
THOMSON
images & beyond

Detectability of Traffic Anomalies in Two Adjacent Networks

Augustin Soule,
Haakon Ringberg,
Fernando Silveira,
Jennifer Rexford,
Christophe Diot

Anomaly detection in large networks

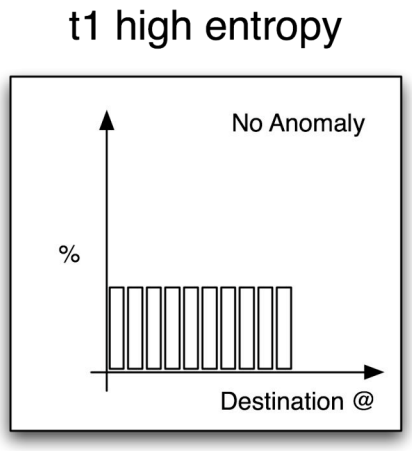
- Anomaly detection is complex for large network

- Network-wide analysis [Lakhina 04] is promising
- Validated against multiple networks at different time
  - Abilene 03, Geant 04, Sprint Europe 03
- Features impacting the anomaly detection are unknown yet

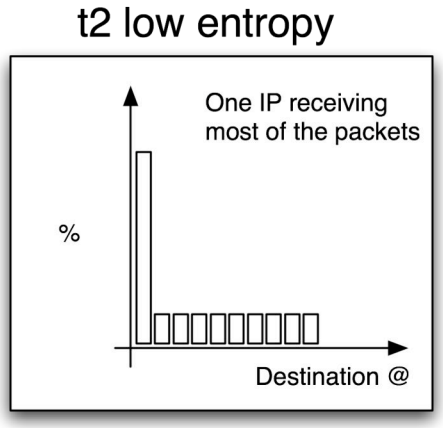
Compare the anomaly observed between two networks

Using entropy for anomaly detection

Normal

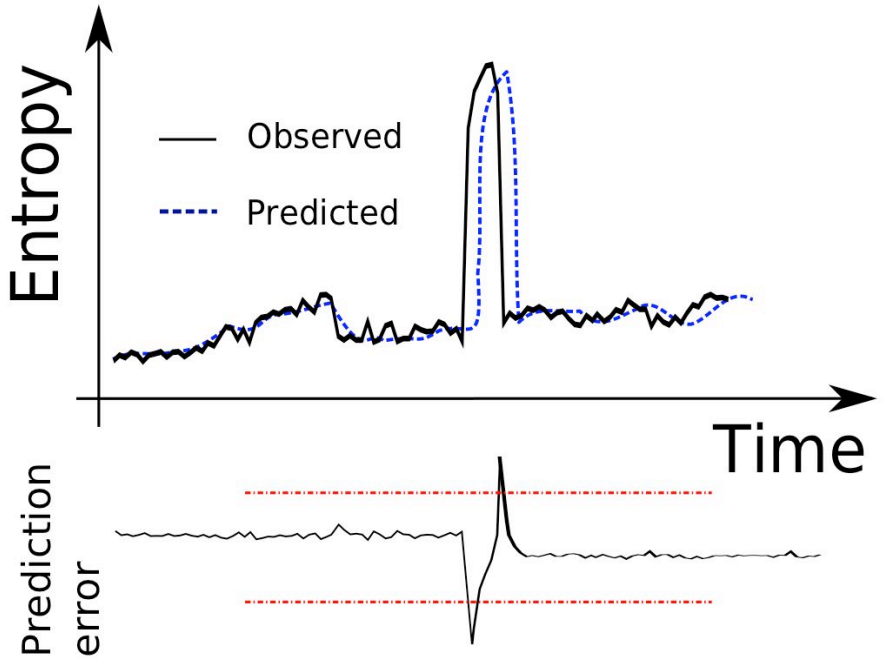


During a DOS attack



- Hypothesis : the distribution changes during an anomaly
- Entropy is a measure of the dispersion of the distribution
  1. Minimum if the distribution is concentrated
  2. Maximum if the distribution is spread
- Four features
  1. Source IP distribution
  2. Destination IP distribution
  3. Source Port distribution
  4. Destination port distribution

