BGP-lens: Patterns and Anomalies in Internet Routing Updates

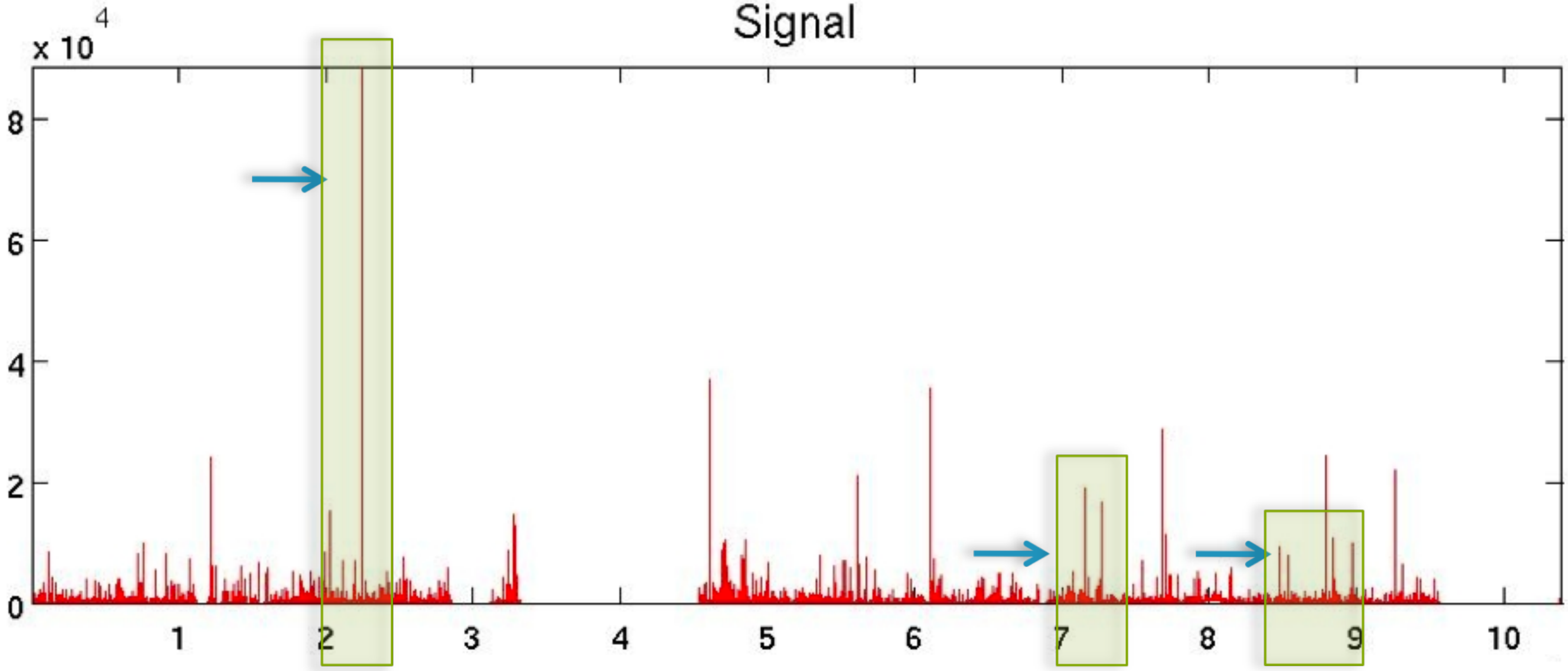
B. Aditya Prakash, Nicholas Valler, David Andersen, Michalis Faloutsos, Christos Faloutsos, SIGKDD'09

Presented by: Jian Wen

What's Happening in BGP?

- Routing information in a BGP network is updated frequently.
  - Why? Link/node failure, router maintenance, misconfigure.
- From these updates:
  - What is the normal pattern?
  - What does the anomalies look like (Route Flapping, Hijacking)?

