INDIANA UNIVERSITY
BLOOMINGTON

Microsoft Research

Finding the Linchpins of the Dark Web: A Study on Topologically Dedicated Hosts on Malicious Web Infrastructures

Zhou Li, Indiana University Bloomington
Sumayah Alrwais, Indiana University Bloomington
Yinglian Xie, MSR Silicon Valley
Fang Yu, MSR Silicon Valley
XiaoFeng Wang, Indiana University Bloomington

The Big Web


Google


amazon.com


WIKIPEDIA
The Free Encyclopedia


facebook


ebay

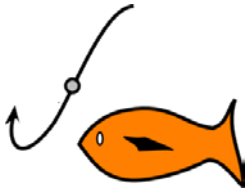

twitter


Y!


YouTube

The Dangerous Web

What bothers you?



Phishing

Scam

Theft

Drive-by Download

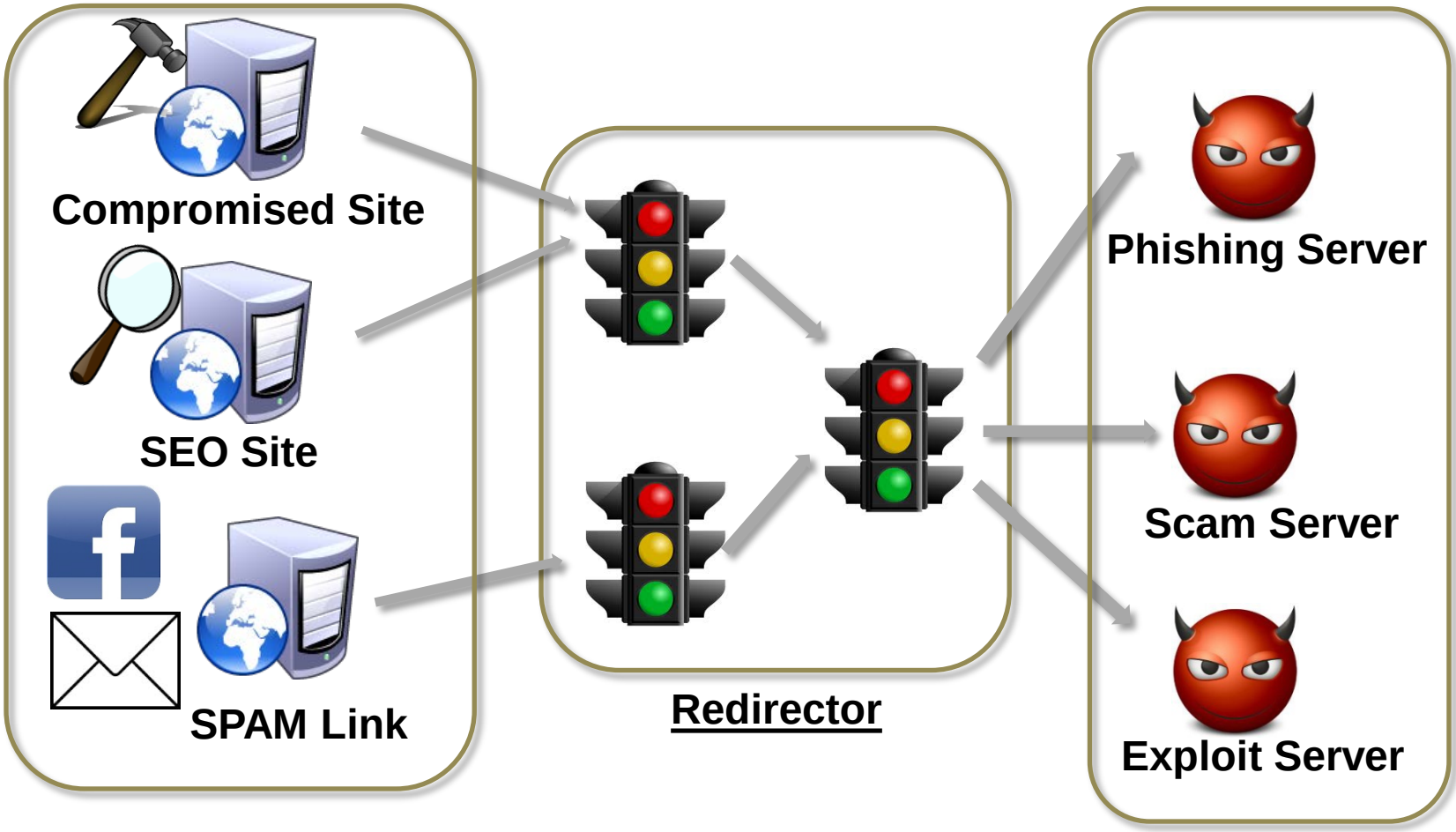
Bot

Who is behind?

