Predicting Prefix Availability in the Internet

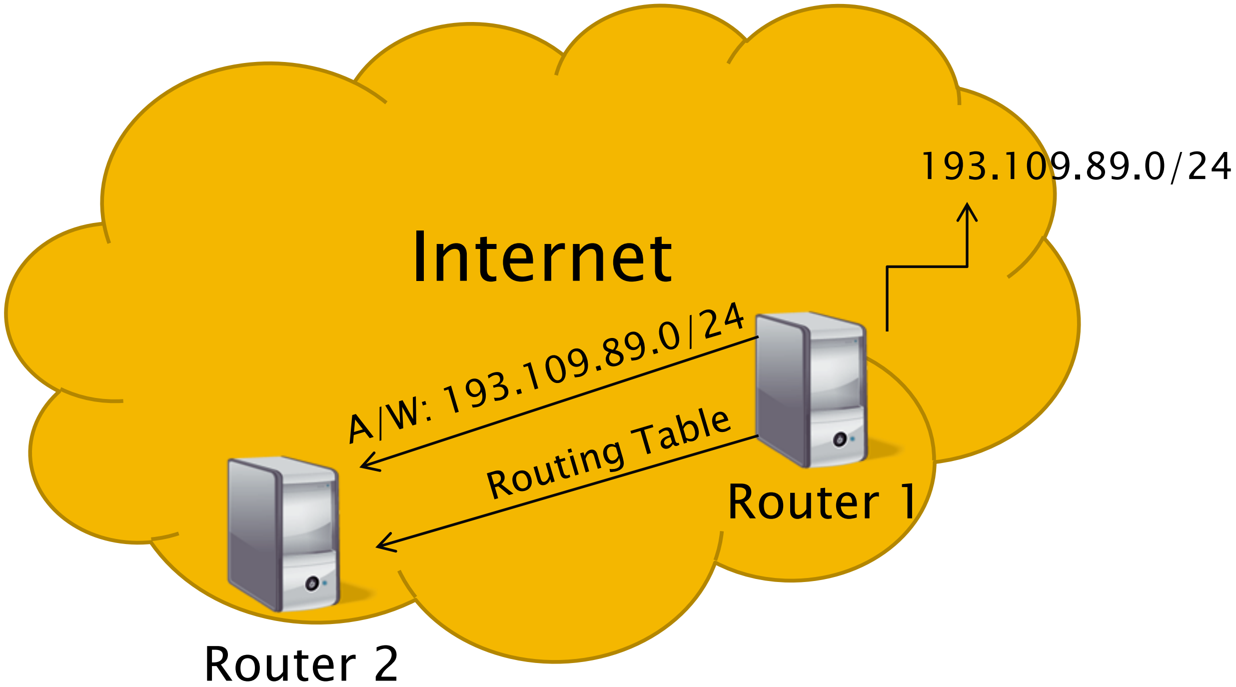
Ravish Khosla, Sonia Fahmy, Y. Charlie Hu, Jennifer Neville

Purdue University

INFOCOM 2010

Background

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



Motivation

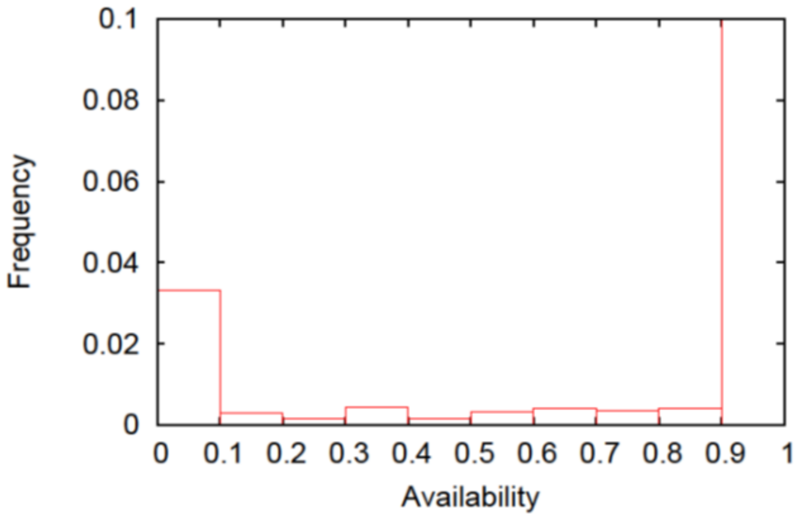
- Prefix Availability: $\frac{\text{Time prefix is reachable}}{\text{Total Time Period}}$
- Availability from various vantage points in Internet should be high, especially for popular websites/services
  - Continuous BGP advertised reachability is a key ingredient
- Measuring availability : non-trivial
  - 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

Availability Range	Frequency
0.9-1.0	94.1 %
> 0.99	94.63 %
> 0.99999	68.75 %

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)