Anomalies



Problem Definition

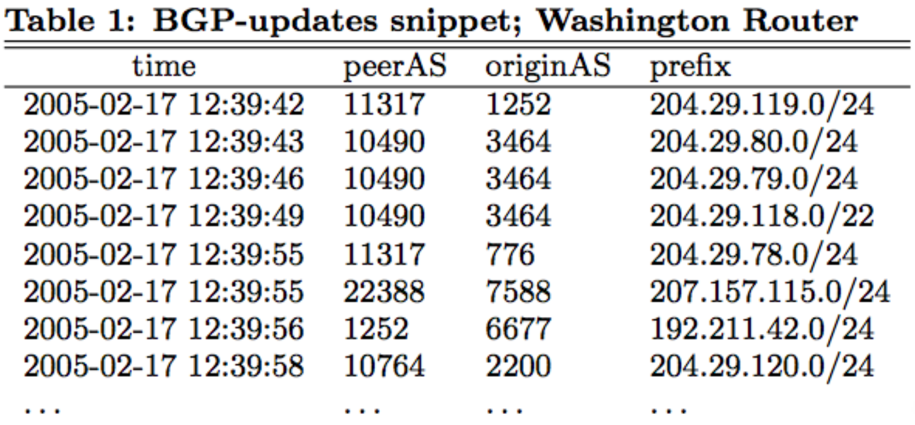
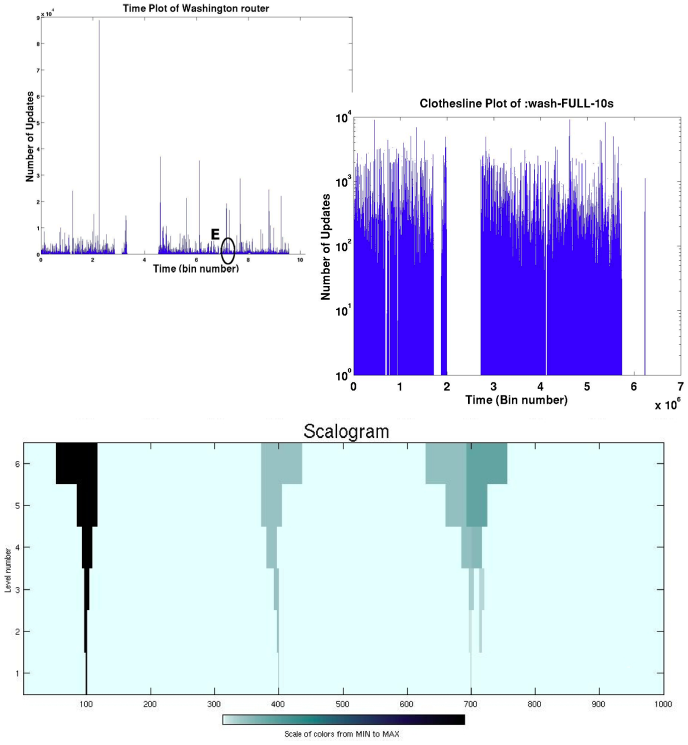


Table 1: BGP-updates snippet; Washington Router

time	peerAS	originAS	prefix
2005-02-17 12:39:42	11317	1252	204.29.119.0/24
2005-02-17 12:39:43	10490	3464	204.29.80.0/24
2005-02-17 12:39:46	10490	3464	204.29.79.0/24
2005-02-17 12:39:49	10490	3464	204.29.118.0/22
2005-02-17 12:39:55	11317	776	204.29.78.0/24
2005-02-17 12:39:55	22388	7588	207.157.115.0/24
2005-02-17 12:39:56	1252	6677	192.211.42.0/24
2005-02-17 12:39:58	10764	2200	204.29.120.0/24
. . .	. . .	. . .	. . .



- Given: BGP updates.
- Problem: Find patterns and anomalies.
- Out Approach: BGP-lens!

Existing Work/Solutions

- Network: BGP measurement and analysis
  - Canonical measurement and models for BGP anomalies and instability behaviors. Not really handy.
  - Detect network-wide BGP anomalies. Not for fine granularity.
  - Visualization and statistic methods. Data Mining?

BGP-lens

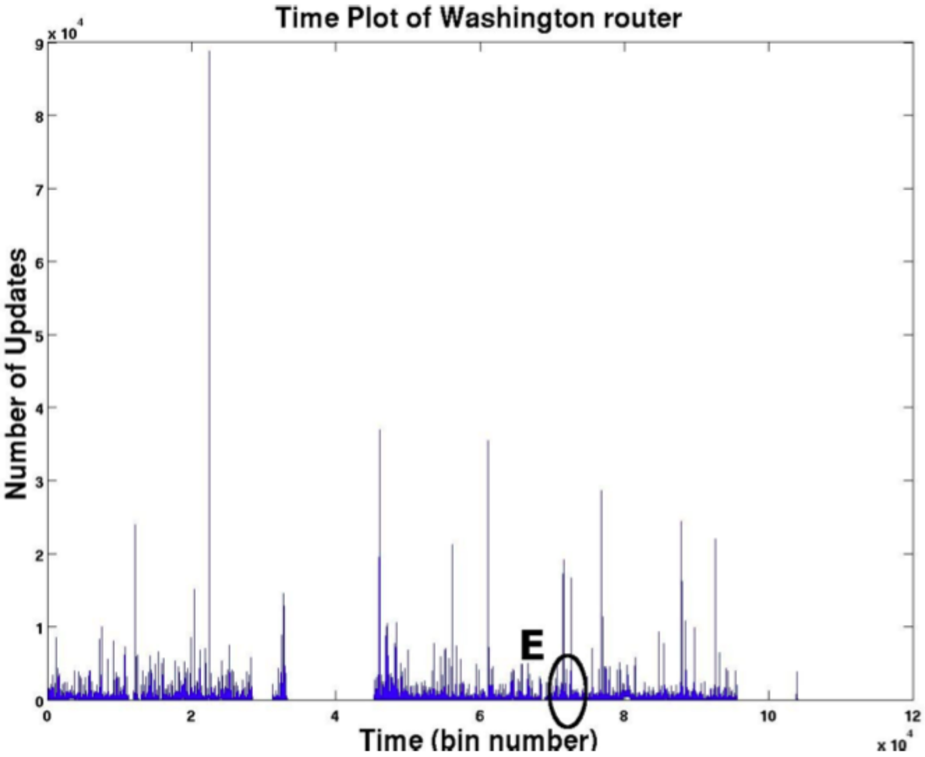
- A novel tool for automatically detecting patterns and anomalies in BGP updates at many different scales of observation.
  - Effective: Can detect both temporal and frequency anomalies.
  - Scalable: The algorithms are linear on the number of time-ticks and thus it can handle large datasets.
  - Admin-friendly: It can work with zero user input; automotive detection.

