Profiling Internet Backbone Traffic: Behavior Models and Applications

Kuai Xu, Zhi-Li Zhang, and Supratik Bhattacharyya

University of Minnesota                   Sprint ATL

August 24, 2005
Why profile traffic?

- Changes in Internet traffic dynamics
  - increase in unwanted traffic
  - emergence of disruptive applications
  - new services on traditional ports
  - traditional service on non-standard ports

- Existing tools
  - rely on ports for identifying or classifying traffic
  - report volume-based heavy hitters
  - look for specific or known patterns

- Need better techniques to discover behavior patterns
  - help network operators secure and manage networks
**Communication patterns**

- Underlying communication patterns of end hosts
  - who are they talking to? how are ports used?
  - how many packets or bytes transferred?
- Can communication patterns reveal interesting behavior?
**Problem settings**

- **Problems**
  - how to characterize communication patterns?
  - are these patterns meaningful?
  - how to automatically discover such patterns?

- **Challenges**
  - vast amount of traffic data
  - large number of end hosts
  - diverse applications

- **A more specific problem setting**
  - use one-way traffic data from single backbone link
  - use only packet header information
  - **no assumption of normal (or anomalous) behavior**
Roadmap of our methodology

- Data pre-processing
  - aggregate packet streams into 5-tuple flows
  - group flows into clusters

- Extract significant clusters
  - data reduction step using entropy

- Classify cluster behavior based on similarity/dissimilarity of communication patterns
  - characterize using information theory
  - clusters classified into behavior classes

- Interpret behavior classes
  - structural modeling for dominant activities
Data pre-processing

- Aggregate packet streams into 5-tuple flows
- Group flows associated with same end hosts/ports into clusters
Roadmap of our methodology

- Data pre-processing
  - aggregate packet streams into 5-tuple flows
  - group flows into clusters

- Extract significant clusters
  - data reduction step using entropy

- Classify cluster behavior based on similarity/dissimilarity of communication patterns
  - characterize using information theory
  - clusters classified into behavior classes

- Interpret behavior classes
  - structural modeling for dominant activities
Extract significant clusters

- Focus on significant clusters
  - sufficiently large number of flows
  - represent behavior of significant interest

- One definition: using a fixed threshold
  - a cluster is significant if containing at least x% of flows
  - how to choose x for all links?

- Our definition: adaptive thresholding using entropy
  - a cluster is significant if “standing out” from the rest
  - use entropy to quantify whether the rest looks random
Entropy-based adaptive thresholding

- An iterative process
  - extract significant clusters until the rest look nearly uniform in size

\[ P(\text{srcIP}) \]

\[ \alpha = \alpha_0 \]

\[ p(\text{cluster}) \geq \alpha \]

\[ \alpha = \alpha / 2 \]

Significant Clusters

The Rest

the rest random?

Stop

Yes

No
Sample results

- Packet traces
  - OC-48 link during 24 hours
  - extract clusters every 5 minutes
Roadmap of our methodology

- Data pre-processing
  - aggregate packet streams into 5-tuple flows
  - group flows into clusters

- Extract significant clusters
  - data reduction step using entropy

- Classify cluster behavior based on similarity/dissimilarity of communication patterns
  - characterize using information theory
  - clusters classified into behavior classes

- Interpret behavior classes
  - structural modeling for dominant activities
Understanding behavior patterns

- Still many significant clusters in each time interval
  - can we characterize their behavior patterns?
  - are there similarities/dissimilarities in behavior?
  - communication patterns provide more insight than volume metrics

- What traffic features should we look at? And how?
  - for each cluster, look at distributions of flows by ports and IP addresses
  - distribution summarized by relative uncertainty
  - each cluster characterized by a point in 3-D space
Relative uncertainty

- Entropy: $H(X) = -\sum p(x_i) \log p(x_i)$
- Maximum Entropy: $H_{\text{max}}(X) = \log [\min(m,N)]$
- Relative Uncertainty of variable $X$

$$RU(X) := \frac{H(X)}{H_{\text{max}}(X)}, ~ RU \in [0, 1]$$
- $RU(X) = 0$: $X$ is deterministic
- $RU(X) = 1$: $X$ is randomly distributed
Behavior characterization

srcPort

Low 0
Medium 1
High 2

dstPort

dstIP

RU(dstPrt)
0 0.2 0.4 0.6 0.8 1

RU(srcPrt)
0 0.2 0.4 0.6 0.8 1

RU(dstdPr)
0 0.2 0.4 0.6 0.8 1

Relative uncertainty
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10

Frequency of data
0 50 100 150 200 250 300 350

Frequency of data
0 50 100 150 200 250 300 350

Frequency of data
0 50 100 150 200 250 300 350

Frequency of data
0 50 100 150 200 250 300 350
Behavior classifications

- Behavior classes (BC)
  - summarize three feature distributions into 27 classes
  - \([0, 0, 0] \ldots [2, 2, 2]\), for convenience \(BC_0\) to \(BC_{26}\)

