WHAT STARTS HERE CHANGES THE WORLD
THE UNIVERSITY OF TEXAS AT AUSTIN

Towards Automated Performance Diagnosis in a Large IPTV Network

Ajay Mahimkar, Zihui Ge, Aman Shaikh, Jia Wang, Jennifer Yates, Yin Zhang, Qi Zhao

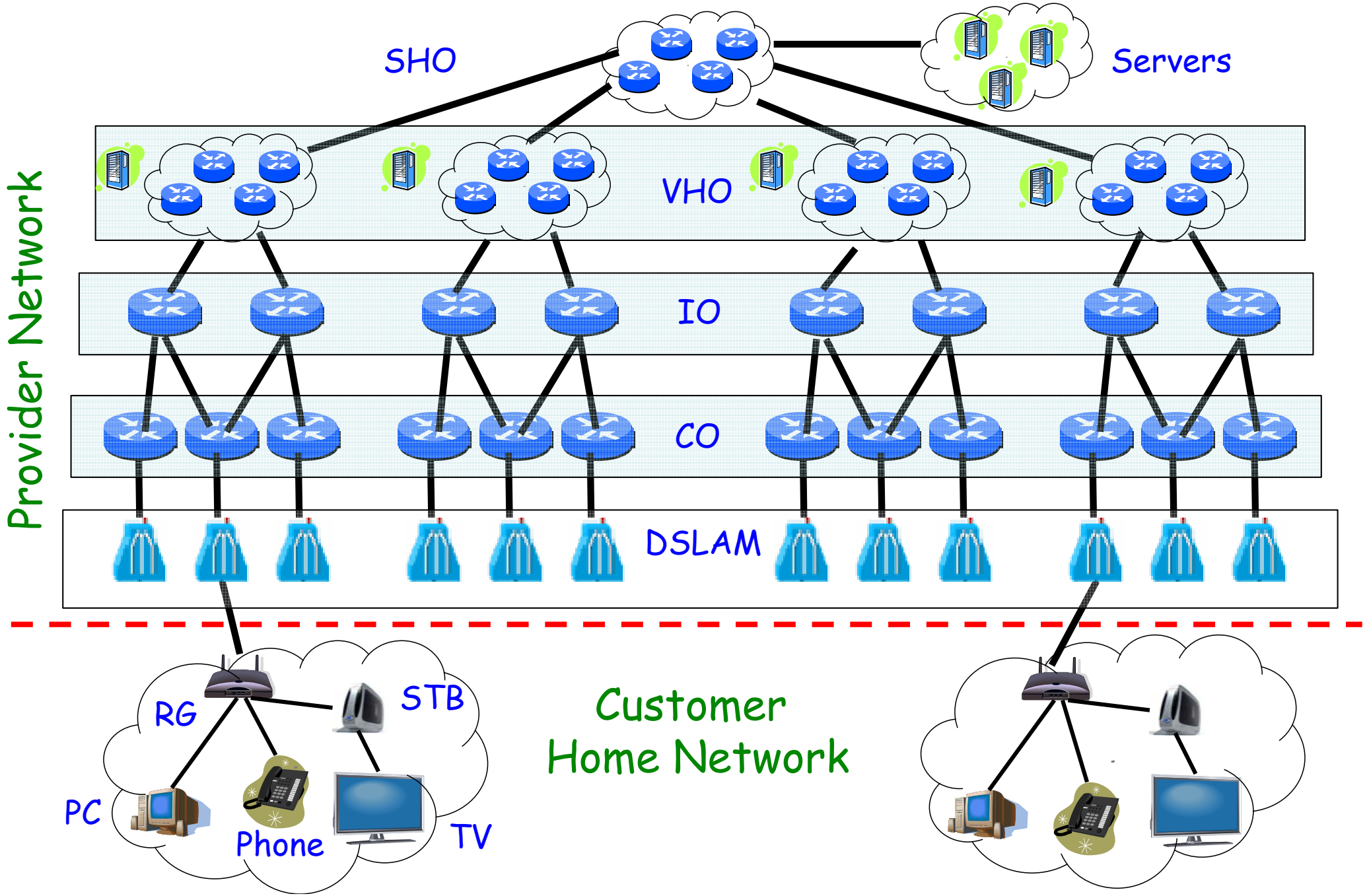
UT-Austin and AT&T Labs-Research
mahimkar@cs.utexas.edu

