Expanding IoT Honeypots to Attract IPv6 Attacks

Phil Roberts proberts@globalcyberalliance.org

Leslie Daigle ldaigle@globalcyberalliance.org

Quick Overview of GCA Honeyfarm Research

- GCA is a not-for-profit with offices in the US, UK, and Europe working around the globe to help deliver a secure and trustworthy Internet
- We have operated a (v4 only so far) honeyfarm with over 200 open-source sensors collecting data for over 4 years
  - We record every telnet and ssh attack (IP addresses, ports, embedded URLs, scripts, etc.)
  - Rolling out new honeypot technology presently that will add detection of http and https attacks
  - 10 million unique attackers recorded over the history of the project, on average from around 3000 unique ASNs daily
  - More than 5 billion attacks recorded
- Partner with network operators to clean up unwanted traffic originating from their networks
- Partner with research organizations who would like to do research using our data
  - First research report (Max Planck Institute): https://dl.acm.org/doi/10.1145/3618257.3624826?_gl=1*yiwyjv*_gcl_au*NjM4MzM3NzA4LjE3MjEzMzA2Mzk.
  - We have partnered with a half a dozen research organizations interested in analyzing our data repository and are always interested in investigating collaboration with other research organizations
- This talk will be about the steps needed to do similar work in IPv6. Our paper describes these findings in more detail: https://globalcyberalliance.org/reports_publications/expand-honeypots-to-include-ipv6/
- Thanks to Microsoft for supporting this research

How our honeypots work

- Imitate an IoT device (a linux-like device), accessible on the public Internet (we attract attacks looking for high-powered devices also)
- We require logins, but accept almost any login credentials
- The honeypot records all the actions of the attackers, but launches no attacks
- Most of the malicious activity we detect is distribution of malware
- We make no particular efforts to be found, rely on malware distributors scanning or using lists of previously scanned devices
- Distributed around the globe in all sorts of networks

Where we want to go

- IPv6 usage is trending upwards, but IPv4 deployment is pervasive
- At some point, IPv6 only networks will emerge
- Some tidbits of malicious activity is seen over IPv6 now
- Maybe more instrumentation would see more?
- It would be good to be ahead of the curve in detecting malicious activity directed over IPv6 and directed towards IPv6 interfaces on these networks/devices
- We would like to collaborate with others with similar interests and concerns

How honeypots attract attention (in v4 world)

- Just by existing really, honeypots want to be found and attacked
- Scans
  - Multiple security organizations scan the Internet daily
  - Many botnets scan the Internet daily
  - Some scan it multiple times a day, each scan searching for something different
  - Our sensors are scanned over 100 times daily, sometimes by known security research organization, sometimes by others
  - The ability to scan the entire (IPv4) Internet easily and efficiently gives attackers fresh data to help them focus their attacks

Problems with attracting attention (in v6 world)

- Remember honeypots want to be found and attacked
- IPv6 space just can’t be exhaustively scanned (theoretical time)
  - - one security researcher who scans calculated this to be @10^25 years
- IPv6 interface id space can’t be scanned (theoretical time)
  - Even if you know a /64 prefix, for example, you can’t realistically find all the configured interface addresses – it would still take years

How to get your device found?

- Ignore the IETF's very good advice about privacy and security when configuring v6 addresses on your honeypot
- RFC 7721 described 4 privacy and security vulnerabilities
  - Activity correlation over time, location tracking, scanning, exploitation of specific device vulnerabilities
  - but we're only interested in one (scanning) as we really want attackers to find our honeypots
  - So, don't use a hard-to-find address (random or cryptographically generated, for example)

Discoveries of bad or sloppy practices (or how to get your device found part 2)

- Too easy to find devices near public resources
  - Someone will stand up a public server of some sort
  - Servers and devices in the same network will be numbered near the address of the public resource, so scanning nearby address may find other targets for attack
- Observed hosts numbered with just the prefix and the rightmost bits
- Operators do not enable standard protections on v6 interfaces that they have enabled on v4 interfaces (of course a honeypot would not want these protections)
- Mimicking these observed sloppy practices may help your honeypot get discovered

Prior research

- Full bibliography in the paper, here are some key prior publications our work has depended on:
  - Richter, Gasser, and Berger, “Illuminating Large-Scale IPv6 Scanning in the Internet”
  - Murdock, et.al., “Target Generation for Internet-wide IPv6 Scanning”
  - De Coster and Kijewski, “Internet Spelunking: IPv6 Scanning and Device Fingerprinting”

Key findings

- Hitlists
  - Developed by researchers for conducting studies and measurements
  - Conducted with a high degree or care wrt ethical concerns
  - Published lists for researchers to use
  - Similar techniques might be employed by adversaries
- Observed IPv6 scans
  - There aren't many
  - Some are known security company sources
  - Others look more like penetration tests than broad scanning to find vulnerable hosts
- Results from reported scanning
  - Not so much scanning to find new host
  - Scanning known hosts (from, for example, published IPv6 hitlists) looking for already known vulnerabilities (to clean them up and reduce vulnerabilities)

Some observations from our personal research

- Over half of our engineering team live in places where IPv6 Internet is not an available
- Address assignment practices (from networks)
  - Every prefix assignment we observed (/56, /60, /64) was unchanged over the 9 mos of the work
- Address generation practices (on devices)
  - Constant CGAs (all devices apparently)
  - Constant (on a few devices), seemingly easy to find addresses: prefix::final 16 bits (easy to scan if someone knows the prefix – as far as I can tell this address is never used
  - Variable CGA (on most devices) that changes at least every 24 hours, sometimes as frequently as every time (a laptop) sleeps

Building and deploying an IPv6 honeypot

- Building an IPv6 version of our honeypot is a straightforward undertaking
  - Enabling IPv6 interfaces
  - Updating data collection to record IPv6 information
- Deploying IPv6 honeypots is a little more work
  - Need to be deployed in networks that have IPv6 capabilities
  - Need to make oneself found using the ways to get oneself discovered above:
    - Deploy a stable v6 server that is publicly known through the DNS and have a honeypot at a nearby address
    - Deploy honeypots with conspicuous addresses (using only low-order bits of IPv6 address)
    - Get one's honeypot on an IPv6 "hit list"

Best practices in IPv6 deployment (for those not trying to get their honeypots detected)

- Use random interface identifiers where possible (on any resource that doesn’t need to be publicly reachable)
- Make frequent address changes where possible
- Frequent network renumbering adds even more protection (this has been observed in networks in Germany for example)
- Discard addresses that are no longer used

Questions?