Individual Hacker

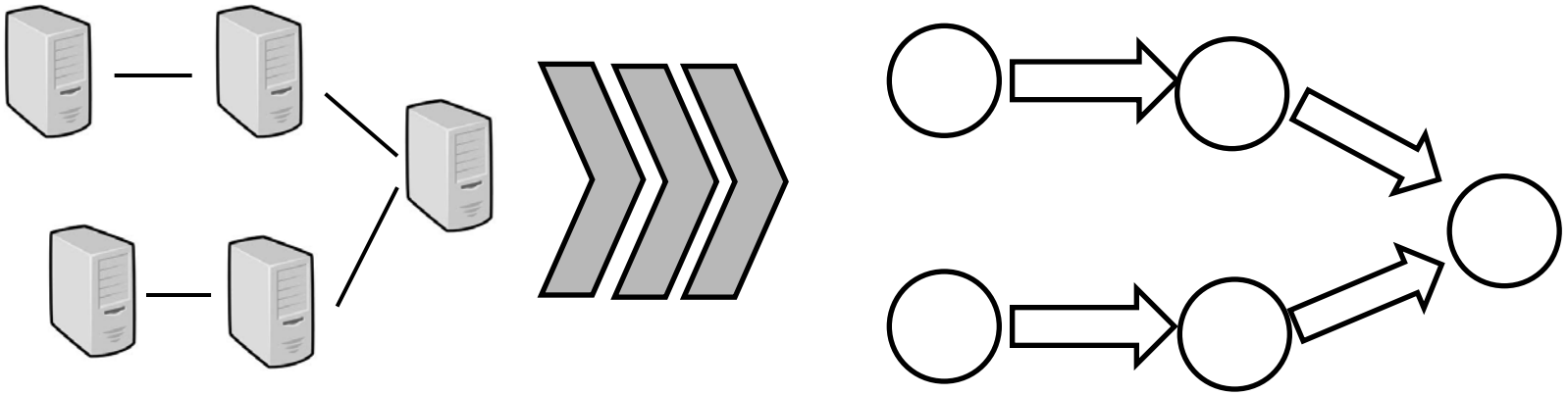
Criminal Organization

Malicious Web Infrastructure



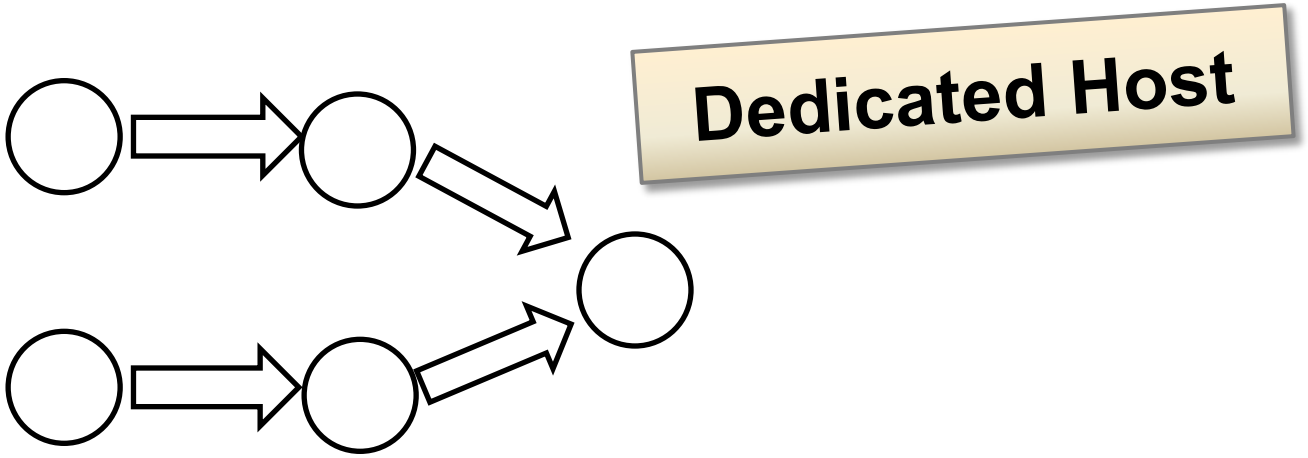
Our Work - I

- What is the topological view of malicious Web infrastructure?
  - ✓ Build redirection graph



Our Work - II

- What are the linchpins?
  - ✓ Topologically Dedicated Malicious Hosts: All redirection paths are malicious
  - ✓ Stay in “dark” side, far from “bright” side



Our Work - III

- How can we find the linchpins?
  - ✓ Topologically detector
  - ✓ Not relying on semantics of attacks
- Study of Traffic Direction System (TDS)
  - ✓ Landscape, Lifetime, Parking

Outline

- ❑ Build graph 
- ❑ Find the linchpins
- ❑ Study Traffic Direction System (TDS)

Data Collection

- Dynamic Crawler
  - ✓ Firefox extension
  - ✓ Redirection: JavaSc… …n, …

Client-side redirection

```
<script>
var a = "<iframe src =...>";
document.write(a);
</script>
```