ACM Sigcomm 2009

Internet Protocol Television (IPTV)

- Television delivered through IP network
  - Encoded in series of IP packets
- Rapid deployment by telecom companies
  - New services: quadruple-play (digital voice, TV, data & wireless)
  - More flexibility and interactivity for users
- One of the largest commercial IPTV deployments in US
  - By 2008, more than 1 million customers spanning 4 time-zones
  - Supports advanced features
    - Digital video recording (DVR), Video on demand (VoD)
    - High definition (HD) channels, Choice programming

IPTV Service Architecture



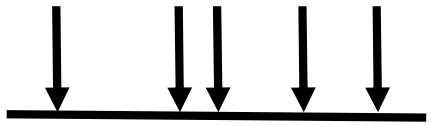
IPTV Characteristics

- Stringent constraints on reliability and performance
  - Small packet loss or delay can impair video quality
- Scale
  - Millions of devices (routers, servers, RG, STB)
  - Number is growing
- Complexity
  - New service
  - New application for native IP multicast
    - Operational experience with IP multicast is limited

Problem Statement

- Characterize faults and performance in IPTV networks
  - What are the dominant issues?
  - Is there spatial correlation between different events?
  - Is there daily pattern of events?

- Detect and troubleshoot recurring conditions
  - Temporal – repeating over time
    - E.g., recurring poor picture quality at TV



  - Spatial – re-occurring across different spatial locations
    - E.g., software crashes at multiple set-top-boxes within a region

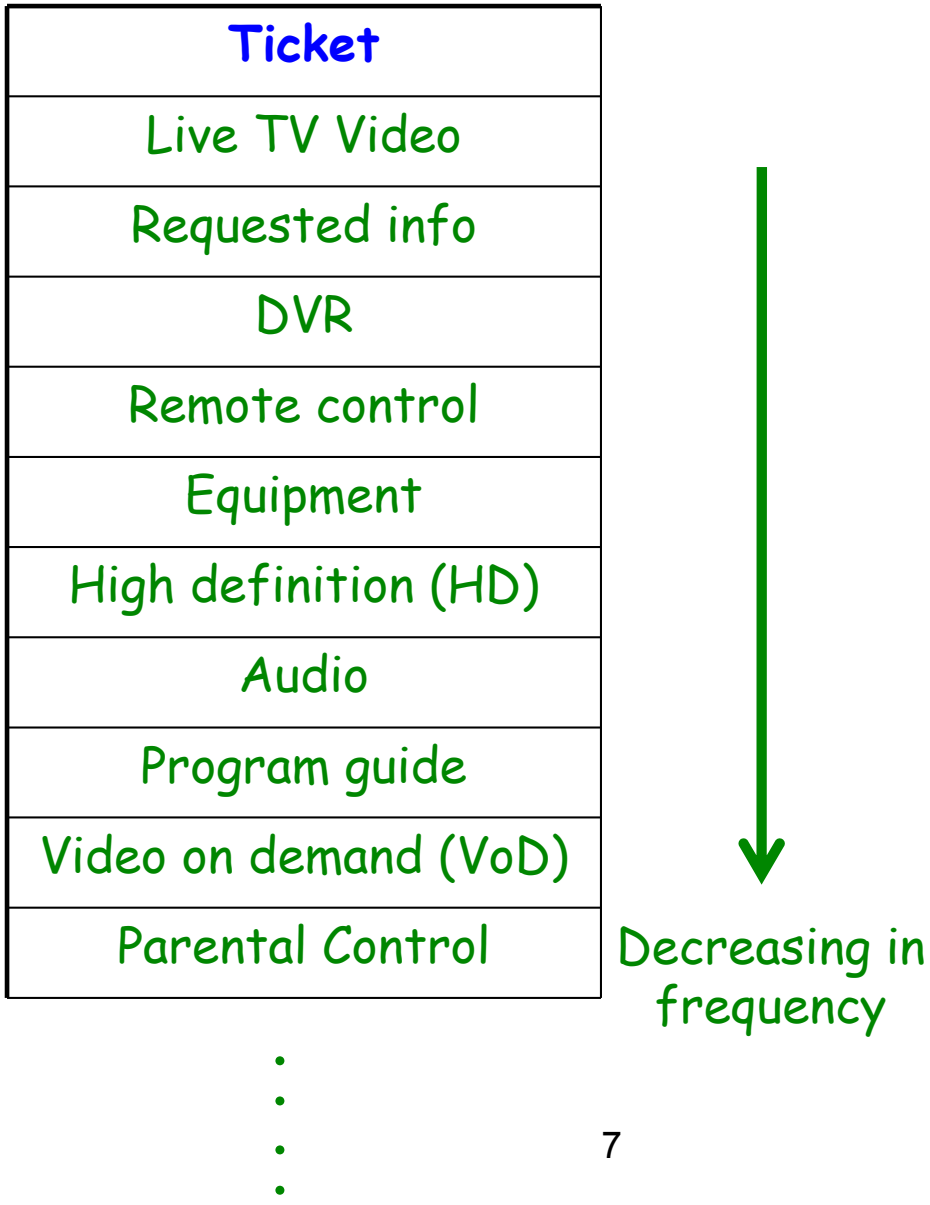
Lots of alarms. How do you identify recurring conditions of interest to network operators?