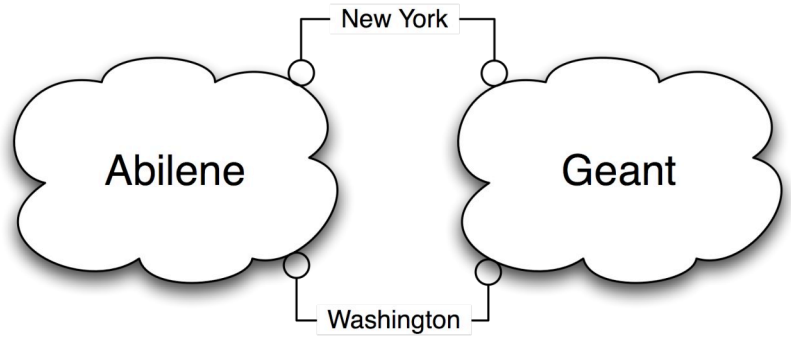
Detecting anomalies



- Kalman filter method [Soule 05]
- Method Overview
  1. Use a model to predict the traffic
  2. Innovation = Prediction error
- High threshold avoid false positive

THOMSON
images & beyond

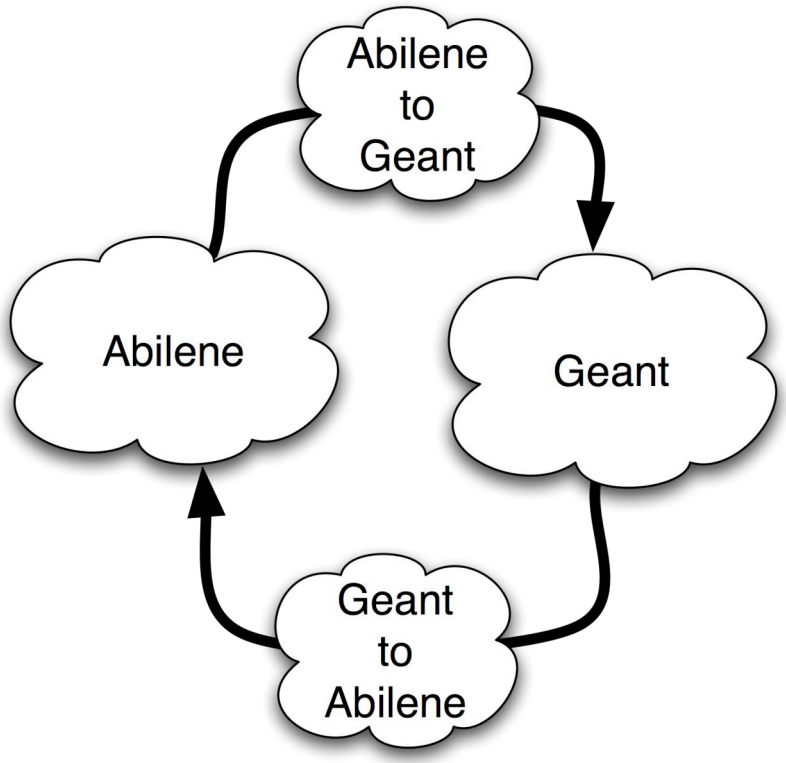
Collected dataset



- Abilene and Geant monitoring
- Collected three month of data
  1. BGP
  2. IS-IS
  3. NetFlow
- Isolate twenty consecutive days of complete measurement
- Connected through two peering links

	Sampling	Temporal aggregation	Anonymization
Abilene	1/100	5 min	11 bits
Geant	1/1000	15 min	0 bits