Roadmap

- Tool Components and Observations in BGP-lens
    - The Clothesline Effect - Temporal Analysis
    - The Tornado Plots - Frequency Analysis
- Automating Discovery
- Scalability
- User-interface: BGP-lens as an administrative tool
- BGP-lens at work

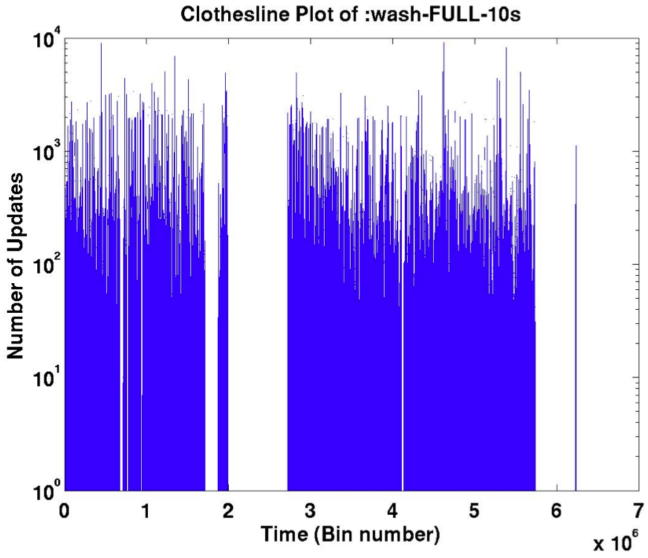
Temporal Analysis: Clothesline

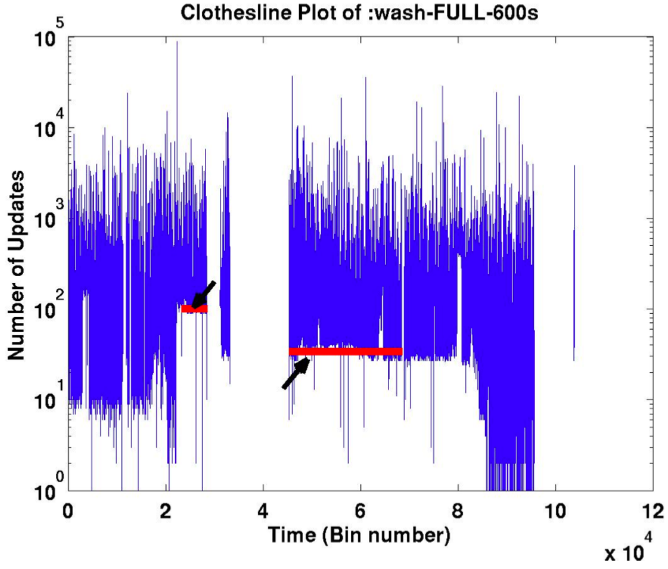
- Linear-linear plots fail to show short duration spurts.
  - Threshold method cannot deal with the huge variations.
  - FFT cannot work here due to the burstiness of the updates.



Temporal Analysis: Clothesline

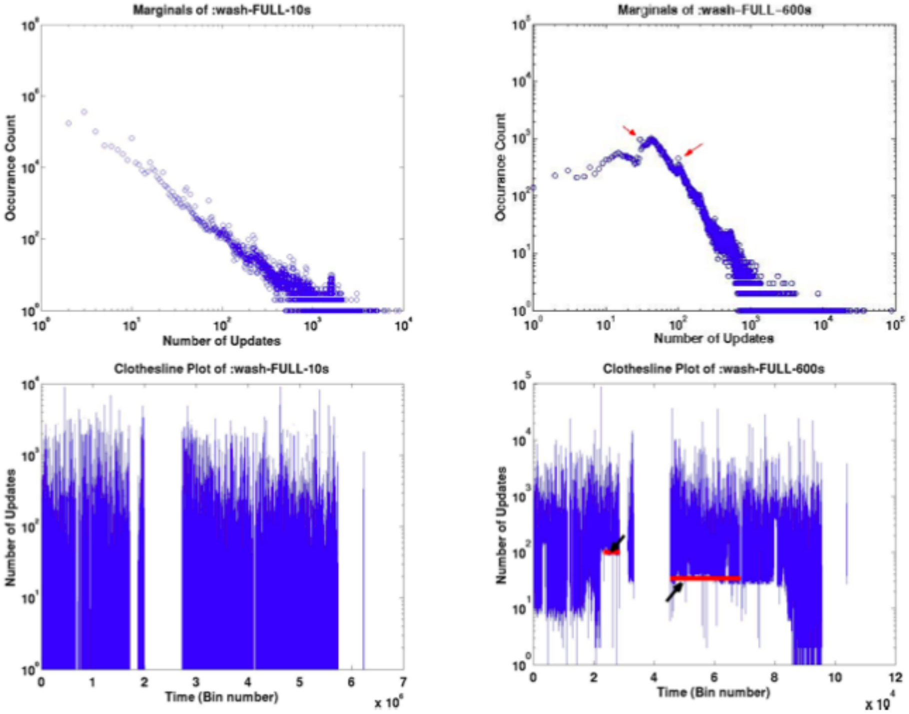
- Instead of using linear-linear plots, we use log-linear plots.
  - No striking outliers any more;
  - The “bin size”, or the window size for the measurement, now means a lot: clothesline!
  - Clothesline: a periodic update stream over a prolonged time period (so it may be Route Flapping).


