New Approaches to Characterizing Scam-Hosting Connectivity

Le Nolan*, Robert Beverly*, Joel Young {lenolan,rbeverly,jdyoung}@nps.edu

Motivation

1. On-line scams (pharmacy sales, phishing sites) continually evolve
2. Most recently, using multiple levels/ types of indirection (HTTP, DNS)
3. Existing passive traffic analysis techniques rely on IP addresses, communication structure, redirection patterns, etc – can be evaded
4. Traffic characteristics should be agnostic to evasion

Facts

1. Prior work finds significant redirection and traffic proxying by botnets
2. Scam content hosted by bot CDNs and by countries with poor connectivity

Hypothesis

Transport-layer traffic analysis of intermediate and landing pages reveal poor connectivity?

How connected are scam servers?

Scam Connectivity "Quality"

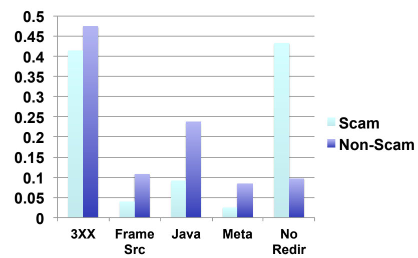
1. We're agnostic to IP, DNS names, registrars, etc.
2. Collect *Transport-layer* traffic features that reveal:
   - Asymmetric bandwidth
   - Busy bots and/or poorly connected hosts
3. More detailed than NetFlow-style statistics:
   - Retransmits (in/out)
   - RSTs/FINs (in/out)
   - Congestion Window (min, zero)
   - 3WHS and per-segment RTT variance
   - Packet inter-arrival jitter

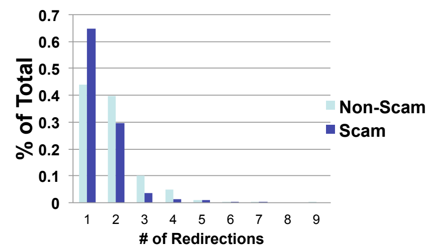
Experiment

- Web-crawl: Alexa Top 10K and 35K known-scam URLs from spam sink
- Record transport layer information of each HTTP GET (including redirections):
- Find statistical discriminators between scam and non-scam hosts

Redirection Summary

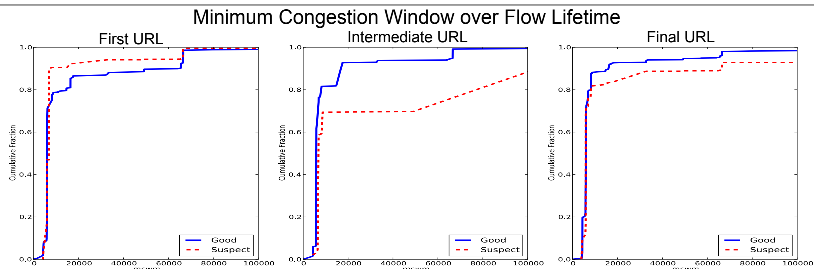
- Scam URLs = 23,762, 1.45 per
- Non-Scam URLs = 3,075, 1.8 per
  - Does redirection information still aid in discrimination?

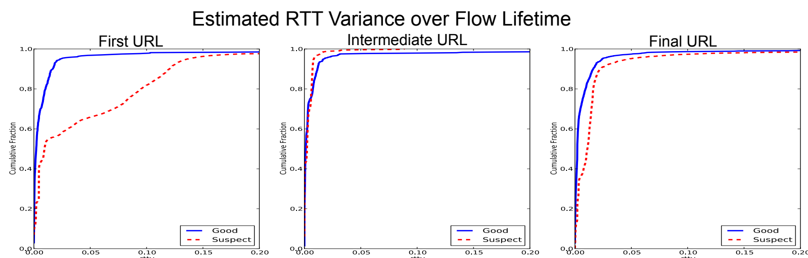




Transport-Layer Features

- Very different distributions (scam/ non-scam) depending on redirection stage (initial, intermediate, terminal)
- Confirms previous observations that bots perform redirection





Classification

•Using data with 50% "good", 50% "scam":

Method	Acc	Sens	Spec	PPV	NPV
Bayes	0.760	0.715	0.808	0.795	0.731
SVM	0.874	0.816	0.935	0.929	0.830
Decision Tree	0.937	0.943	0.931	0.934	0.940

* Supported by: Cisco Systems Unive