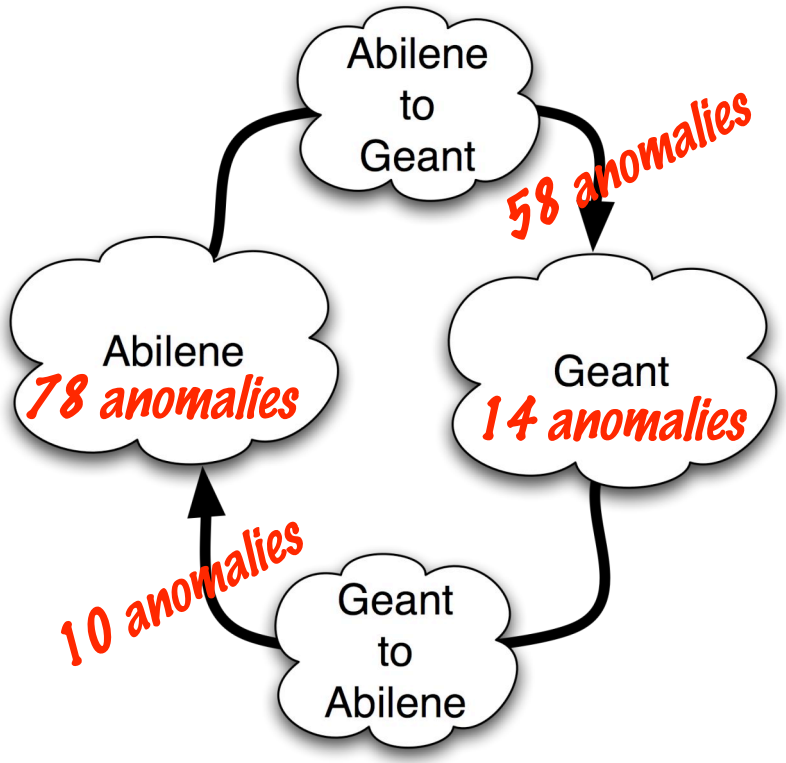
Abilene and Geant



- Use routing information to isolate
  1. Traffic from Abilene to Geant
  2. Traffic from Geant to Abilene

- Detect anomalies inside each dataset using the same threshold parameter, but different data-reduction parametes

Anomalies detected



- Compare the anomalies sent versus the anomalies observed
  1. Expected for G2A and A
  2. Surprising for G and A2G

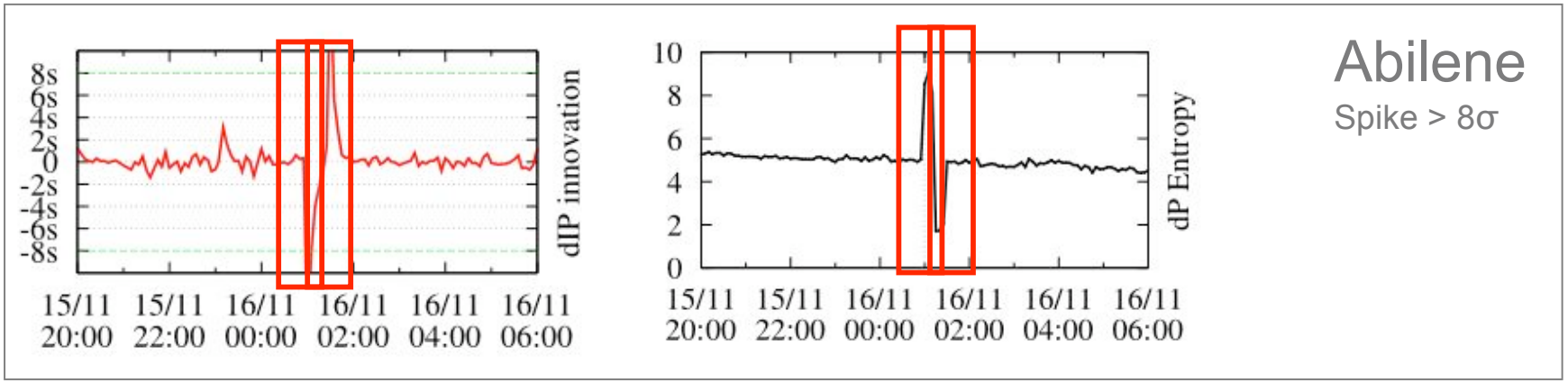
- Amount of traffic ?
- Sampling ?
- Anonymization ?
- Threshold ?
- Method ?
- Model ?

