IP Profiling Analytics & Mission Impacts

[redacted]

Tradecraft Developer
CSEC – Network Analysis Centre

[redacted]

May 10, 2012

SIGINT

Canada

Example IP Profile Problem

Target appears on IP address, wish to understand network context more fully

Example Quova look-up & response for [redacted]

Lat. 60.00 Long: -95.00 (in frozen tundra W. of Hudson Bay)

City: unknown

Country: Canada,

Operator: Bell Canada, Sympatico

Issues with IP look-up data:

is it actually revealing, or is it opaque

is the data even current, or is it out-of-date

was the data ever accurate in the first place

Objectives

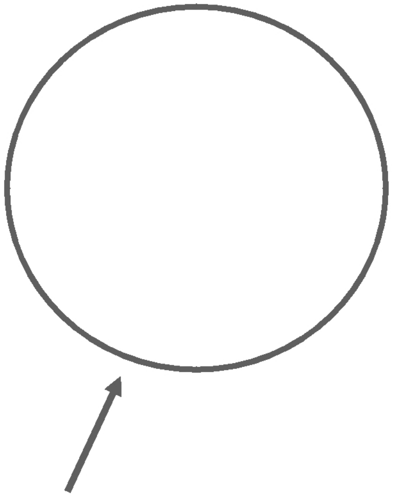
Develop new analytics to provide richer contextual data about a network address

Apply analytics against Tipping & Cueing objectives

Build upon artefact of techniques to develop new needle-in-a-haystack analytic – contact chaining across air-gaps

Analytic Concept – Start with Travel Node

Begin with *single* seed Wi-Fi IP address of intl. airport



Assemble set of user IDs seen on network address over two weeks

Profiling Travel Nodes – Next Step

Follow IDs backward and forward in recent time

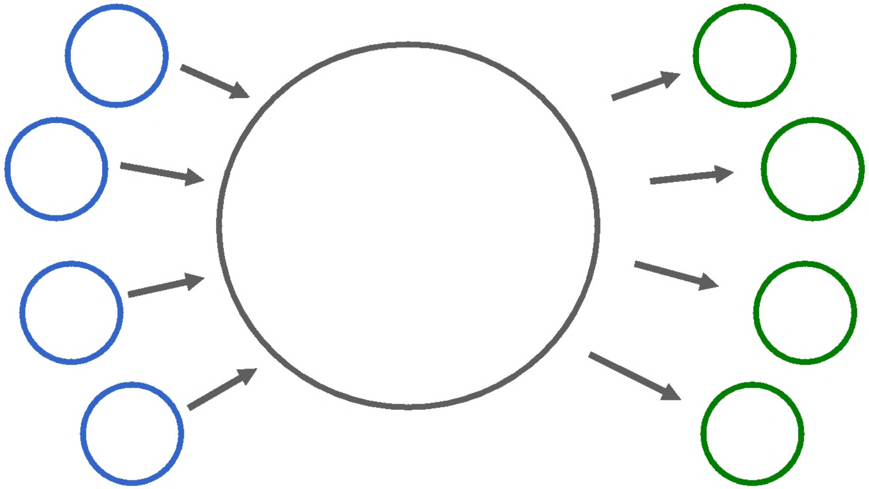
Earlier IP clusters of:

- local hotels
- domestic airports
- local transportation hubs
- local internet cafes

- etc.

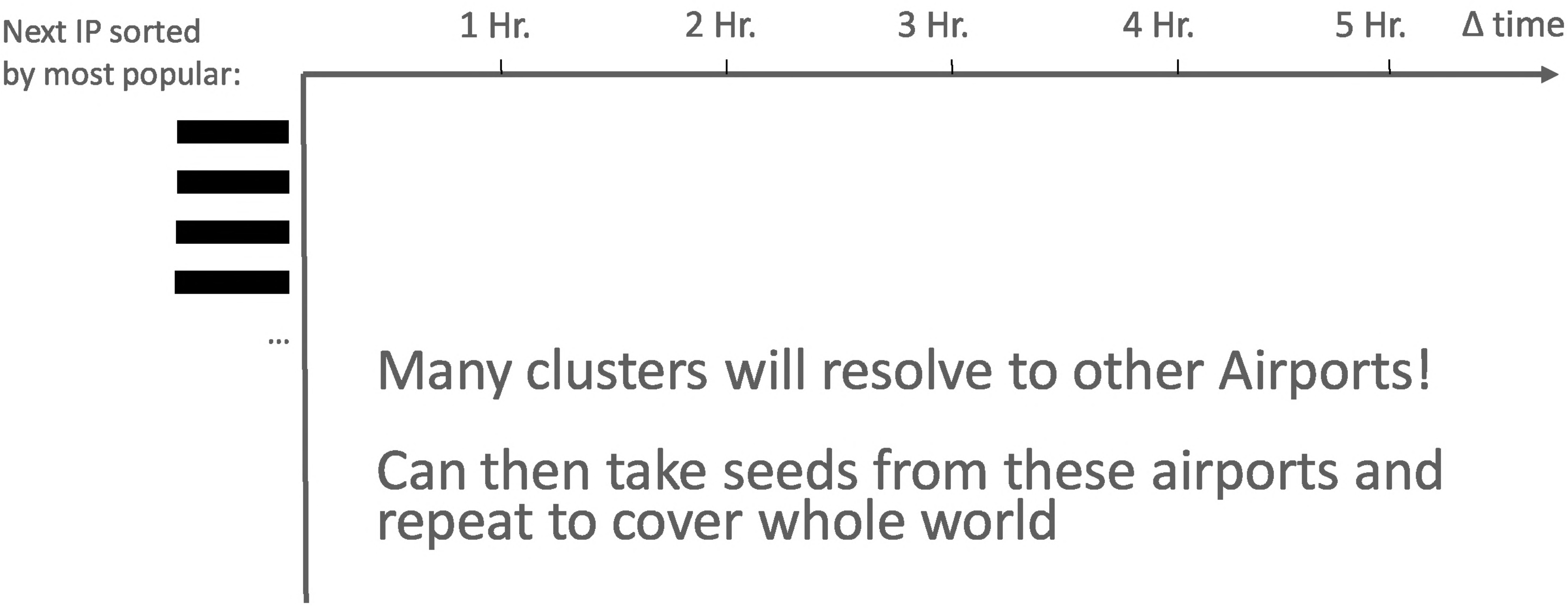
Later IP clusters of:

- other intl. airports
- domestic airports
- major intl. hotels
- etc.