- What is the difference between behavior classes?
  - are there common vs. rare behavior classes?
  - are BCs have many or a few clusters?
  - are memberships in BCs stable?
**Temporal Properties**

- **Metrics**
  - Popularity: how many time slots do we see a BC in?
  - Avg. number of clusters: how many clusters in each BC?
  - Membership volatility: does a BC contain the same clusters over time?

---

![Temporal Properties Diagram](image_url)

- **Common Behavior**
  - Volatile members

- **Rare Behavior**
  - Membership volatility

- **Avg. clusters**
  - Popularity
Summary of behavior classifications

- Behavior classes classify clusters based on communication patterns
- Behavior classes have distinct temporal properties
- Clusters have stable behavior over time

How can we interpret observed behavior?
Roadmap of our methodology

- Data pre-processing
  - aggregate packet streams into 5-tuple flows
  - group flows into clusters

- Extract significant clusters
  - data reduction step using entropy

- Classify cluster behavior based on similarity/dissimilarity of communication patterns
  - characterize using information theory
  - clusters classified into behavior classes

- Interpret behavior classes
  - structural modeling for dominant activities
Structural modeling

- Each cluster has hundreds or thousands of flows.
  - an exhaustive approach is not practical
  - need a compact summary

- Dominant state analysis
  - dominant activities of the clusters

- An example: a web server from srcIP perspective
  - $RU_{\text{srcPort}} \leq RU_{\text{dstIP}} \leq RU_{\text{dstPort}}$
  - feature dependency: srcPort, dstIP, dstPort
Dominant state analysis

- Observations
  - clusters within the same BCs have similar structural models
  - they could have different dominant states (or activities)

<table>
<thead>
<tr>
<th>BCs</th>
<th>Structural models</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC2</td>
<td>srcPort(.)-&gt;dstPort(.)-&gt;dstIP(*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>srcPort(1025)-&gt;dstPort(137)-&gt;dstIP(*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>srcPort(1081)-&gt;dstPort(137)-&gt;dstIP(*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>srcPort(1153)-&gt;dstPort(1434)-&gt;dstIP(*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>srcPort(220)-&gt;dstPort(6129)-&gt;dstIP(*)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>scan activities</td>
<td></td>
</tr>
</tbody>
</table>
**Additional flow features**

- Flow, packet and byte counts
  - average counts of packets and bytes per flow
# Canonical behavior profiles

<table>
<thead>
<tr>
<th>Profile</th>
<th>Interpretation</th>
<th>BC</th>
<th>Freq.</th>
<th>Flow feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server/service</td>
<td>servers talk to a large number of clients</td>
<td>srcIP BC{6,7,8} dstIP BC{18,19}</td>
<td>frequently occurring</td>
<td>diverse packets and bytes</td>
</tr>
<tr>
<td>Heavy hitter</td>
<td>hosts talk to many or several IP addresses (typically servers)</td>
<td>srcIP BC{18,19} dstIP BC{6,7}</td>
<td>frequently occurring</td>
<td>diverse packets and bytes</td>
</tr>
<tr>
<td>Scan/exploit</td>
<td>hosts attempt to spread malicious exploits</td>
<td>srcIP BC{2,20}</td>
<td>highly volatile</td>
<td>single packet, same bytes</td>
</tr>
</tbody>
</table>
Case Studies

- Identify interesting events using typical profiles
  - server profiles on high ports, e.g., 60638
  - p2p traffic on alternative ports
  - exploit activities on unknown ports, e.g., an end host probing random dstIPs on dstPort 12827

- Rare behaviors
  - behavior patterns that rare happen are interesting
  - case study: exploit traffic from NAT boxes

- Deviant behaviors
  - clusters change from its usual BCs to a different
  - case study: a web server under DoS attack
Conclusions

- Develop a systematic methodology to automatically discover and interpret communication patterns
- Use information-theoretical techniques to build behavior models of end hosts and applications
- Apply dominant state analysis to explain traffic behavior
- Discover typical behavior profiles as well as rare and deviant behaviors
Future work

- Correlating behavior profiles across multiple links
- Validate behavior profiles using additional features, e.g., packet payload
- Integrate traffic profiling framework with a real-time monitoring system