IPTV Measurement Data

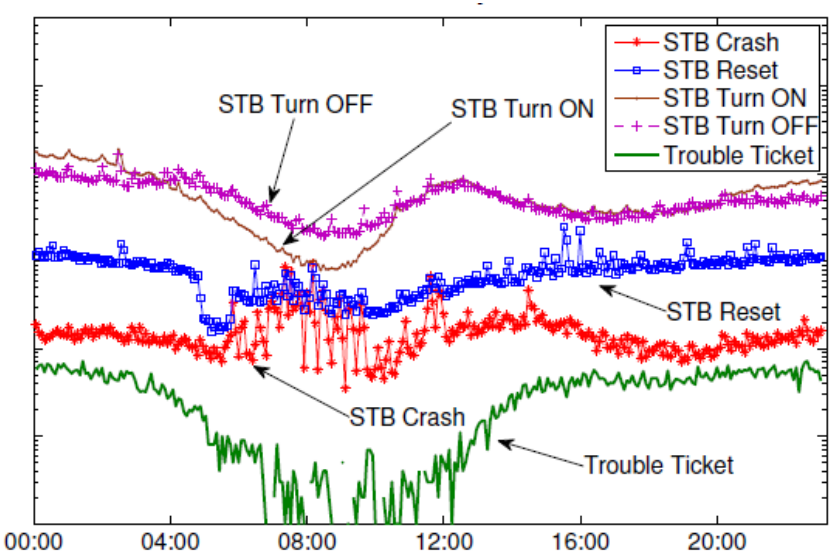
- Customer care call records
  - Trouble tickets related to billing, provisioning, service disruption
- Home network performance / activities
  - User activities: STB power on/off, STB resets, RG reboots
  - Performance/Faults: STB software crashes
- Provider network performance
  - Syslogs at SHO, VHO
  - SNMP (CPU, memory, packet counts) at SHO, VHO, IO, CO
  - Video quality alarms at monitors

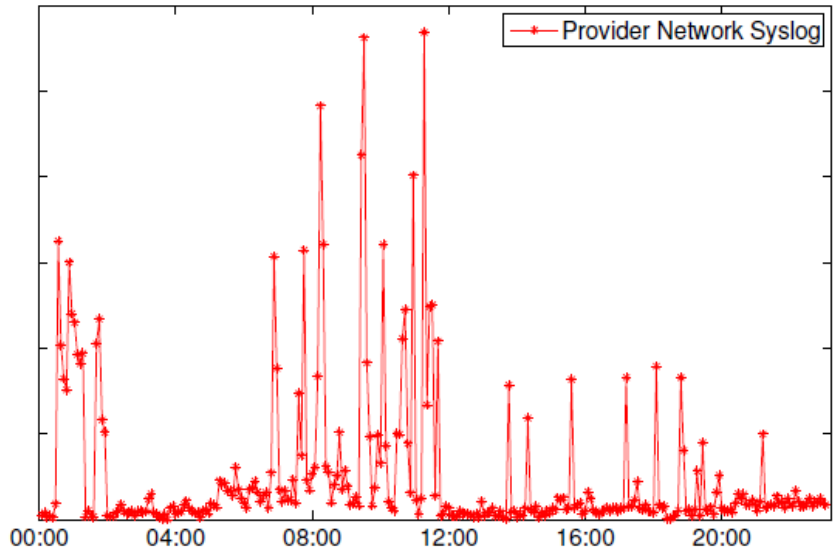
IPTV Characterization

- Data analyzed over 3 months
- Customer Trouble Tickets
  - Performance related issues
  - Sample Live TV video tickets
    - Blue screen on TV
    - Picture freezing
    - Poor video quality
  - Small degree of spatial correlation