Clothesline Plot of :wash-FULL-10s


Clothesline Plot of :wash-FULL-600s

Catch the Clothesline: Marginals

- Outliers in the "marginal" distribution usually correspond to clotheslines.
- Marginal distribution plot
  - Log-log scale;
  - PDF of Occurrence count on Number of updates

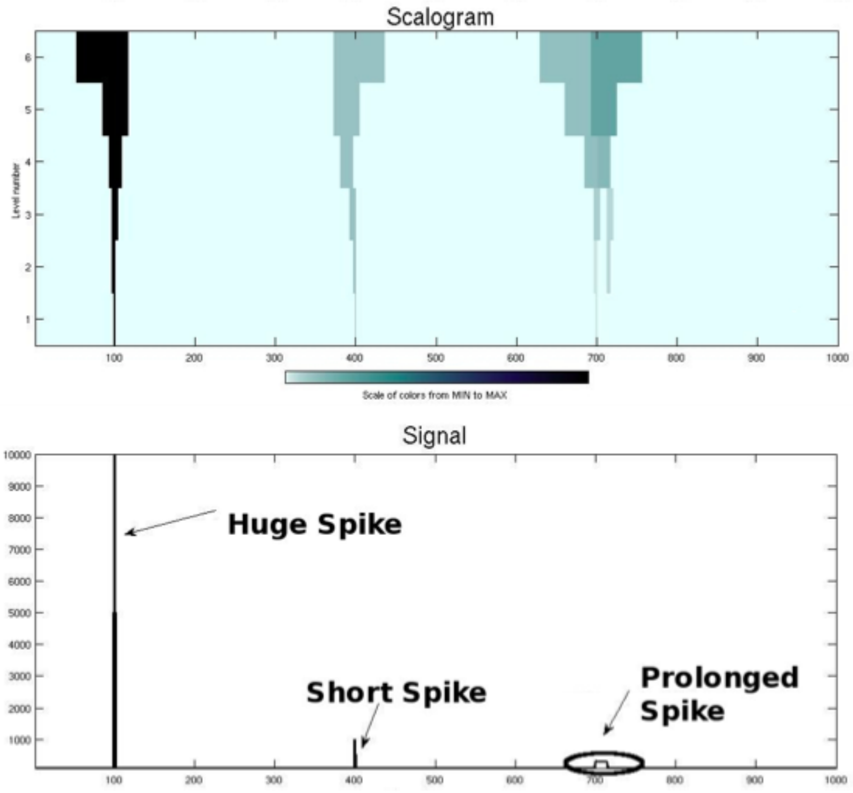


(a) Bin Size 10s: Marginal Plot (top) and Clotheslines Plot

(b) Bin Size 600s: Marginal Plot (top) and Clotheslines Plot

Frequency Analysis: Tornado

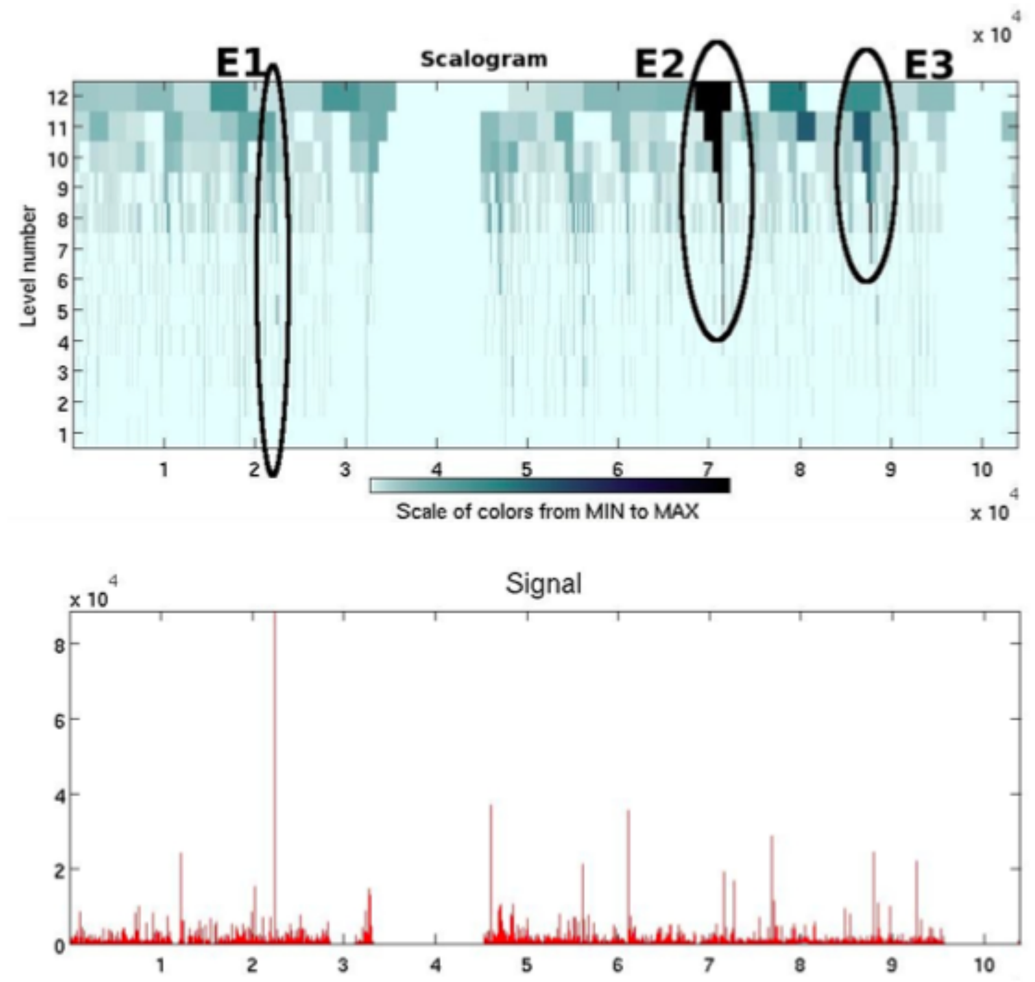
- Due to the self-similar nature of the data, Fourier Transformation doesn't work well for our purpose.
- Discrete Wavelet Transform and scalogram.
- Observations.
  - Pronounced spikes correspond to "tornadoes" that touch down.
  - Darker tornado => Larger spike.
  - Non-touch-down tornado => Prolonged spike.