Undetected anomalies

- Examples of anomalies detected in a network but undetected in the other.

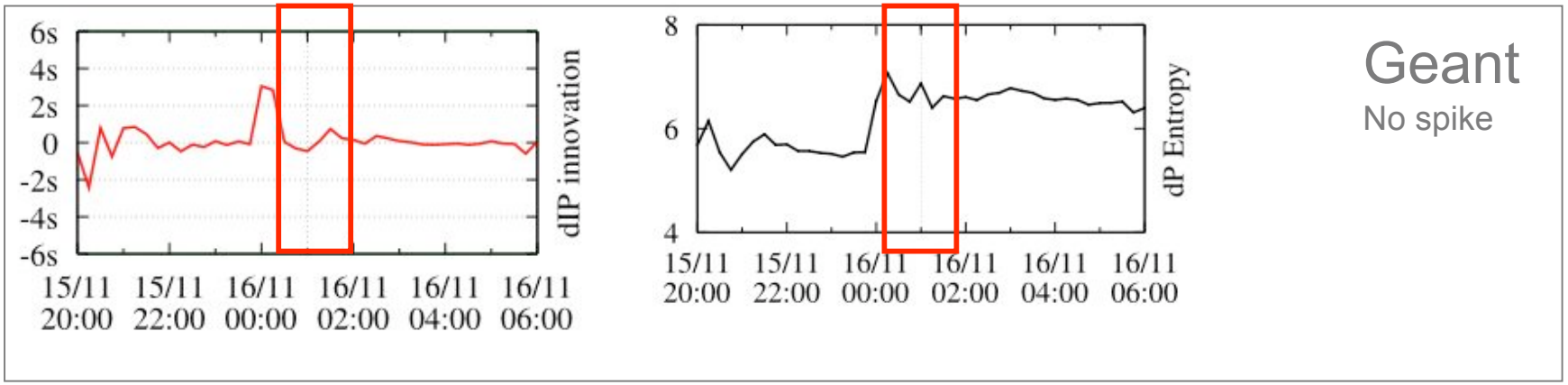
1. Impact of Sampling & Method
2. Impact of customer's Traffic Mix
3. Impact of anonymization

