The Dark Oracle: Perspective-Aware Unused and Unreachable Address Discovery

Evan Cooke*, Michael Bailey*, Farnam Jahanian*, Richard Mortier†

*University of Michigan
†Microsoft Research
Honeynet Monitoring

• *Unused* Internet addresses form the basis for a broad class of security systems: *honeynet monitoring*
  – Honeypots
  – Darknets
  – Telescopes

• Goal today is to dramatically expand the scope of this class of systems by redefining what we mean by *unused Internet addresses*
A Closer Look At Address Usage

- 67% of IPv4 addresses not announced via BGP
A Closer Look At Address Usage

- 67% of IPv4 addresses not announced via BGP
- 58% of addresses in an organization not internally routed
A Closer Look At Address Usage

- 67% of IPv4 addresses not announced via BGP
- 58% of addresses in an organization not internally routed
- 65% of addresses allocated to a DHCP server not used
Honeynets

• Sample of existing projects:
  – Honeynet Project ~ /32 (1 address)
  – honeyd ~ /24 (255 addresses)
  – Potemkin Honeypot Farm ~ /16 (65K addresses)

• Observations:
  – Very small % of available addresses
  – Contagious blocks of unused addresses
  – Only globally advertised and globally reachable addresses
Automatic Address Discovery

• Our Goal:
  1. Redefine *unused* addresses as *dark* addresses
  2. Develop a methodology for automatically discovering dark addresses
  3. Route traffic to those addresses to honeynet monitoring systems

• Vision: Pervasive Honeynets

  Give honeynets the visibility of 1,000’s honeypots distributed inside and outside the network
Redefining Dark Space

- Honeynets typically configured to monitor *globally advertised and globally reachable* unused addresses
Unused and Unreachable Addresses

- We propose a *dark address* should be defined as an *unused* or *unreachable* address.
Unused and Unreachable Addresses

We propose a dark address should be defined as an unused or unreachable address
With this expanded definition we can adopt the perspective of a specific network and construct *incoming* and *outgoing* honeynets.
Architecture

- Leverage distributed address allocation data

Diagram:
- External Routing (e.g., BGP)
- Internal Routing (e.g., OSPF)
- Host Config. (e.g., DHCP)
- Address Manager
- Organizational Network
- Honeynet/HoneyFarm
Finding Unused Addresses

• Why is it hard to get more unused addresses:
  1. Address allocation information distributed across devices, applications, domains
  2. Address allocations change quickly (e.g. mobile users)
Potential Sources of Global Addresses

- Many unused addresses at each step in the address allocation process (unused = black)
Leveraging Internal Data

- Find globally unique and local routable addresses
Address Allocation Data Sources

- We find three major classes of potential sources of address allocation information
  1. External Routing Data
     - e.g. BGP
  2. Internal Routing Data
     - e.g. OSPF, ISIS, RIP
  3. Host Configuration Data
     - e.g. DHCP, LDAP
Address Misclassification

- But, what if you make a mistake and misclassify an address as dark or used?
- Two possible ways this can happen:
  1. State between data source and data reader become out of sync
  2. Inaccurate data source
- Can help prevent by:
  - sampling often
  - add delay to the used to dark transition
The Dark Oracle

• We implemented a prototype address discovery platform call the Dark Oracle

• Inputs
  – BGP feed from RouteView project as source of external routing data
  – OSPF listener as source of internal routing data
  – DHCP server monitor as source of host config data

• Outputs
  – A locally-accurate map of dark addresses
The Honeynet

- We passively captured packet to the dark addresses from the Dark Oracle
- Traffic was capture using:
  1. **Darktrap**: A userland program to differentiate dark traffic from live traffic from a route *span port*
  2. **Blackhole** Route: A fall-through route
Evaluation Plan

• Analyze the stability, density, and quantity of the addresses from the three data sources
• Deploy the Dark Oracle on academic, enterprise, and ISP networks
• Compare the resulting visibility to an existing honeynet
Data Source Eval: External Routing Data

- Bogon List
- RIPE + ARIN Allocations
- BGP from RouteViews

- More accurate information reveals significantly more and better external reachability information
Data Source Eval: External Routing Data

- Over the 1 month period the average change in addresses per 4-hours was 0.01-0.001% of IPv4 (~/16)
- Very small error incurred by sampling less frequently (0.04% error if sampled once per day)
We monitored DHCP server that managed 5120 addresses, 70% of those addresses never allocated.

An address newly classified as dark remained for an average of 8.85 days.
Dark Oracle Deployment

- Deployed Dark Oracle on a live network with ~10,000 hosts
- Darktrap run on 3Ghz system connected to a span port from a centrally located Cisco Catalyst 6500 (DHCP/BGP addresses)
- Blackhole route (OSPF addresses)
Addresses Discovered

Number of Prefixes, Dark IPs, and % of IPs Dark for Each Source
Daily Snapshot in September 2005

- Huge Outgoing Honeypot
- Large % of Unused Addresses At Each Level
Honeynet Detection Results

Unique Source IPs Observed with Different Darknet Techniques

1 week on a live /15 network with >10,000 hosts

- Bogon Darknet
- BGP Darknet
- DHCP Darknet
- IGP Darknet
- /24 Internal Darknet

Traditional
/24 Darknet

- All IPs
- TCP Only
- UDP Only
- ICMP Only

NSDI 2006
Honeynet Detection Results

Unique Local Source IPs Observed with Different Darknet Techniques

1 week on a live /15 network with >10,000 hosts

- Infected/Misconfigured Sources From the Same Network!
Conclusion

• By becoming **perspective-aware** and monitoring external reachability data we can construct *incoming* and *outgoing* honeynets.

• Automatic address discovery provides *order-of-magnitude* more unused addresses compared to existing honeynet allocation systems.

• **Pervasive** honeynet are better at detecting targeted attacks and are resistant to fingerprinting.
Questions