CATEGORIZING AND ANALYZING DISCRETE DARK TRAFFIC CLASSES

Michael Collins, USC-ISI
mcollins@isi.edu

Stephen Schwab, USC-ISI
schwab@isi.edu

July 13-14, 2021

This material is based upon work supported by DARPA under Contract No. W911NF19C0058. The views, opinions, and/or findings expressed are those of the author(s) and should not be interpreted as representing the official view or policies of the Department of Defense or the U.S. Government.

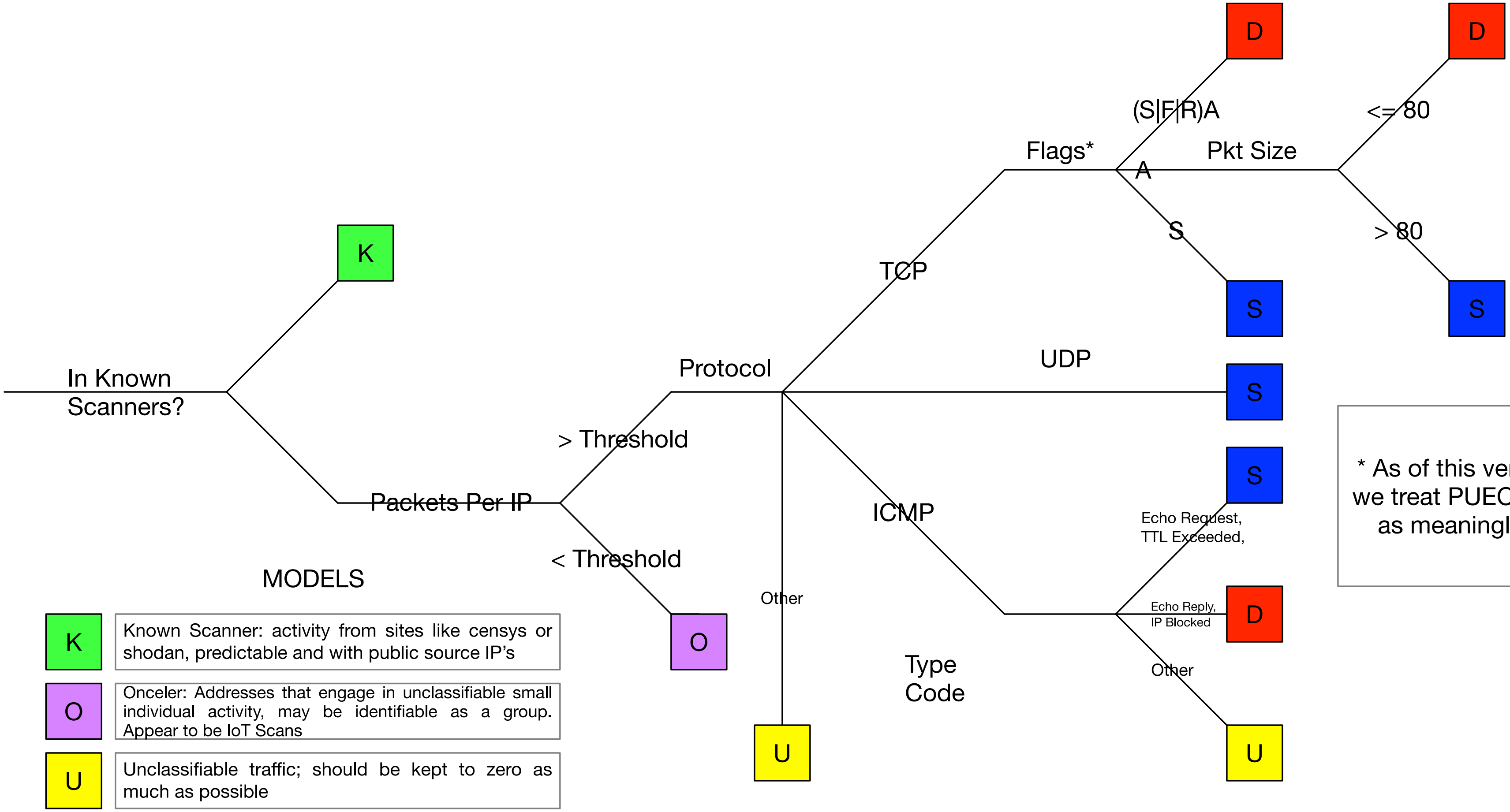
Introduction

- We developed a deeper partitioning system which breaks traffic into more specific categories
- We split out *known* scanners versus more explicitly *hostile* scanners
  - Within the second, we have further categories
- We will discuss these different categories and why they matter

Context

- ISI: 3 Discrete /24’s
- Worked with 2 months of traces in 2020
  - 2020/11/01-2020/12/31
- Data analyzed using SiLK toolkit
  - Primarily for arbitrary IP address collections