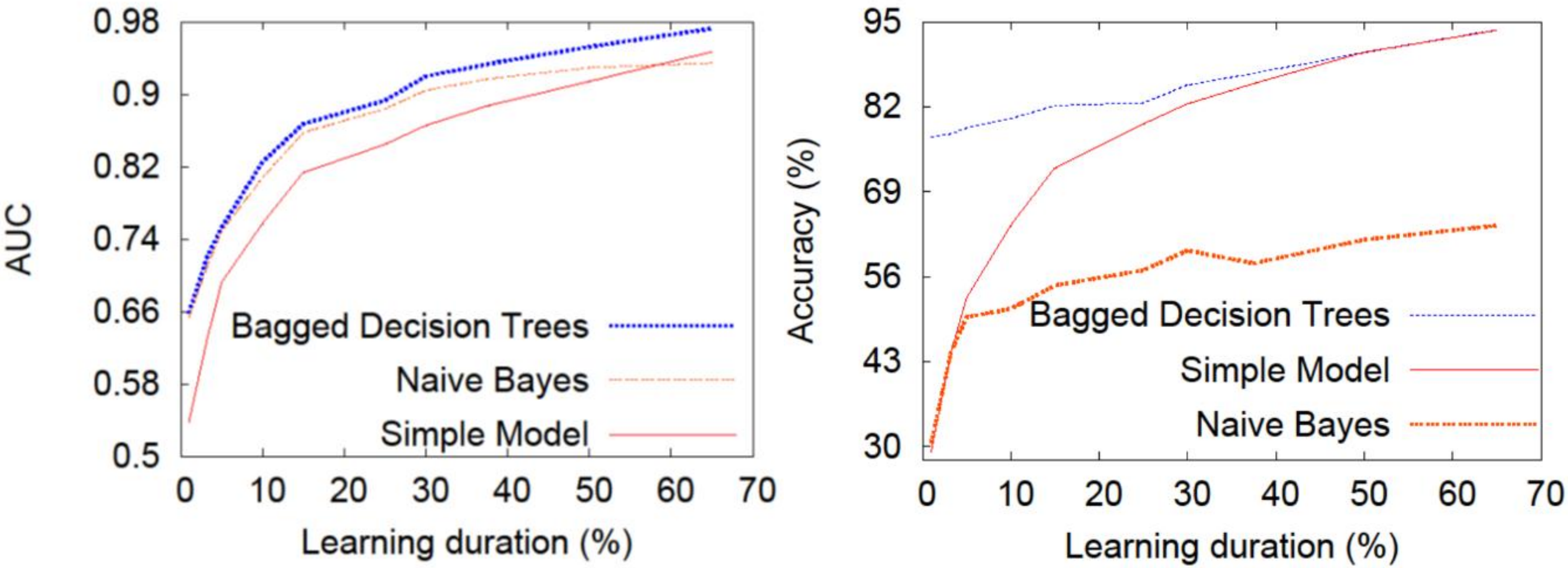
Prediction Results

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

Month	Accuracy (%)	AUC
Jan. 05	67.83	0.7005
Jan. 07	72.50	0.7094
Feb. 08	77.80	0.7483
Mar. 09	83.24	0.7605

- Bagged decision trees perform the best in terms of AUC and good accuracy
- Recent months are more predictable

Effect of Learning Duration



- 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

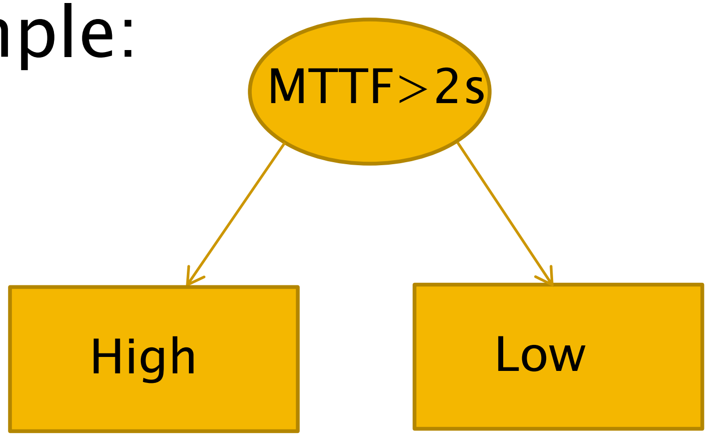
- Studied effect on performance by considering various attribute subsets
- Results
  - Past availability used alone is a bad predictor of future availability
  - Prefix length and update frequency are weaker prediction attributes
  - MTTF and MTTR are the strongest attributes for prediction

Backup: Naïve Bayes prediction

- Assumption: Attributes are conditionally independent given the class label
- P(Class Label|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

Backup: All Prediction Results