- Data Source

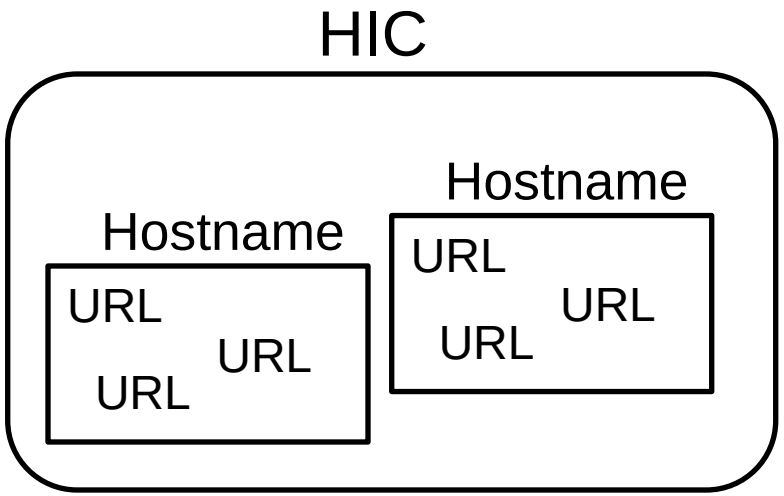
Feed	Type	Start	End	# Doorway URLs
Microsoft	Malicious	3/2012	8/2012	1,558,690
WarningBird[1]	Malicious	3/2012	5/2012	358,232
Twitter Search	Mostly good	3/2012	8/2012	1,613,924
Alexa	Mostly good	2/2012	8/2012	2,040,720

[1] WarningBird: Detecting Suspicious URLs in Twitter Stream, Lee et.al, NDSS'12

Building Redirection Graph

- Data Labeling
  - ✓ Malicious (Forefront)
  - ✓ Legitimate (Whitelist)
  - ✓ Unknown

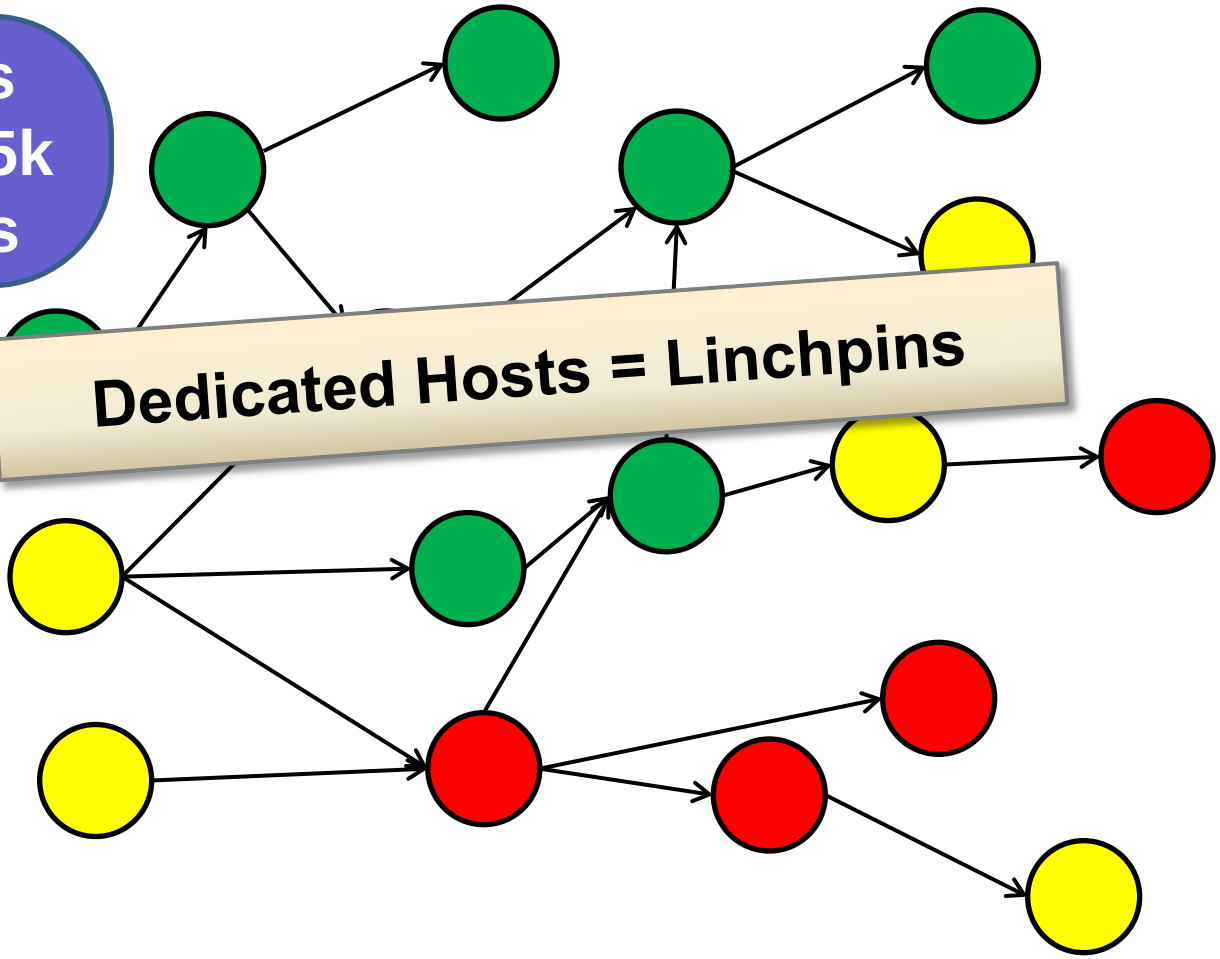
2M Nodes, 9M Edges



- Node Constructing
  - ✓ Node: Hostname-IP Cluster (HIC)
- Edge Constructing
  - ✓ Link 2 nodes if there is an URL redirection

Building Redirection Graph (Cond.)