(a) Synthetic: scalogram (top) and time-plot

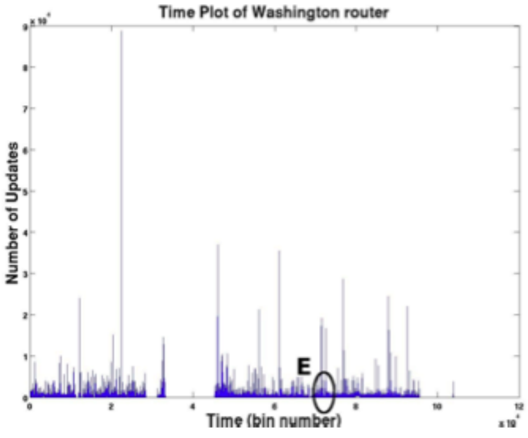
Real "Tornados"

- E1: A huge touch-down spike (one hour' prefix hijacking).
- E2: A dark non-touch-down spike (eight hours' sustained update activities).

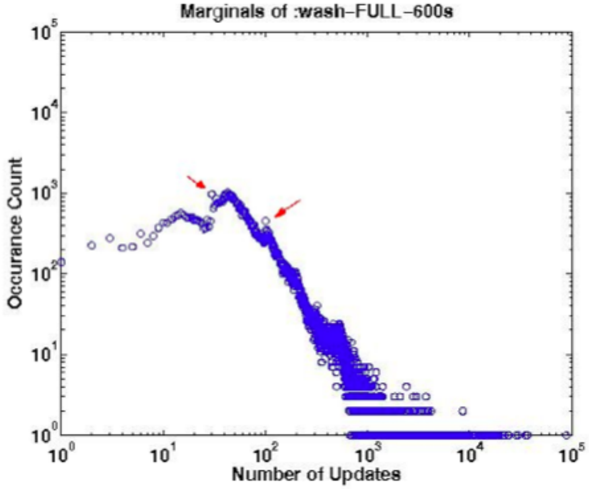


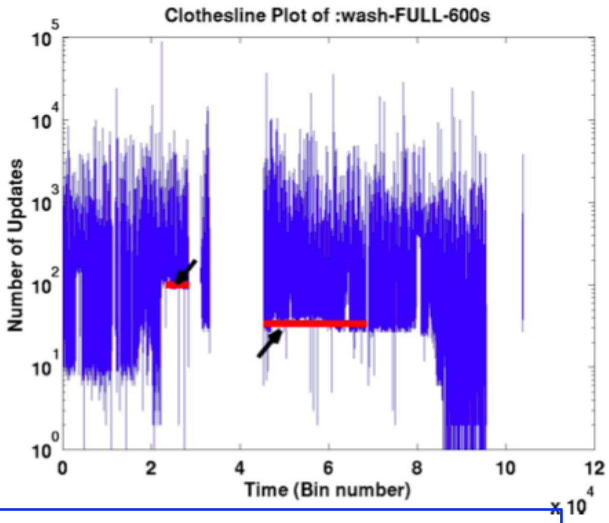
(b) Real: scalogram (top) and time-plot

Automating the Discovery Clotheslines

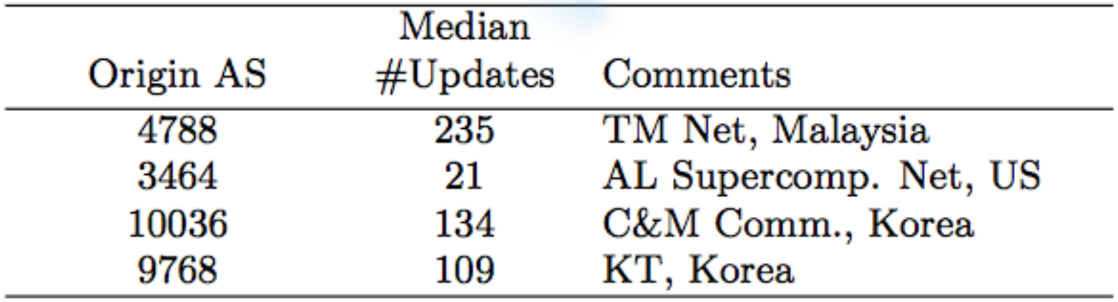


(a) Time-plot





Get marginal plot, find outliers. → Find longest time interval for outliers.



