Modeling Interventions in Complex Networks

Steven Hofmeyr (LBNL), Stephanie Forrest (UNM)

U.S. Department of Energy

Introduction

- Problems
  - Malware spread on complex networks difficult to predict
  - Effect of proposed security policies is unclear
- Approach
  - Model malware spread and intervention policies
  - Mathematical analysis and simulations

Modeling Platform: ASIM

ASIM is an agent-based model that mimics the growth and topology of the internet at the autonomous systems level using the following information:

- Geography
- Economics
- Traffic

By combining ASIM with models of malware spread, we end up with a reasonable approximation of real world malware distributions on the internet.

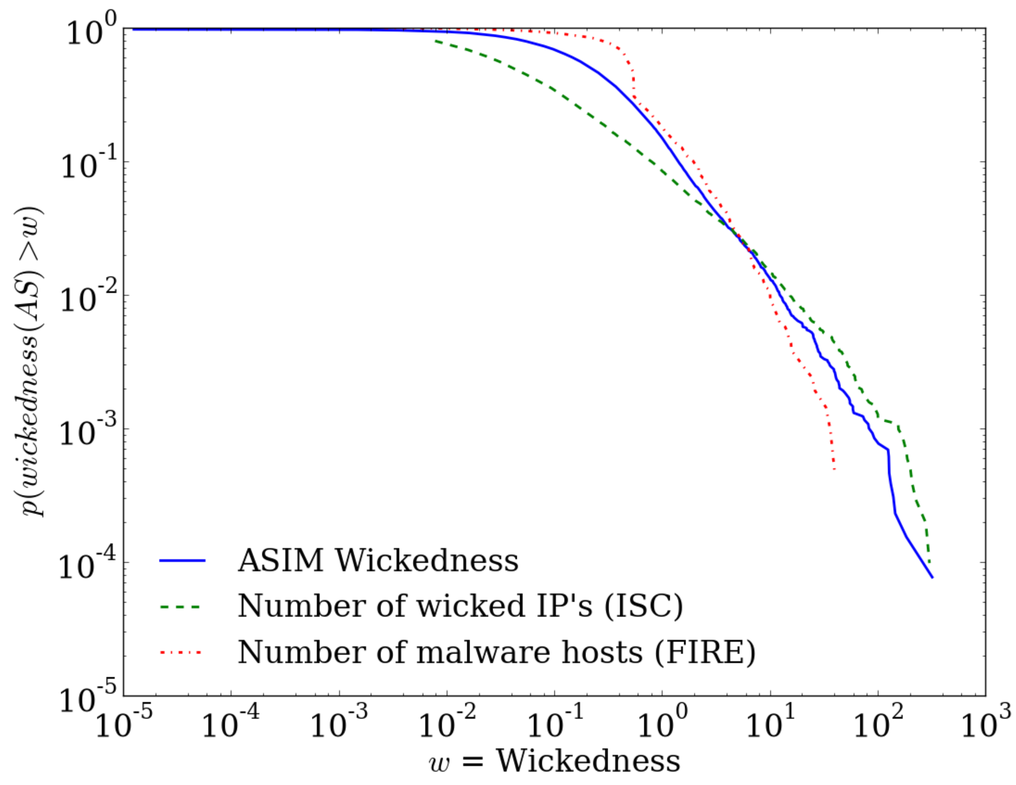


Figure: Comparison of malware distributions over the AS graph from the ASIM simulation, FIRE data, and the ISC data

Graduated Response and Drive by Downloads

To model the effect that search providers can have on infection rates, we created an epidemiological model of drive by downloads and search providers.

ρ_s	Probability of an uninfected website becoming infected.
γ_s	Probability of an infected website recovering.
ω_i	Popularity of a website i.
ρ_c	Probability of an uninfected client becoming infected upon visiting an infected site
γ_c	Probability of a infected client recovering

Figure: Model of website and client infections.

- Intervention policies
  - Blacklisting (removal from search results)
  - Graduated Response (reduction of search rank)

Data Sources

- BGP Dumps (AS topology over time)
  - Routeviews `http://www.routeviews.org/`
  - RIPE `http://ripe.net`
- Malware Data
  - SANS ISC (Internet Storm Center) `http://isc.sans.org`
  - FIRE `http://maliciousnetworks.org`

Results: ASIM

Filtering ingress traffic (including transit) is more than twice as effective as filtering egress traffic alone