>70% malicious paths through 15k dedicated hosts



Legitimate
(Totally good)

Dedicated
Malicious
(Totally bad)

Non-dedicated
Malicious
(Mixed)

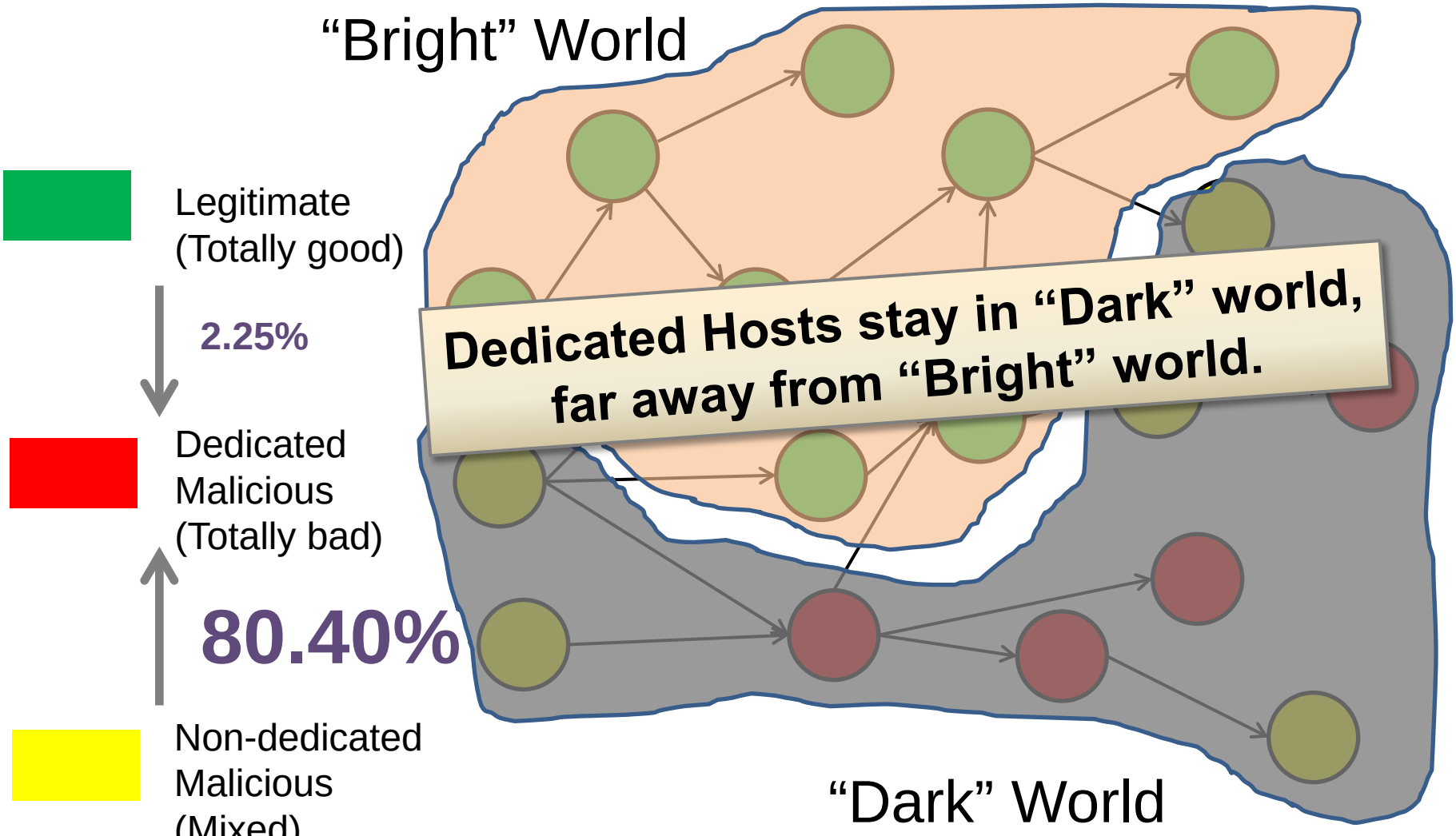
Outline

- ❑ Build graph
- ❑ Find the linchpins 
- ❑ Study Traffic Direction System (TDS)