Initial Partition



* As of this version, we treat PUEC flags as meaningless

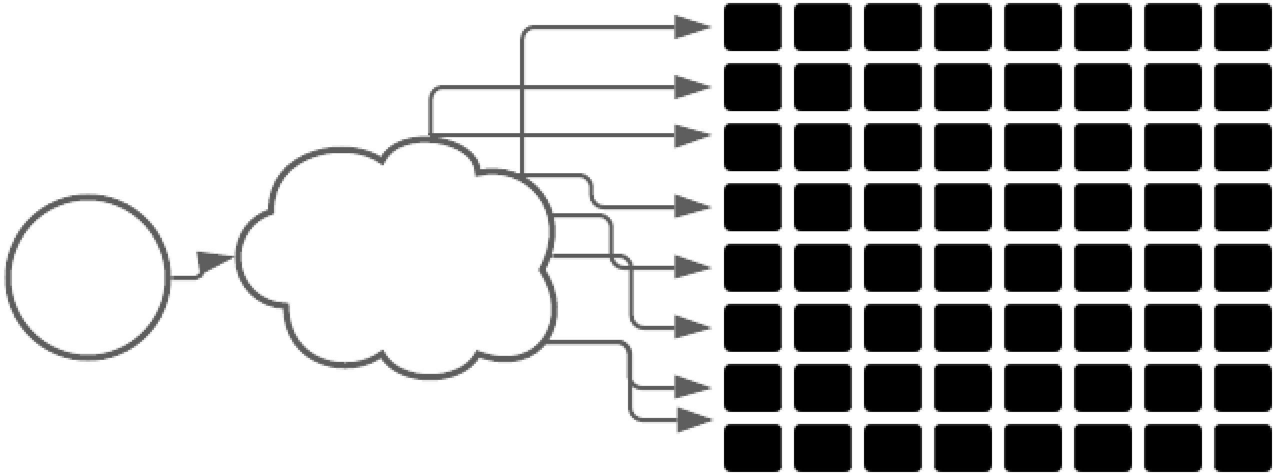
MODELS

🟩 K	Known Scanner: activity from sites like censys or shodan, predictable and with public source IP's
🟪 O	Onceler: Addresses that engage in unclassifiable small individual activity, may be identifiable as a group. Appear to be IoT Scans
🟨 U	Unclassifiable traffic; should be kept to zero as much as possible
🟥 D	Backscatter from DDoS attacks. Mostly SA, but also ICMP 3/10.
🟦 S	"Normal" hostile scanning

Different Scanning Classes

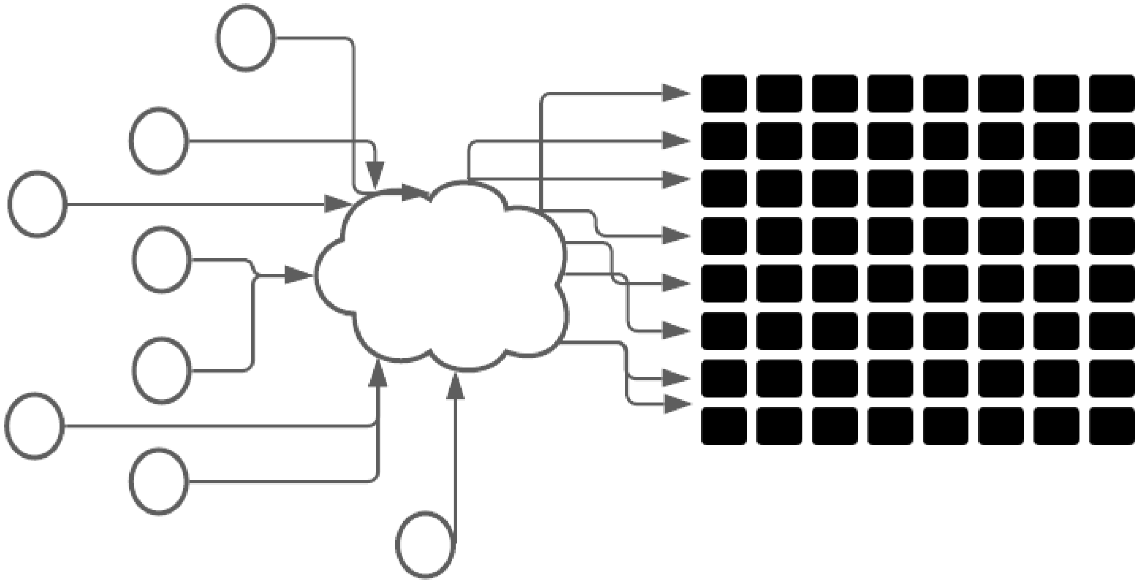
- By protocol
  - TCP: Looking for exploits (Telnet (yes), SSH, SMTP, HTTPS)
  - UDP: Looking for reflectors (NTP, SIP, SNMP, SSDP)
- By Goal
  - Known scanners: looking for vulnerable hosts for public announcement
  - Hostile scanners: looking for hosts to exploit
- By Behavior
  - Knowns/Long: hit all targets over brief time
  - Shorts: appear briefly, then go away