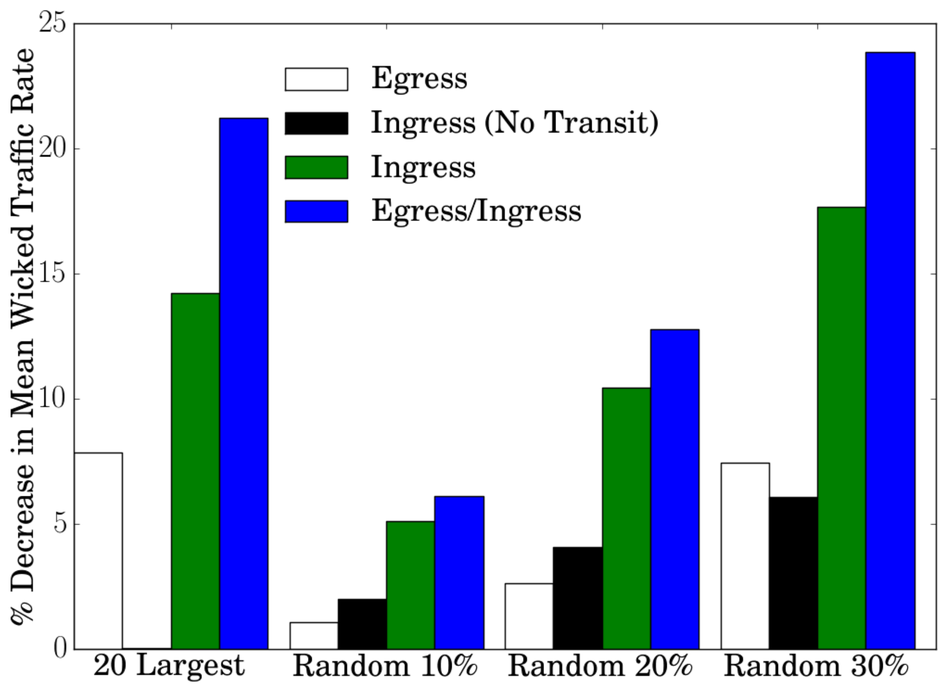


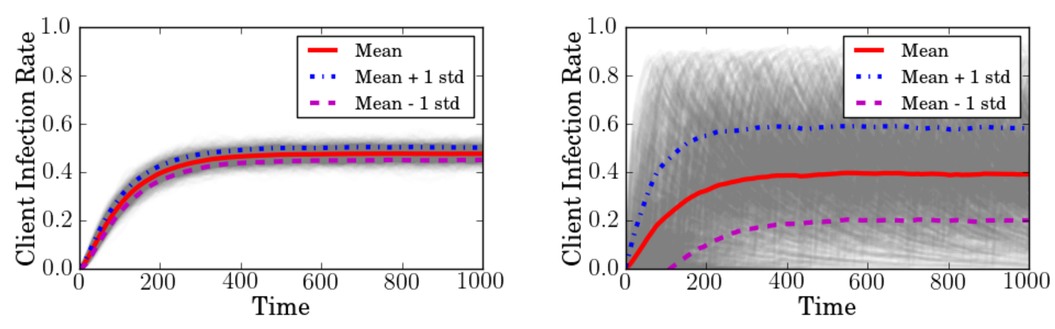
Figure: Comparison of different intervention policies, using outgoing (egress) and incoming (ingress) filtering, for different subsets of the AS graph

Results: Graduated Response and Drive by Downloads

The solution to the steady state client infection rate is

$$\mathbf{P}_c = \frac{\mathbf{P}_s \rho_c}{\gamma_c + \mathbf{P}_s \rho_c}.$$

For realistic distributions, high variance in client infection rates obscures the steady state solution.

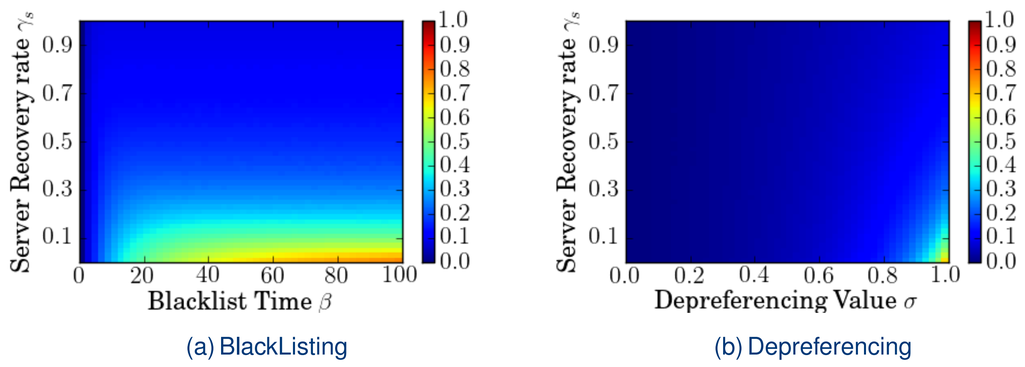


(a) Uniform website popularity

(b) Power-law website popularity

Figure: Variation in client infection rates over time. Individual runs are light gray.

Graduated response vs. Blacklisting:



(a) BlackListing

(b) Depreferencing

Figure: Average client infection rate for various blacklisting times (a) and depreferencing values (b), and website recovery rates. Each data point is the client infection rate averaged over 1000 runs.

For most depreferencing values and blacklist times, graduated response outperforms blacklisting as a intervention policy.

Conclusions

- Agent based models are a powerful tool for modeling interventions in complex networks
- Real world distributions and networks give rise to noisy, unpredictable results, requiring multiple runs to understand steady state dynamics