Finding the Linchpins is Challenging

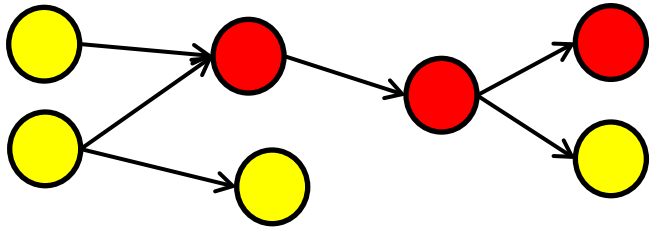
- Serve for different attacks and different sources
- Hide by cloaking
- Mixed with legitimate nodes

Properties of Dedicated Hosts



Topological Detector

- Can we leverage this topological feature? **YES!**
  - ✓ Use known bad and good as seed
  - ✓ Detect dedicated hosts
  - ✓ Expand the result using topology
- Work on different attacks and different sources

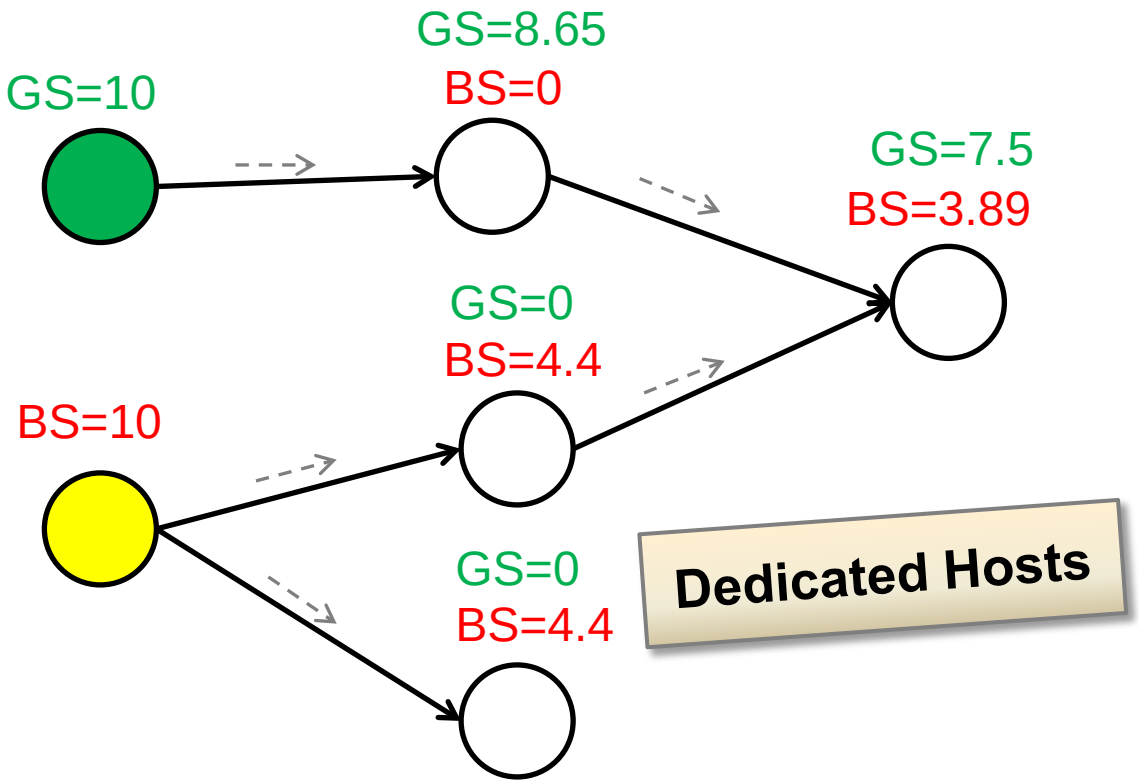


Topology!

Hunting Dedicated Hosts

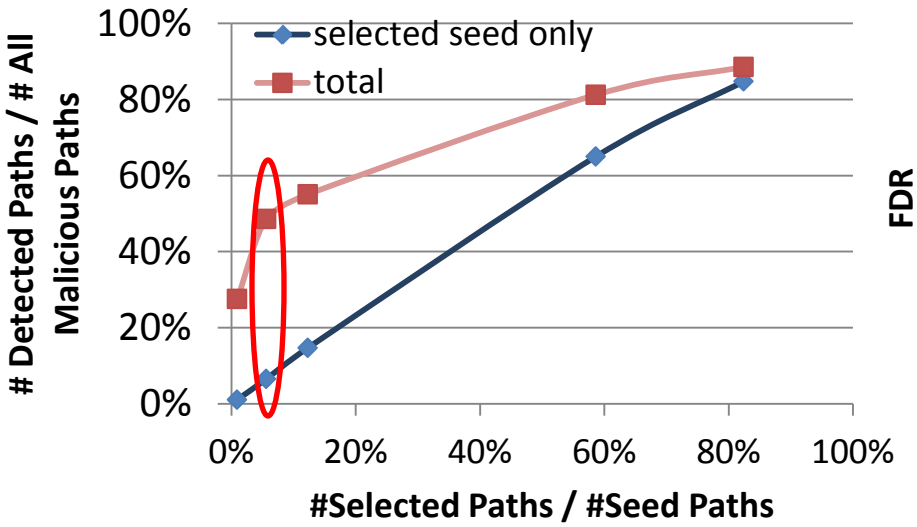
❑ Score Propagation using PageRank { Good / Bad