Daily Pattern of Events





Lot of activity between:

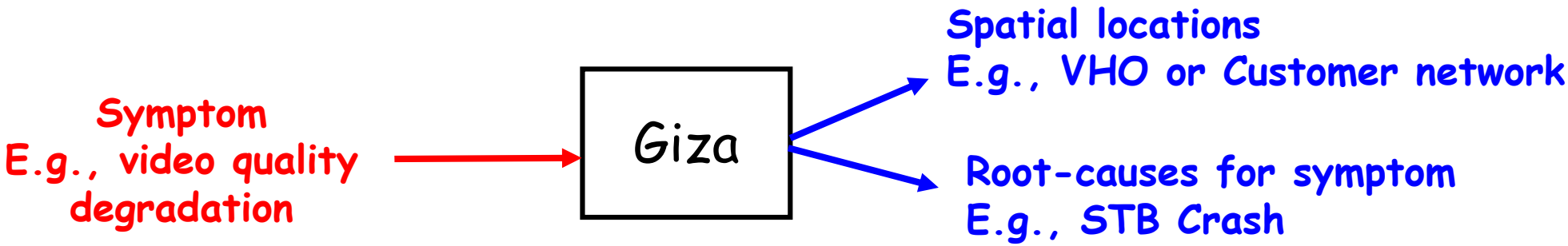
- 00:00 and 04:00 GMT (evening prime time)
- 12:00 and 23:59 GMT (day time)

Relatively quiet period between

- 4:00 and 12:00 GMT (customers are sleeping)
- Number of syslogs at SHO/VHO is high. (provisioning and maintenance)

Giza

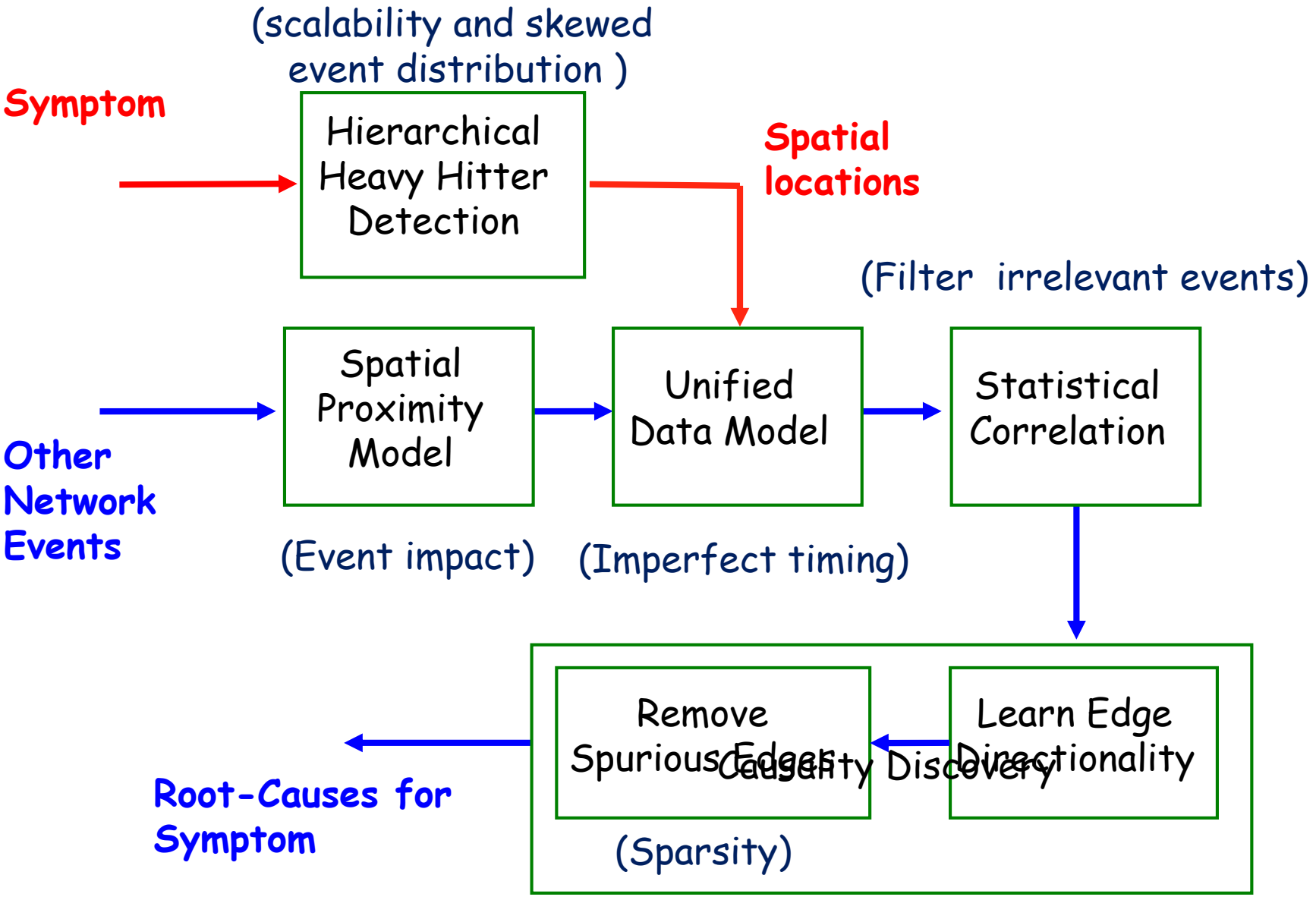
- First multi-resolution infrastructure
  - Detect and troubleshoot spatial locations that are experiencing serious performance problems
  - Detection
    - Output locations (e.g. VHO, or STB) that have significant event counts
  - Troubleshooting
    - Output list the other event-series that are the potential root-causes