IBR Type: Scanning



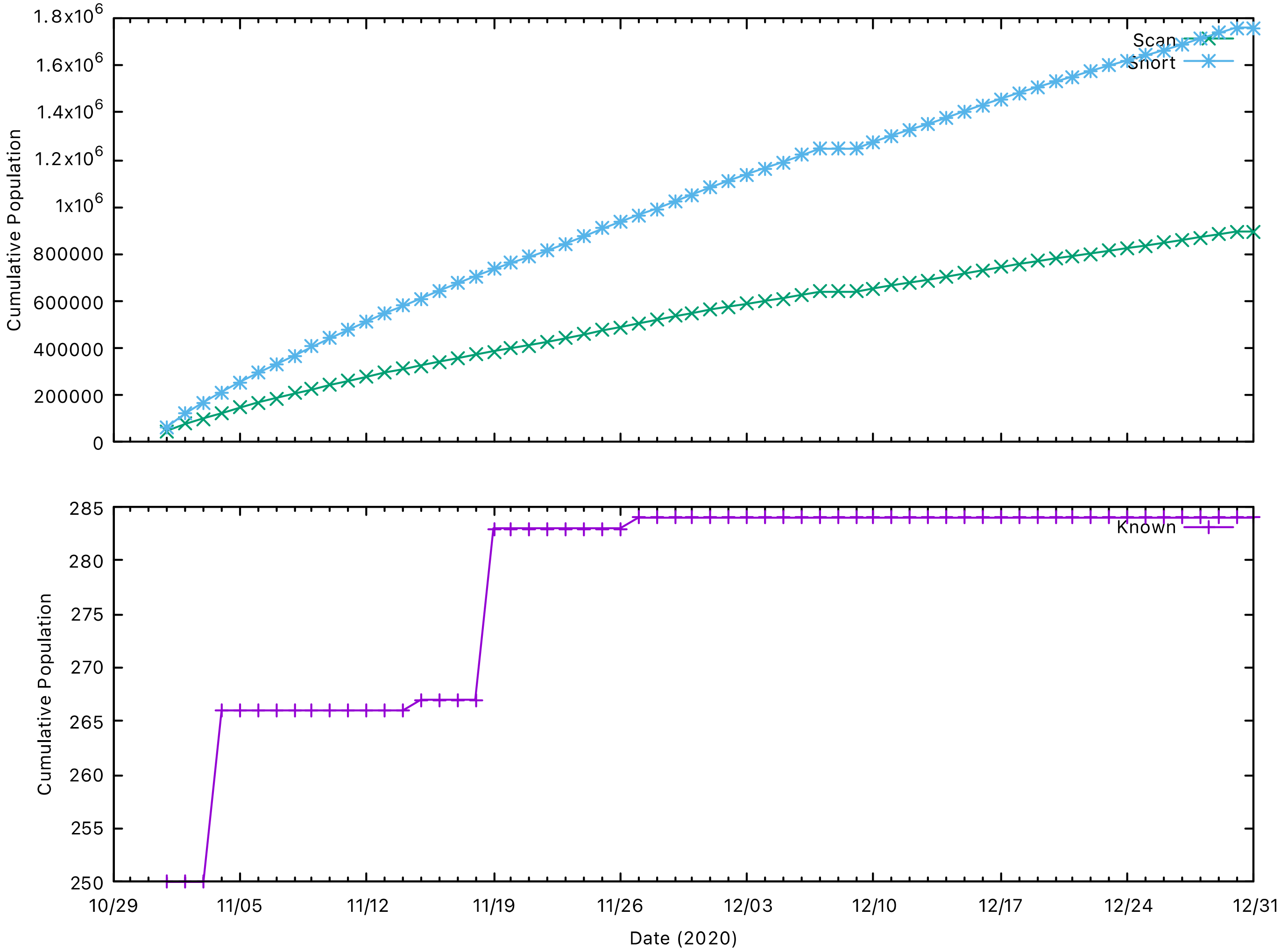
- Single address targeting a high number of distinct destination addresses
- Known scanners: Shodan, Censys and other organizations that announce their scans
  - Fixed addresses, known port destinations
  - May change over time, but the changes are slow and obvious
- TCP scanners: S, odd ACK behaviors
- UDP: All UDP

IBR Type: Short



- Appears to be scanning (SYN only), but very small activity (<4 packets per host)
- Very short lifetime – appear in one day, and then up to two months later haven’t seen repeats

