If It Spreads You Can Catch It: Dynamic Models for Early Epidemic Detection

Eli Meirom, Constantine Caramanis, Shie Mannor, Ariel Orda, Sanjay Shakkottai
Motivation

- Malwares / Cyber-attacks are major security threats
- Early detection
- Detection without identification
  - “Break the cycle”
- *Indications* rather than *signatures*
  - Hard-drive spin up
  - Increased network usage
  - High CPU load
  - ...
The Greater Scope

Malware

Rumor Spreading \ Product adaption

Epidemiology
The Problem

- Provide optimal, continuous, online early identification tool:
  - Normal behavior
  - Contagion / Malware Infiltration!

- No available signature – only weak indications.
  - Extremely noisy setup
  - Single node inference is impossible

- Local network information
An Epidemic – SI dynamics

- SI Model / First Passage Percolation
  - Each edge associated with an exponential clock
Static Framework
Epidemic

Random Event
Intuition
The HotSpot Algorithm
The algorithm converge correctly even in settings where the reporting probability tends to zero and the noise level $\rightarrow \infty$.

The algorithm and proofs are diffusion-model-agnostic – they assume no knowledge of the diffusion mechanics.

- **Interior nodes** – infected nodes that have all their $K$ nearest neighbors infected.

$$\gamma(K) = \frac{\text{#interior nodes}}{\text{#infected nodes}}$$
Theorems

Large infections: $\Theta(N)$ reporting nodes

**Theorem:** Assume there exist $(K, \gamma_0 \neq 0)$ such that:

- $\Pr(\gamma(K) \leq \gamma_0) \to 0$
- $K \geq \log(\frac{1}{\gamma_0^{-1}(f + 1)/\log(1 + q)})$

Then, the hotspot algorithm solves the decision problem correctly, and the false positive and false negative probabilities are $o(e^{-cN})$. 
Theorems

- Small infections: $N^{1-\beta}$ reporting nodes

**Theorem:** Assume the reporting probability of truly infected nodes is $N^{-\mu}$. Set $K = [1/\beta]$ and $T = 1$. Assume that there exist $\gamma_0 \neq 0$ such that:
  
  - $\Pr(\gamma(K) \leq \gamma_0) \to 0$
  - $K \mu < 1 - \beta$

Then, the hotspot algorithm solves the decision problem correctly, and the false positive and false negative probabilities are $o(e^{-cN})$.

In particular, if $1 > \beta > 0.5$, the algorithm simply counts the expected number of infected neighboring pairs.
Experiments

The mean number of hotspots as a function of $K$. The plot was generated for an Erdos-Renyi graph.
Experiments

The mean error rate as a function of the number of epidemic sources for Enron email network.
What about the dynamics? (a teaser)
Normal Activity

- Each node generates flags according to a Poisson process of rate 1.
An Epidemic Session

- Upon infection, a *single* additional flag is raised.
  - *Indistinguishable* from normal activity flags
Can you guess?

By the time of successful / unsuccessful detection, the epidemic had already infected 50% of the network...
Conclusions

- Early detection tool for diffusive processes – the HotSpot Algorithm.
- Provided both theoretical guarantees and simulations indicating both *correct classification* and *early detection*.
- Information spreading analysis
  - First passage percolation
- Devised algorithms for dynamic setup:
  - FlagCounter, for “balanced networks”
  - Hub-FlagCounter, for heavy-tailed networks