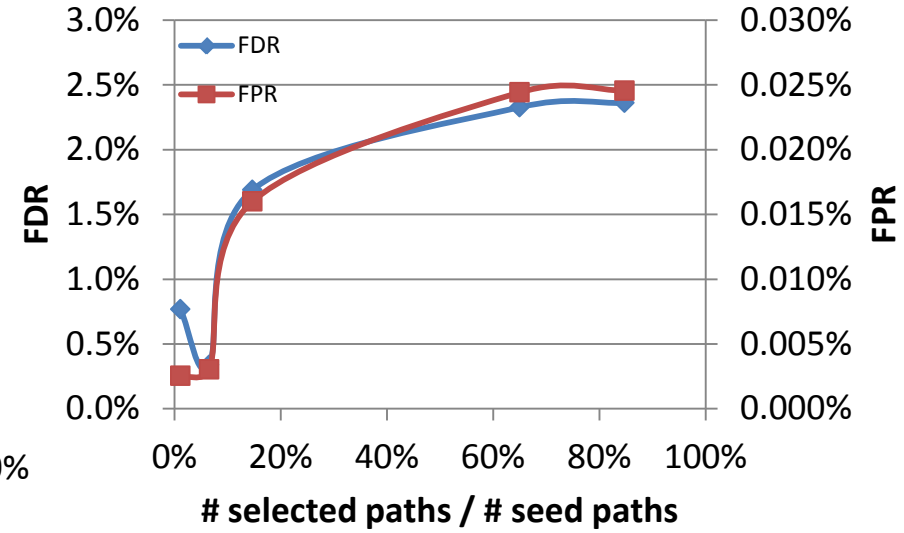
- ✓ Assign score to good seed
- ✓ Assign score to bad seed
- ✓ Score Propagation
- ✓ Choose nodes with high bad score and low good score



Evaluation

- Using x% of known malicious hosts as bad seed
- How many more malicious paths can be identified?





5% bad hosts, 7x expansion rate

0.025% FPR, 0.34% FDR

Evaluation (Cond.)

- Detect new hosts
  - ✓ 6k new malicious hostnames identified
- Detection result can serve as new seed
  - ✓ 12x expansion rate if rolling back the detected result
- Path sharing across different sources
  - ✓ 56% malicious paths from WarningBird feed overlapped with Microsoft feed