Mining Challenges

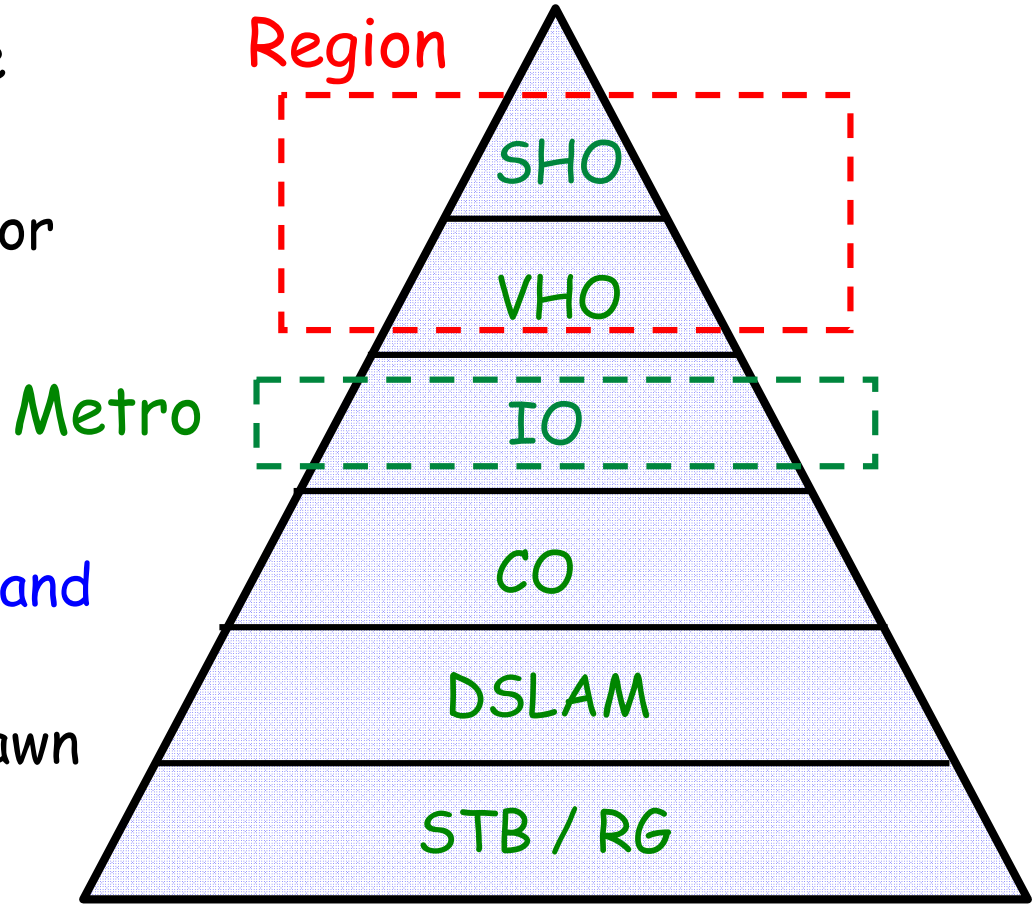
- Massive scale of event-series
  - Each device (SHO, VHO, RG) can generate lots of event-series
  - Blind mining could easily lead to information snow of results
- Skewed event distribution
  - Small frequency counts for majority of events
    - Insufficient sample size for statistical analysis
  - Heavy hitters do not contribute to majority of issues
- Imperfect timing information
  - Propagation delay: From event location to measurement process
  - Distributed events: From root of tree (SHO) towards RG