Origin AS	Median #Updates	Comments
4788	235	TM Net, Malaysia
3464	21	AL Supercomp. Net, US
10036	134	C&M Comm., Korea
9768	109	KT, Korea

Automating the Discovery Clotheslines

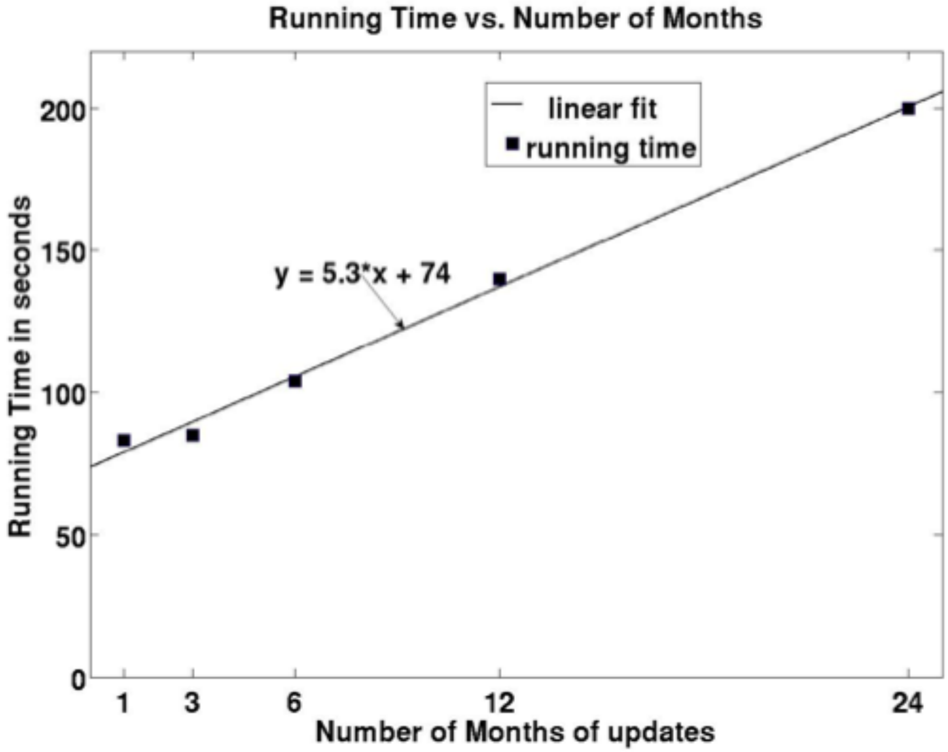
- For each time bin size b=2^i, derive the corresponding marginal plots.
  - Multiple plots corresponding to different *i* value.
- For each marginal plot use the median filtering approach to determine "outliers".
  - Median Filter Approach: reduce the noise and pick the median for output.
- For each outliers found, find the longest time-interval from the corresponding clothesline plot.
- For each time interval found, report the most consistent IPs or ASes etc.

Automating the Discovery Prolonged Spike (Tornadoes)

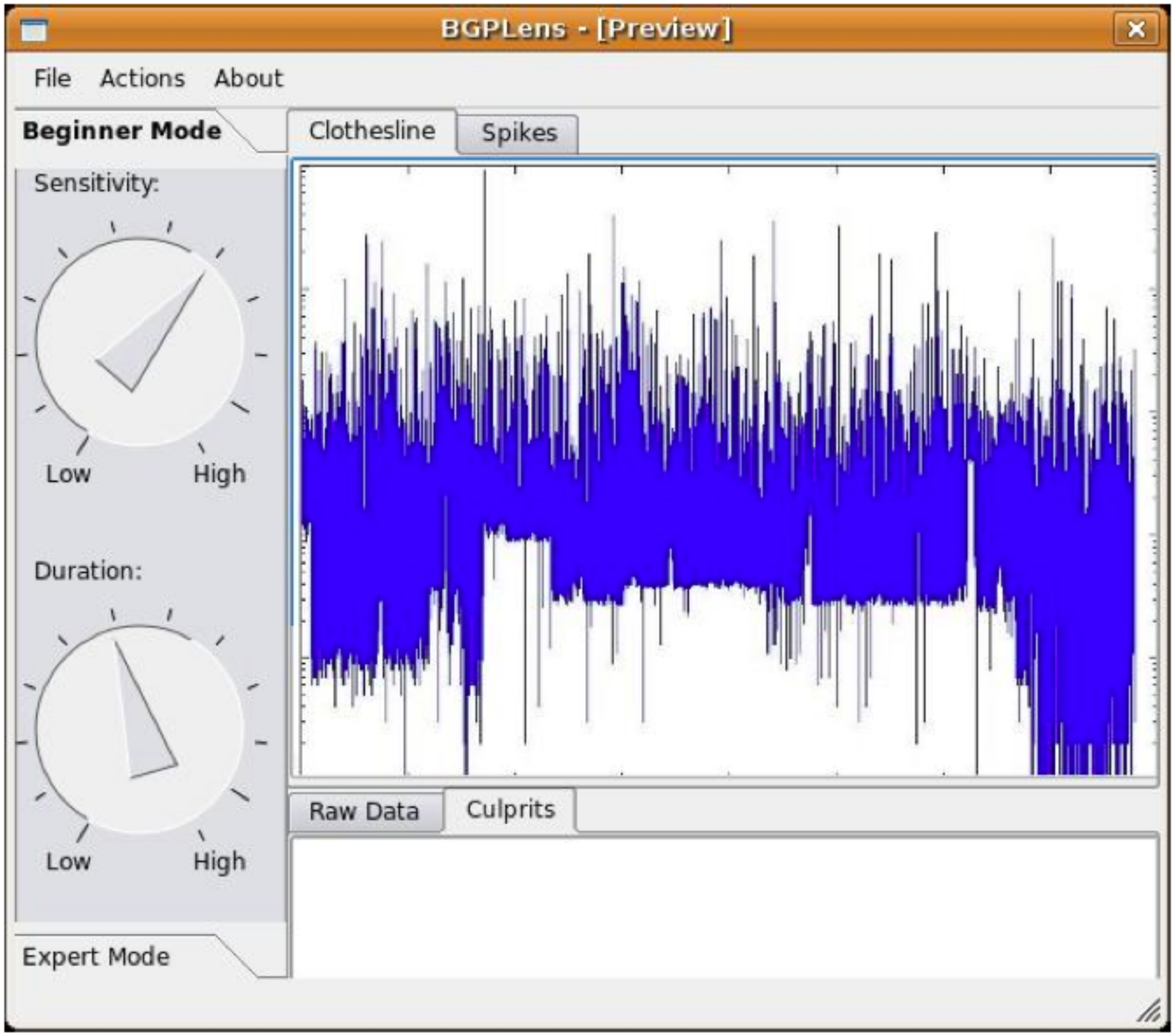
- Require two inputs: sensitivity and duration
  - Sensitivity: the percentage of the DWT coefficients to be considered, which refers to the strength of the spike (recall: larger coefficient -> darker scale cell -> larger spike).
  - Duration: the time threshold for the spike's duration.
- BGP-lens provides the default input of these two parameters.
  - Only consider wavelet coefficients within 60% of the maximum with duration at least $2^{len-8+1}$