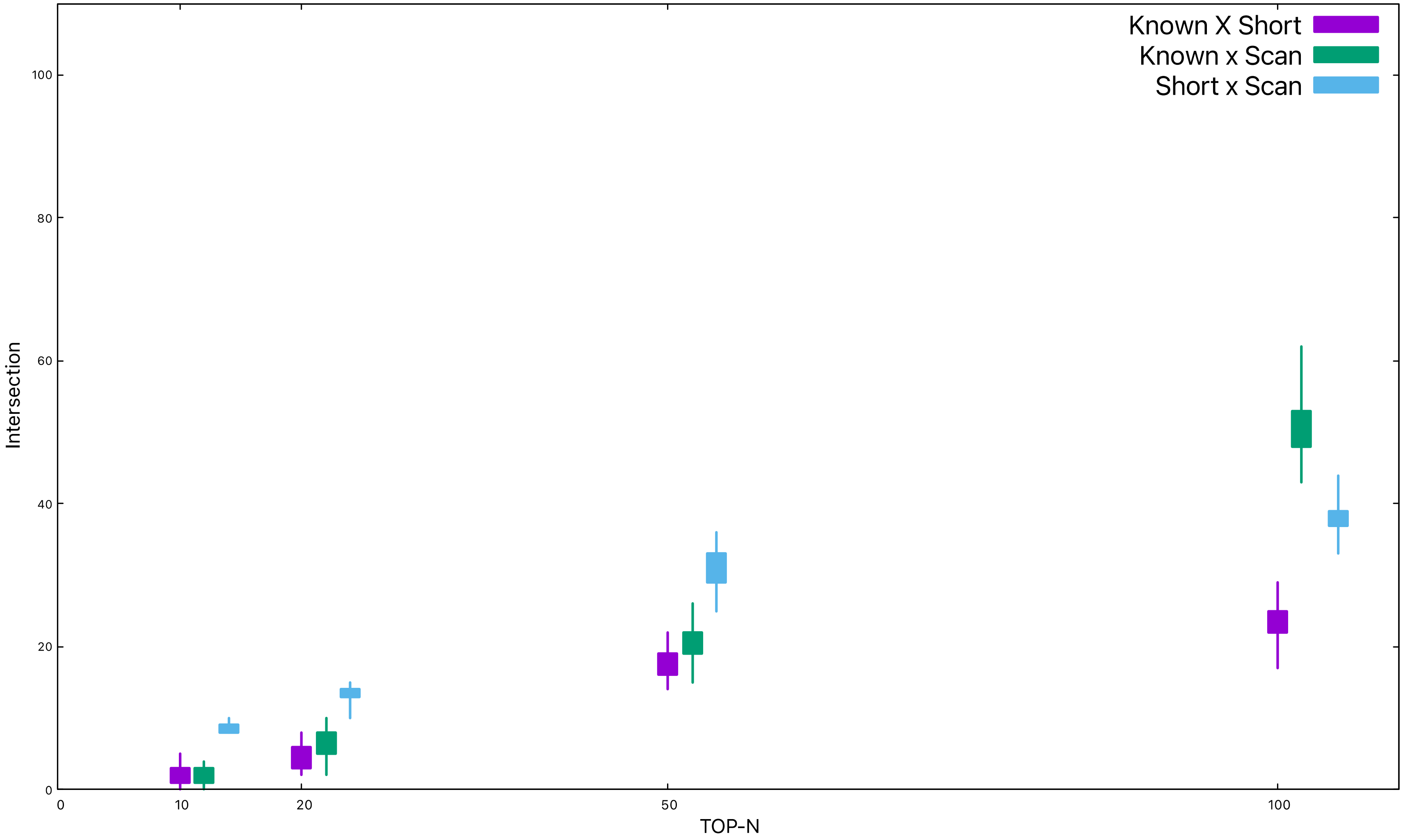
Different Populations Grow Differently



Observations on Different Population

- Note: small flat point around 12/08-12/11 is due to lack of data
- In both scan and short case, there's a constant population increases
  - But shorts have practically no overlap
  - Not sure where scan/short barrier is *behaviorally*
- Generating the known population requires a list of these scanners
  - The sharp increases happen when a known scanner changes their scanning hosts
  - *We don't have a complete set of known scanners*

Different Targets



Known Vs Others

- Knowns are taking look at a different set of vulnerabilities than other scanners
  - *Also different from each other*
  - Known scanners are looking more for RAT ports (1177, 54984)
  - Attackers are more current (?) (5555, 2323, 23)

Conclusions

- Scanning behavior is not monolithic
  - There exist discrete populations within "scanning" which we can identify behaviorally and from point of origin
- The known scanners need to be split off as they operate differently than other scanners
  - Requires out of band investigation as companies come and go
- Split between short and long scanners is an ongoing problem