Example 1 : attack over Port 22



Abilene

Spike > 8σ



Geant

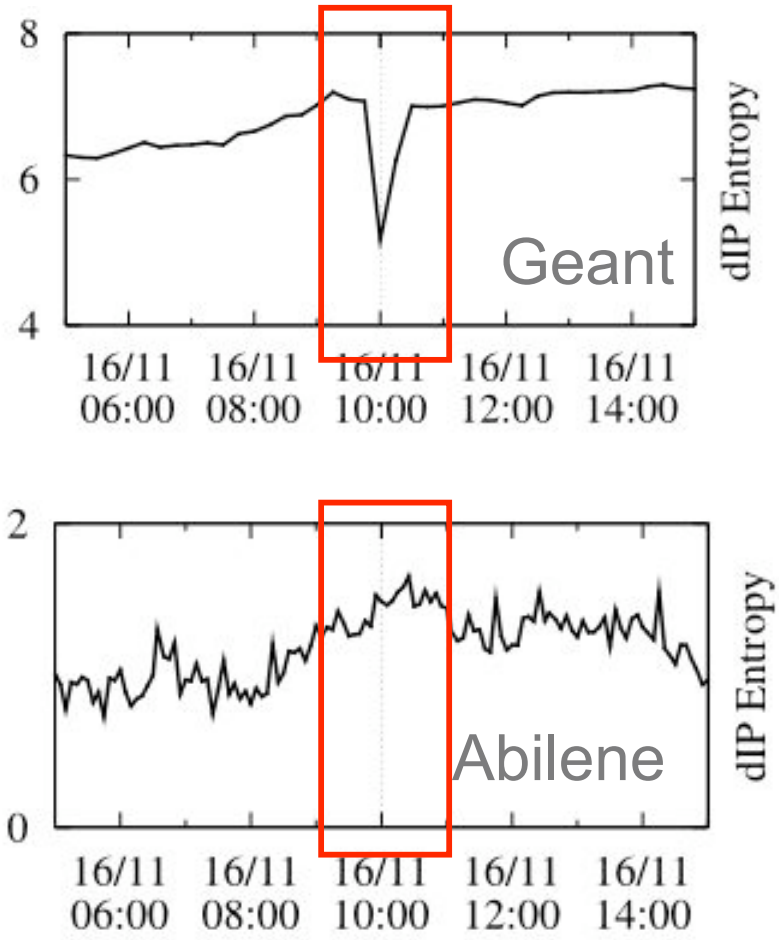
No spike

Sampling affects the perception of anomaly

The effect depends on the type of anomaly

Example 2 : Alpha Flow

Destination IP entropy



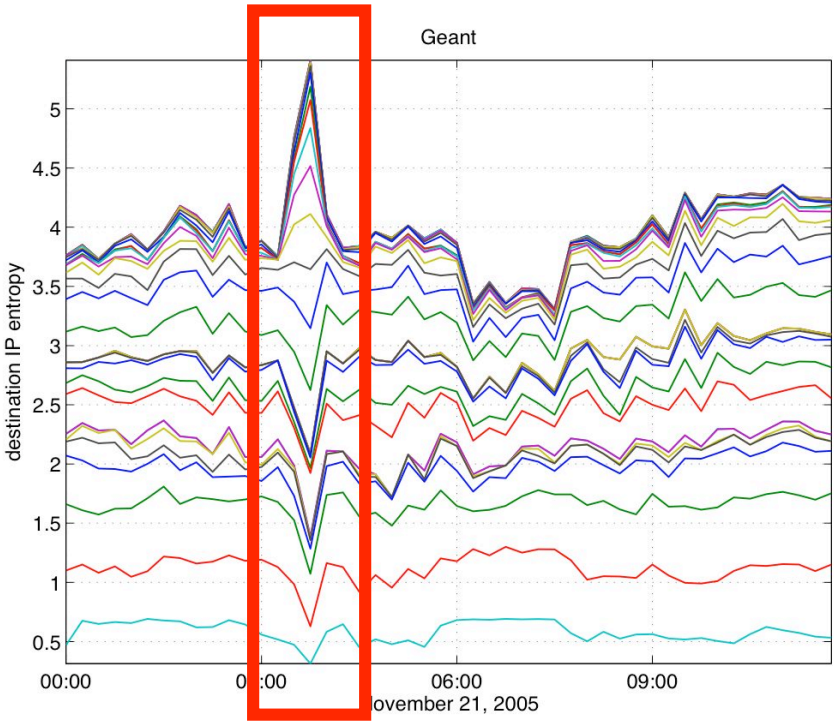
- Large file transfer between two hosts
- Observed in Geant
- Undetectable in Abilene

- In this Abilene the traffic is already concentrated by Web traffic

- The anomaly detectability is impacted by traffic

THOMSON
images & beyond

Example 3 : Scan over an IP subnet



- Attacker doing a subnet scan
- One source host
- Multiple destination hosts
  1. Concentration of source IP
  2. Dispersion of destination IP

- But we observe concentration in the Destination IP entropy

- Anonymization can :
  1. Help to detect anomalies
  2. Impact the anomaly identification

Summary

- First synchronized observation of two networks for anomaly detection
- Identification of various features impacting anomaly detection
  1. Sampling
  2. Traffic mix
  3. Anonymization

- Two anomalies are impacted differently by each features

- What impacts detectability ?

Thanks for listening !