Scalability of BGP-lens

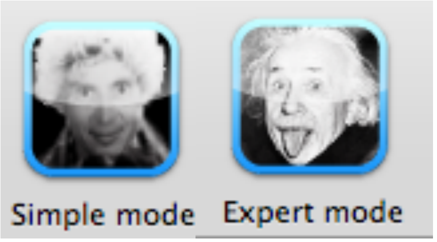
- Top-5 anomalies.
- Two AMD Opteron dual-core 2.4GHz, 48G Mem, Fedora 5
- Data size: > 18 million updates for two years.



User Interface



- Install and run! No more configuration!
- Beginner/ Expert Mode



BGP-lens on Duty: Clotheslines

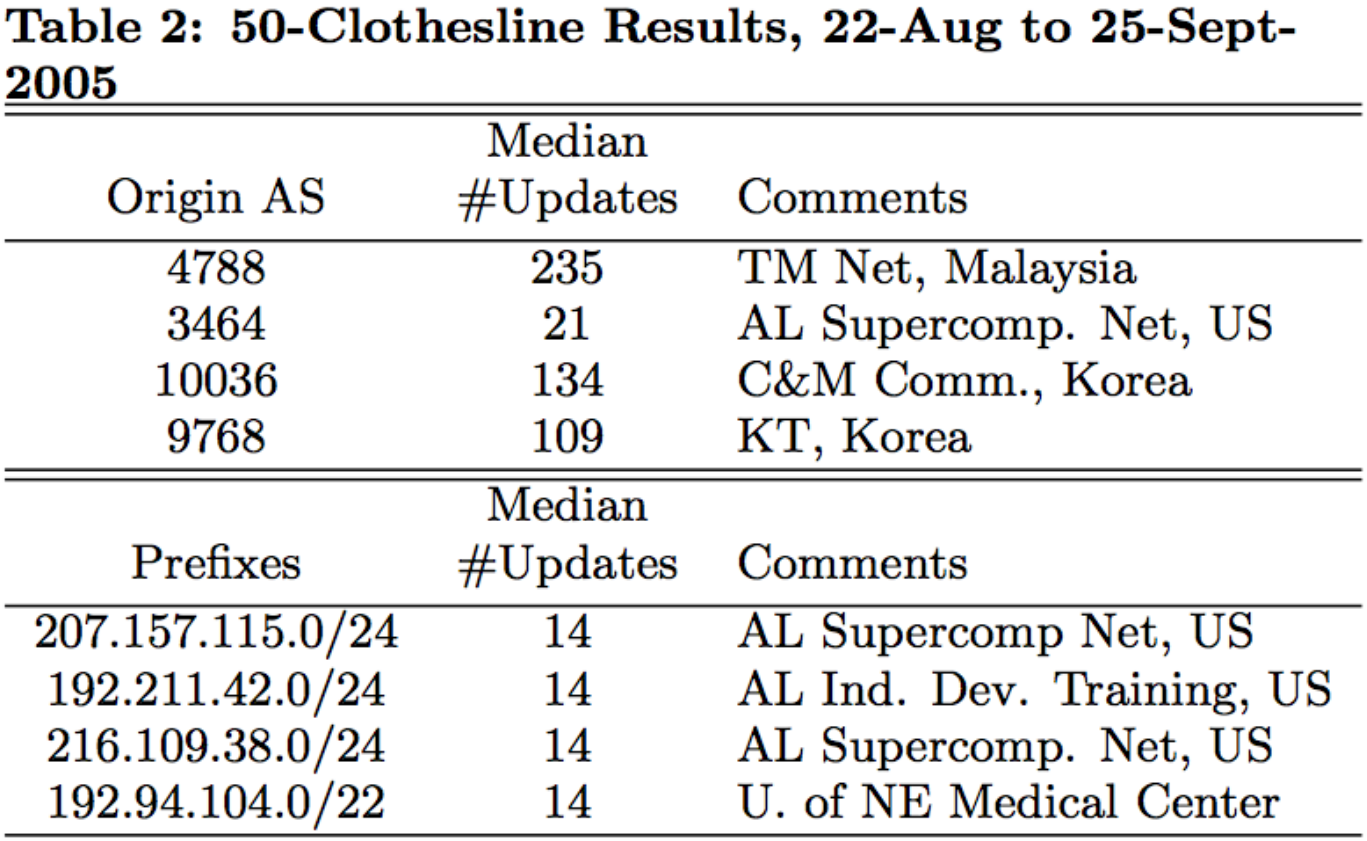


Table 2: 50-Clothesline Results, 22-Aug to 25-Sept-2005

Origin AS	Median #Updates	Comments
4788	235	TM Net, Malaysia
3464	21	AL Supercomp. Net, US
10036	134	C&M Comm., Korea
9768	109	KT, Korea

Prefixes	Median #Updates	Comments
207.157.115.0/24	14	AL Supercomp Net, US
192.211.42.0/24	14	AL Ind. Dev. Training, US
216.109.38.0/24	14	AL Supercomp. Net, US
192.94.104.0/22	14	U. of NE Medical Center

BGP-lens on Duty: Prolonged Spikes

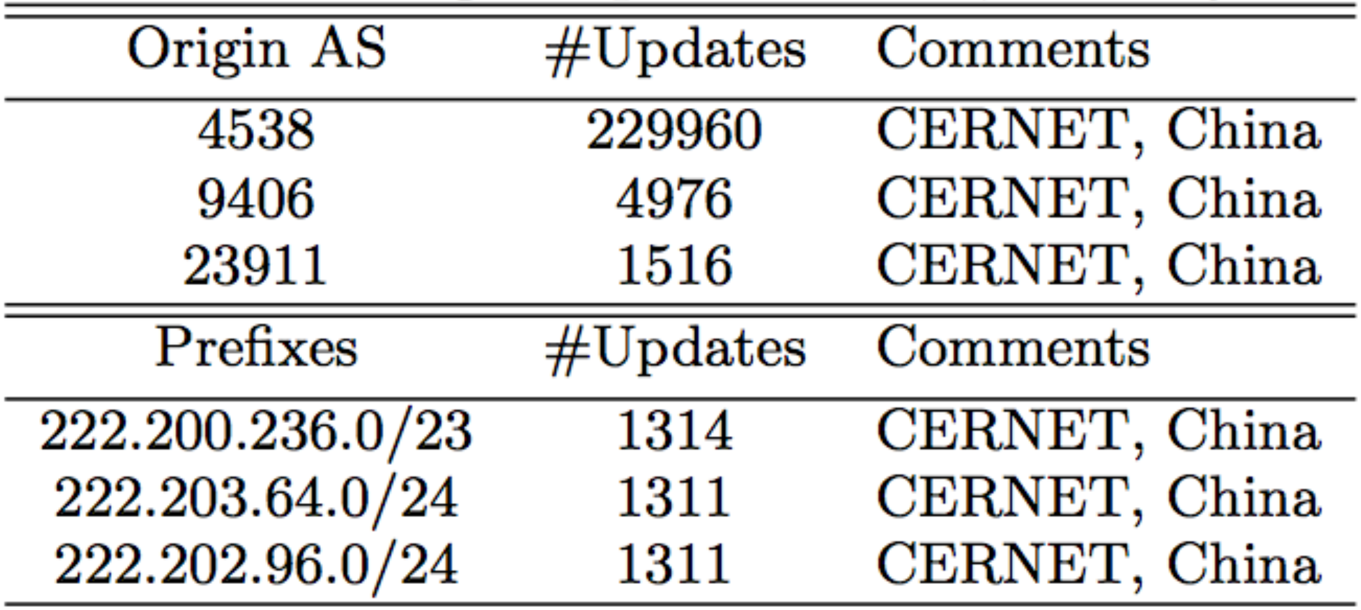


Table 3: Prolonged Spike Results, 12-May-2005

Origin AS	#Updates	Comments
4538	229960	CERNET, China
9406	4976	CERNET, China
23911	1516	CERNET, China

Prefixes	#Updates	Comments
222.200.236.0/23	1314	CERNET, China
222.203.64.0/24	1311	CERNET, China
222.202.96.0/24	1311	CERNET, China

Summary

- BGP-lens: handy tools for administrators to monitor BGP updates.
  - Efficient, scalable, and admin-friendly.
  - Support anomalies detection on both updates bursts and prolonged spikes.
- The paper also covers some interesting observations:
  - Marginals that are mixture of log-normals with a power-law tail.
  - Self-similarity of BGP updates data corresponding to a 75-25 b-model slope.

Future Work

- On-line Monitoring Tool?
  - Incremental algorithms.
  - Arbitrary time instance and duration.