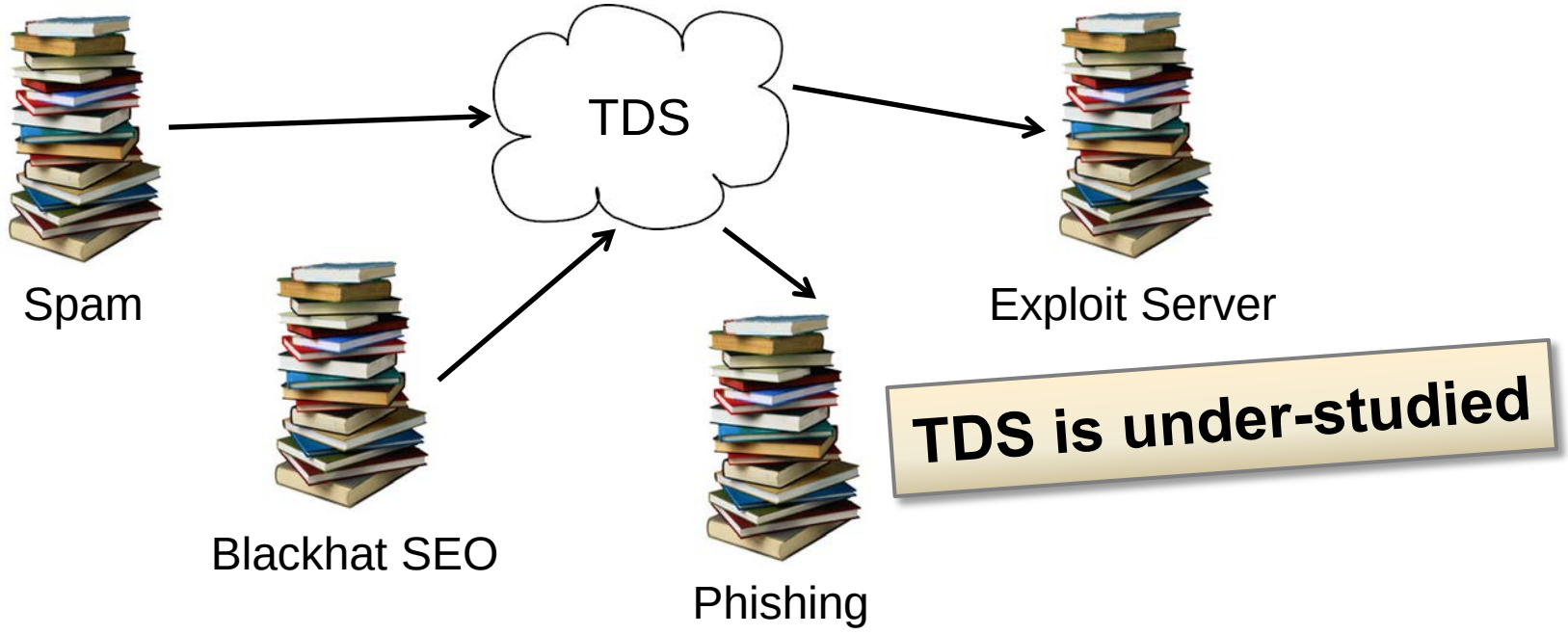
Outline

- ❑ Build graph
- ❑ Find the linchpins
- ❑ Study Traffic Direction System (TDS)

Traffic Direction System (TDS)

- Underground traffic brokers who buy traffic from generators (e.g., malicious doorways) and sell to consumers (e.g., exploit servers)

- >50% of the malicious paths go through TDS.



TDS Landscape

- 71% TDS use Sutra TDS kit.
- 26% Dynamic DNS, 14% Free Domain Providers.
- Inbound traffic: 97% from doorway, 6% from non-doorway redirectors.
- Outbound traffic of Active TDS: 49% to exploit servers, 3% to scam sites.

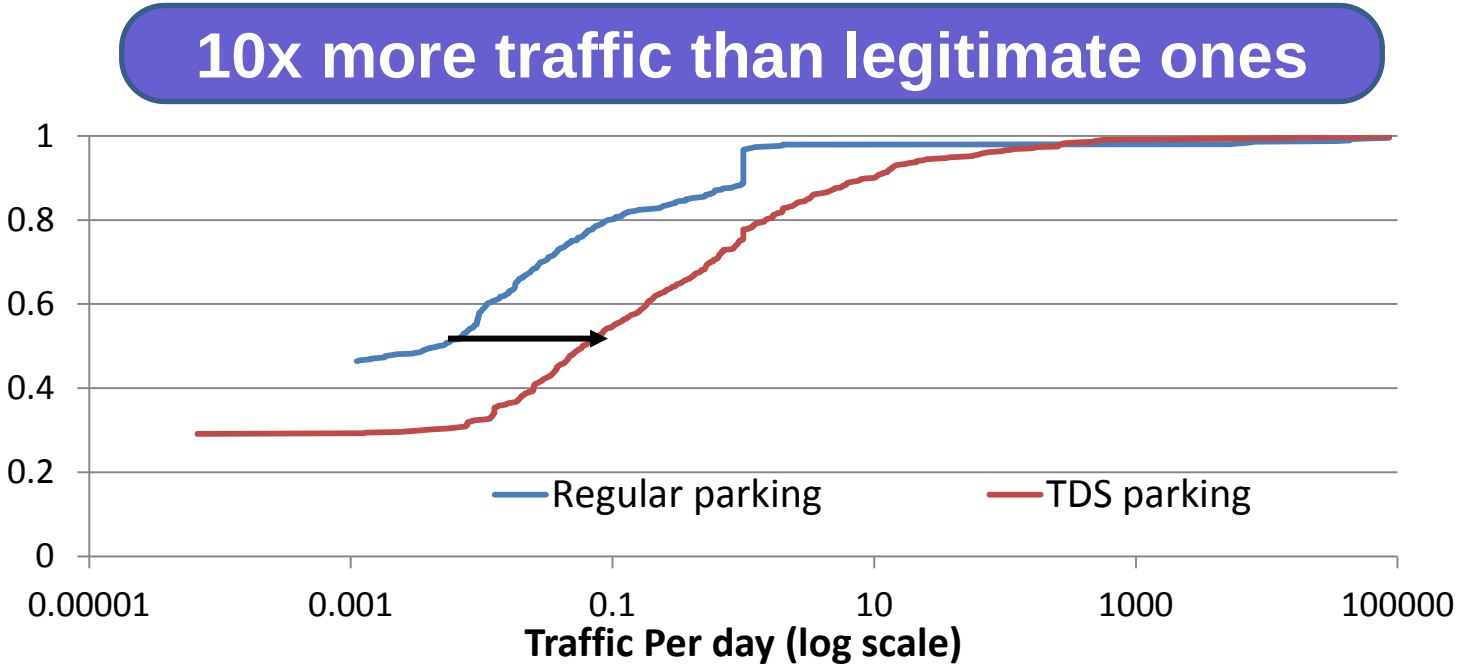
TDS Lifetime

- Query Passive DNS Database from SIE
- Lifetime: interval of host when A record bounding to bad IP
- Result
  - ✓ 65 days (exclude DDNS/ Free Domain Providers)
  - ✓ Much longer than exploit servers, 2.5 hours[1].

[1] Manufacturing Compromise: The Emergence of Exploit-as-a-Service, Chris et. al., CCS'12

TDS Parking

- Even suspended TDSes monetized by attackers
  - ✓ 20% TDS hosts parked.
  - ✓ Continue to receive high volume of traffic after parking.
  - ✓ 62% traffic to ad-networks.