Giza


(scalability and skewed event distribution)
Symptom
Hierarchical Heavy Hitter Detection
Spatial locations
(Filter irrelevant events)
Spatial Proximity Model
Unified Data Model
Statistical Correlation
Other Network Events
(Event impact)
(Imperfect timing)
Remove Spurious Edges
Learn Edge Directionality
Causality Discovery
(Sparsity)
Root-Causes for Symptom

Hierarchical Heavy Hitter Detection

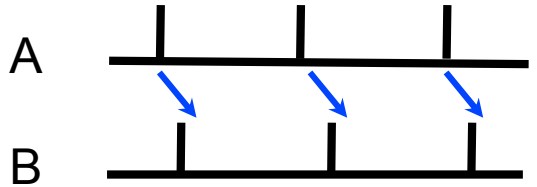
- Identify spatial regions where symptom count is significant
  - E.g., customers in Texas have poor video quality
- Significance test
  - Account for (i) event frequency and (ii) density concentration
  - Null hypothesis: children are drawn independently and uniformly at random from lower locations



IPTV Pyramid Model

Causal Graph Discovery

- First, learn edge directionality
  - Idea: use approximate timing to test statistically precedence
    - Using lag correlation
    - Compute cross-correlations at different time-lags
    - Compare range of positive lags with negative lags
  - Addresses
    - Imperfect timing
    - Auto-correlation within each event-series



A statistically precedes B

- Second, condense correlation graph
  - Idea: identify smallest set of events that best explains symptoms
    - Using L1-norm minimization and L1-regularization
  - L1-regularization achieves sparse solution

