SIE IPv4 Darknet
DUST
San Diego, May 2012
Space

- There's lots of it

[ picture deleted ]
[ for reference, look for “darknet hilbert heat map” on google]

Who has a good recent diagram?
Are we really running out?

- IP counts increasing somewhat linearly - IPv6 emerging

Source: Internet Systems Consortium (www.isc.org)
Typical research

- tcpdump > dataset
- analysis < dataset > results
- cp results presentation
What we do

- How to efficiently distribute data?
- We efficiently encapsulate and redistribute
- ... in real time.
What's there

• 500k+ addresses, 10 networks
• >1000 pps
One way we get it

- ISP router cross-connects to SIE switch
- Router ends up broadcasting on SIE VLAN
- Cisco config-fu:

```plaintext
cisco(config)# router static
  address-family ipv4 unicast
    XX.XX.0.0/16 10.255.10.254
  arp vrf default 10.255.10.254 0202.0404.0606 ARPA
interface GigabitEthernet0/2/0/3.14
  description SIE Dark Net
  ipv4 address 10.255.10.1 255.255.255.0
  dot1q vlan 14
```
How to redistribute

- NMSG
  - Google protocol buffers
  - Encapsulation
  - Source Identifiers
  - Broadcast network plumbing
  - Net->File->Replay capability

Sender:
nmsgtool -dddd -V ISC -T pkt -i sie.14+ -m 1280 -s DESTIP/50140

Receiver:
nmsg-pkt-inject -I DESTIP/DESTPORT -o sie.14
ip route add blackhole X.Y.Z.0/24

nmsgtool -D -V ISC -T pkt -i eth0 -m 1280 -unbuffered \
  -s DESTIP/50140 -z -b 'net X.Y.Z.0/24'

nmsgtool -D -V ISC -T pkt -i eth0 -z -w FILE.nmsg -t 3600 -k kick.sh

Would love to get flow or Null0 traffic.
Uses

• Commercial:
  • Backscatter analysis – target watch
  • Probe sources mapping to botnets or “sources of interest” for IDS people.

• Research:
  • Test theories/predictions on live data
  • Combine with other data (netflow, bgp, passiveDNS?, others)
  • Loosely-coupled multi-processor approach
Levels of darkness

• V1 – black – no response
• V2 – dark-gray – limited response
  • Think sinkhole: reset after TCP handshake
• V3 – blue - Honeypot VPN
  • Darknet offers NAT transport to remote honeypot server(s) to get infected.
  • Infected server uses remote IP resources for study after initial infection session closed.
Challenges

• Anonymizing? (PII)
  • Not yet, we rely on privacy agreement
  • Can make your own anon wrapper
  • Can make 3rd-party summary tools
    – Standardized 5060/445/80/53/ICMP triggers and event correlation.- encouraged by Alberto
    – Real-time feedback of event reports from ISPs

• Timing
  • We can preserve timing at capture, but replay and distribution in PCAP has timers set to current when regenerated.
Challenges

- Some ISPs have only flow data available – perhaps we should make another type?
- Getting more data
- How do you collect data?
- What formats do you use?

Email: sie @ isc.org
Future

• Let's take some common methods and tools and publish them so that anyone can apply them to their darknets and share classification results.

• Let's show ISPs what good can come from their contributing data in real time to make available to researchers. Possible feedback loop for them.