Conclusion

- Dedicated Hosts are critical in Malicious Web Infrastructure
  - ✓ Serve >70% malicious paths.
- Detect Dedicated Hosts only using Topological information.
  - ✓ No need to know semantics of specific attacks.
  - ✓ 7x expansion rate, <1% FPR & FDR
- TDS Hosts deserve in-depth study
  - ✓ Key role in malicious Web infrastructure
  - ✓ Long lifetime (65 days)
  - ✓ Traffic monetized even after suspended

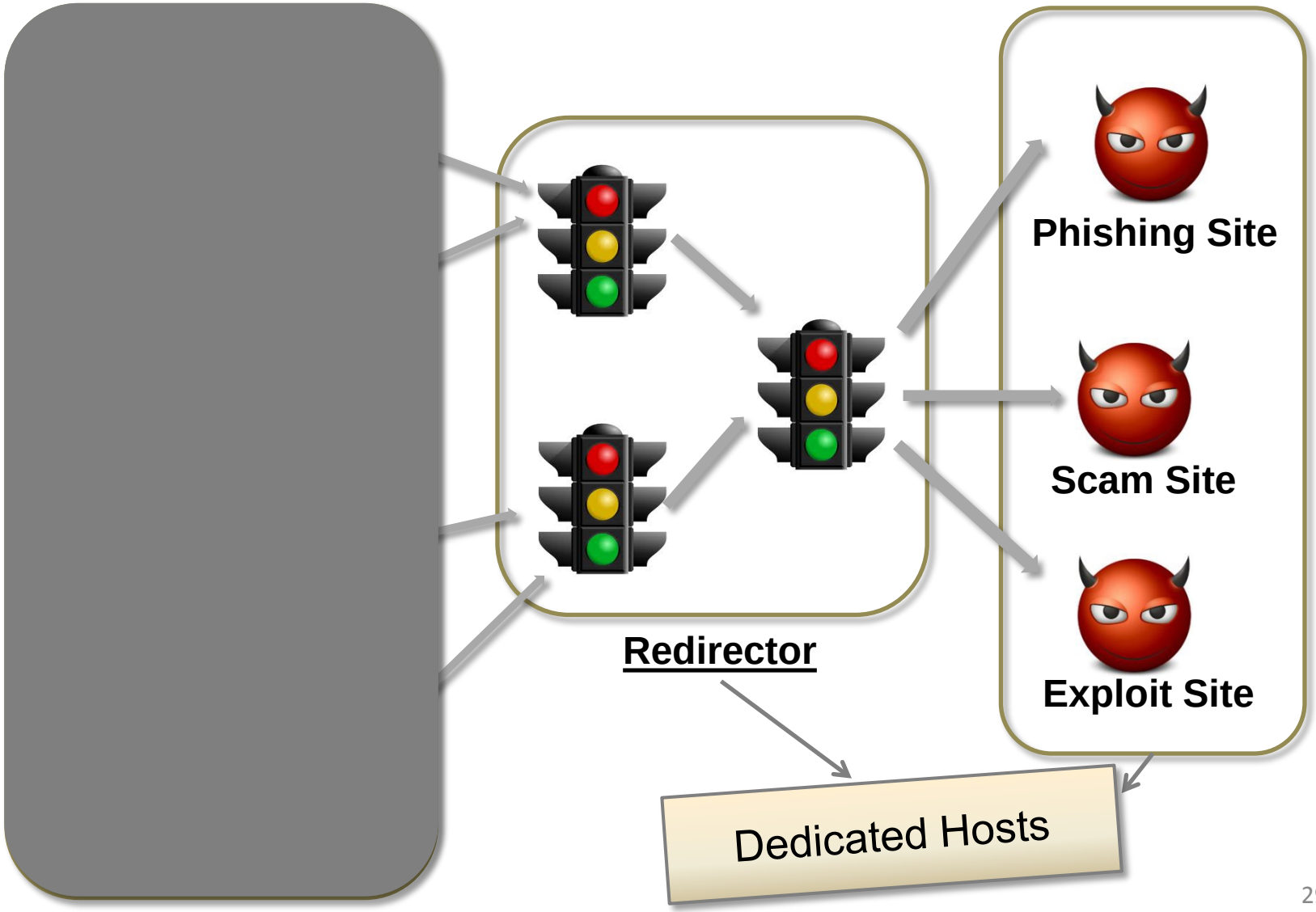
Thank You!

Q & A

Validation

- Forefront reporting
- Content and URL clustering
- Safebrowsing reporting

Dedicated Malicious Hosts



Limits of Prior Works

- Code Analysis [Cova'10, Curtsinger'10]
  Obfuscate Code
- Code Execution [Provos'08, Wang'06]
  Detect Emulator
- URL Pattern [Zhang'11, John'11]
  Regenerate URL Pattern
- Redirection Chain [Li'12, Lee'12, Lu'11]
  Depend on Semantics of Attacks