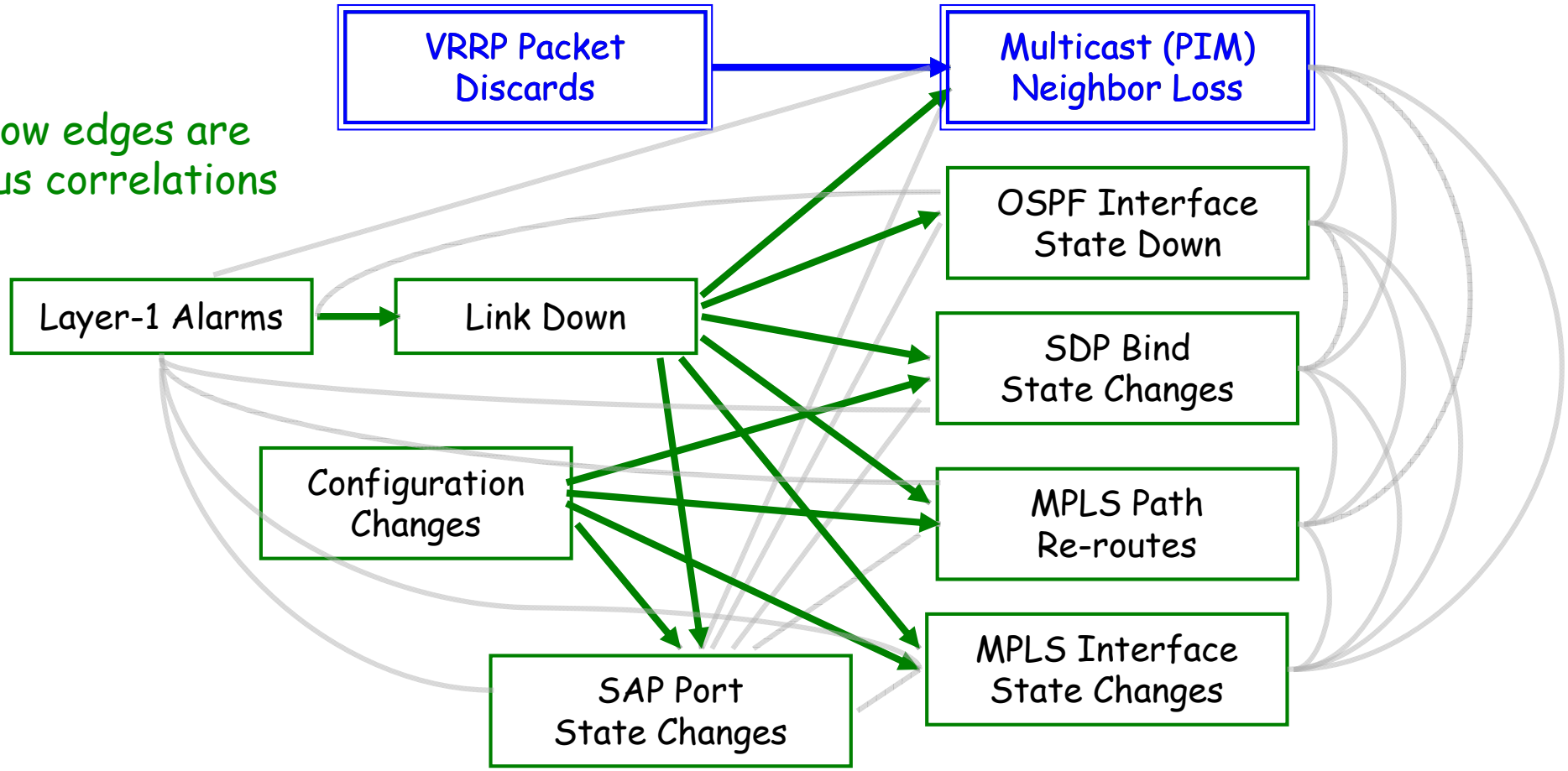
Giza Experiments

- Validation
  - Select customer trouble tickets as input symptom
  - Apply Giza to identify potential root causes
  - Ground truth: mitigation actions in customer trouble tickets
  - Result: Good match between ground truth and Giza output

    Majority of tickets explained by home network issues
- Case Study: Provider network events
  - Identify dependencies with customer trouble tickets
  - Result: Giza discovered unknown causal dependency
- Causality Discovery Comparison
  - Ground truth: networking domain knowledge
  - Result: Giza performs better than WISE (Sigcomm'08)

Case Study: Provider Network Events

Shadow edges are spurious correlations



- Dependency between VRRP packet discards and PIM timeouts was unknown
- Behavior more prevalent within SHO and VHOs near SHO
- We are investigating with operations team

Comparison to Related Literature

	Sherlock Kandula et al. Sigcomm'07	WISE Tariq et al. Sigcomm'08	Orion Chen et al. OSDI'08	NICE Mahimkar et al. CoNEXT'08	NetMedic Kandula et al. Sigcomm'09	Giza Mahimkar et al. Sigcomm'09
Original Focus	Enterprise Networks	Content Distribution Networks	Enterprise Networks	ISP Backbone Networks	Enterprise Networks	IPTV Networks
Eliminate spurious dependencies		YES				YES
Achieve sparse solution						YES
Automated edge directionality	YES	YES	YES		YES	YES
Multi-resolution analysis						YES

Conclusions

- First characterization study in operational IPTV networks
  - Home networks contribute to majority of performance issues
- Giza - Multi-resolution troubleshooting infrastructure
  - Hierarchical Heavy Hitter detection to identify regions
  - Statistical correlation to filter irrelevant events
  - Lag correlation to identify dependency directionality
  - L1-norm minimization to identify best explainable root-causes
- Validation and case studies demonstrate effectiveness
- Future Work
  - Troubleshooting home networks
  - Network-wide change detection

Thank You !