)}{t} \cdot \frac{1}{\text{number of live users}} \right]
  $$

  - Arrival rate of user $i$
  - # of selections per link within $i$’s lifetime
  - $W_i(n, t)$
Simulations

Pareto lifetimes with shape parameter = 3

Pareto lifetimes with shape parameter = 1.5
Agenda

• Motivation and background
  – Peer churn and Palm-Khintchine Theorem
• General Edge-Creation Model
• Edge Arrival Process
  • In-Degree
• Wrap-up
Main Theorem

• **Theorem 2:** Under Assumptions 1-2 and uniform selection, given that a user is alive in the system, its expected in-degree at fixed age $s > 0$ converges as $n \to \infty$ to a monotonically increasing function of age:

$Z_v(t)$

\[
\lim_{n \to \infty} E[X_n(s)] = kE[U(R) - U(R - s)]
\]

\[
= k \int_0^\infty (E[U(x) - U(x - s)]) \, dH(x)
\]
Simulations

Exponential lifetimes

Pareto lifetimes with shape parameter $= 3$
Wrap-up

• A generic modeling framework for understanding user join/departure, edge arrival, and in-degree
• Closed-form results on the edge-arrival process to each user and the transient in-degree
  – Proofs in technical report
• Open problems:
  – Non-uniform selection
  – Non-stationary user churn