Predicting Prefix Availability in the Internet

Ravish Khosla, Sonia Fahmy, Y. Charlie Hu, Jennifer Neville
Purdue University
INFOCOM 2010
Background

- Border Gateway Protocol (BGP)
  - Inter-domain policy based routing protocol
  - Advertises IP prefixes belonging to Autonomous Systems (ASes)
Motivation

- **Prefix Availability**: Time prefix is reachable Total Time Period

- Availability from various vantage points in Internet should be high, especially for popular websites/services
  - Continuous BGP advertised reachability is a key ingredient

- Measuring availability: non-trivial
  - Measurement infrastructure

- This work: Predictive approach of BGP (control-plane) availability
Predicting Future Availability

- Is future availability = past availability?
  - Can we observe prefix’s updates for some time and predict its availability?
- Fairly true if observation duration equal to prediction duration
- Often prediction desired for much longer duration than observation period
- **Contribution**: Build statistical prediction models to predict availability
  - Prefixes convey information about other “unrelated” prefixes
Methodology

- Datasets from RouteViews
  - Jan. 05, Jan. 07, Feb. 08 and Mar. 09
- Predict availability classes of a *combination*: (peer, prefix) tuple
- Classes: High/Low with 0.99999 threshold

### Availability Histogram for Mar. 09

<table>
<thead>
<tr>
<th>Availability Range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9–1.0</td>
<td>94.1 %</td>
</tr>
<tr>
<td>&gt; 0.99</td>
<td>94.63 %</td>
</tr>
<tr>
<td>&gt; 0.99999</td>
<td>68.75 %</td>
</tr>
</tbody>
</table>
Methodology (Contd.)

- Prefix attributes
  - Prefix length, Update Frequency, Mean Time to Failure (MTTF) and Mean Time to Recovery (MTTR)

- Applying prediction models
  - Learn using attributes and availability of combinations for training period
  - Apply on other combinations with attributes computed from training period e.g. 1 week of a month
  - Predict availability for test period e.g. remaining 3 weeks
  - Validate prediction results using known availability, computed from RouteViews
Methodology (Contd.)

- **Models studied**
  - Simple Model
    - Predict availability of combination as its past availability
  - Naïve Bayes
  - Decision trees with and without bagging

- **Prediction metrics**
  - Accuracy
  - Area under Receiver Operating Characteristic (ROC) Curve (AUC)
Prediction Results

- Bagged decision trees learned from one week (~25%) of the month

<table>
<thead>
<tr>
<th>Month</th>
<th>Accuracy (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 05</td>
<td>67.83</td>
<td>0.7005</td>
</tr>
<tr>
<td>Jan. 07</td>
<td>72.50</td>
<td>0.7094</td>
</tr>
<tr>
<td>Feb. 08</td>
<td>77.80</td>
<td>0.7483</td>
</tr>
<tr>
<td>Mar. 09</td>
<td>83.24</td>
<td>0.7605</td>
</tr>
</tbody>
</table>

- **Bagged decision trees** perform the best in terms of AUC and good accuracy
- Recent months are more predictable
• Bagged decision trees also perform best for all learning durations
Conclusions and Future Work

- Availability prediction
  - Future availability = Past availability works fairly well when training period = prediction period
  - For shorter learning periods, use statistical learning based prediction models
    - Bagged decision trees work the best
  - Prediction models can be built using random Internet prefixes
- Future Work: Study potential improvement in prediction accuracy using prefixes in the same AS or BGP Atom
Questions
Backup: Importance of attributes

- Studied effect on performance by considering various attribute subsets
- Results
  - Past availability used alone is a bad predictor of future availability
  - Prefix length and update frequency are weaker prediction attributes
  - MTTF and MTTR are the strongest attributes for prediction
Backup: Naïve Bayes prediction

- Assumption: Attributes are conditionally independent given the class label
- \( P(\text{Class Label} | \text{Attributes}) \) computed using Bayes rule
- Individual probabilities are learned using information from the training set
Backup: Decision Trees

- Example:

  - Bootstrap Aggregating (Bagging):
    - Take many bootstrap samples with replacement
    - Learn various trees from the samples
    - Apply all of them and take majority vote

MTTF > 2s

High
Low
Backup: All Prediction Results

Accuracy (%)

AUC

Simple Model
Naive Bayes
Bagged Decision Trees

Jan.05 | Jan.07 | Feb.08 | Mar.09

Jan.05 | Jan.07 | Feb.08 | Mar.09

3/15/2010