IP Hopping Forward in Time

Follow IDs forward in time to next IP & note delta time



[redacted]

Many clusters will resolve to other Airports!

Can then take seeds from these airports and repeat to cover whole world

Ditto for going backward in time, can uncover roaming infrastructure of host city: hotels, conference centers, Wi-Fi hotspots etc.

Data Reality

The analytic produced excellent profiles, but was more complex than initial concept suggests

Data had limited aperture – Canadian Special Source

major CDN ISPs team with US email majors, losing travel coverage

Behaviour at airports

little lingering on arrival; arrivals using phones, not WiFi

still, some Wi-Fi use when waiting for connecting flight/baggage

different terminals: domestic/international; also private lounges

Very many airports and hotels served by large Boingo private network

not seen in aperture; traffic seems to return via local Akamai node

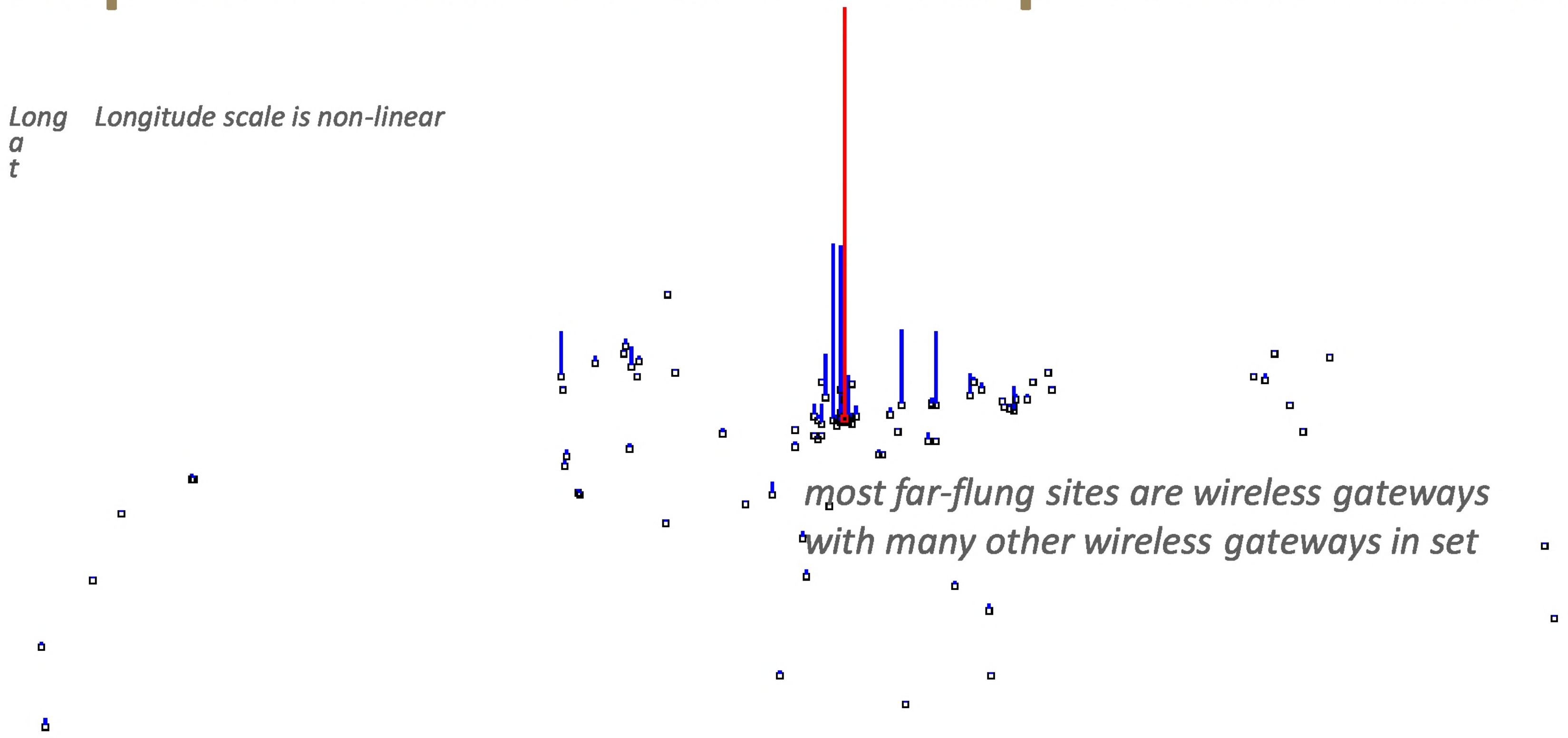
Tradecraft Development Data Set

Have two weeks worth of ID-IP data from Canadian Special Source – [redacted]

Had program access to Quova dataset connecting into Atlas database

Had seed knowledge of a single Canadian Airport WiFi IP address

Hop Geo Profile From CDN Airport Intl. Terminal



Longitude scale is non-linear

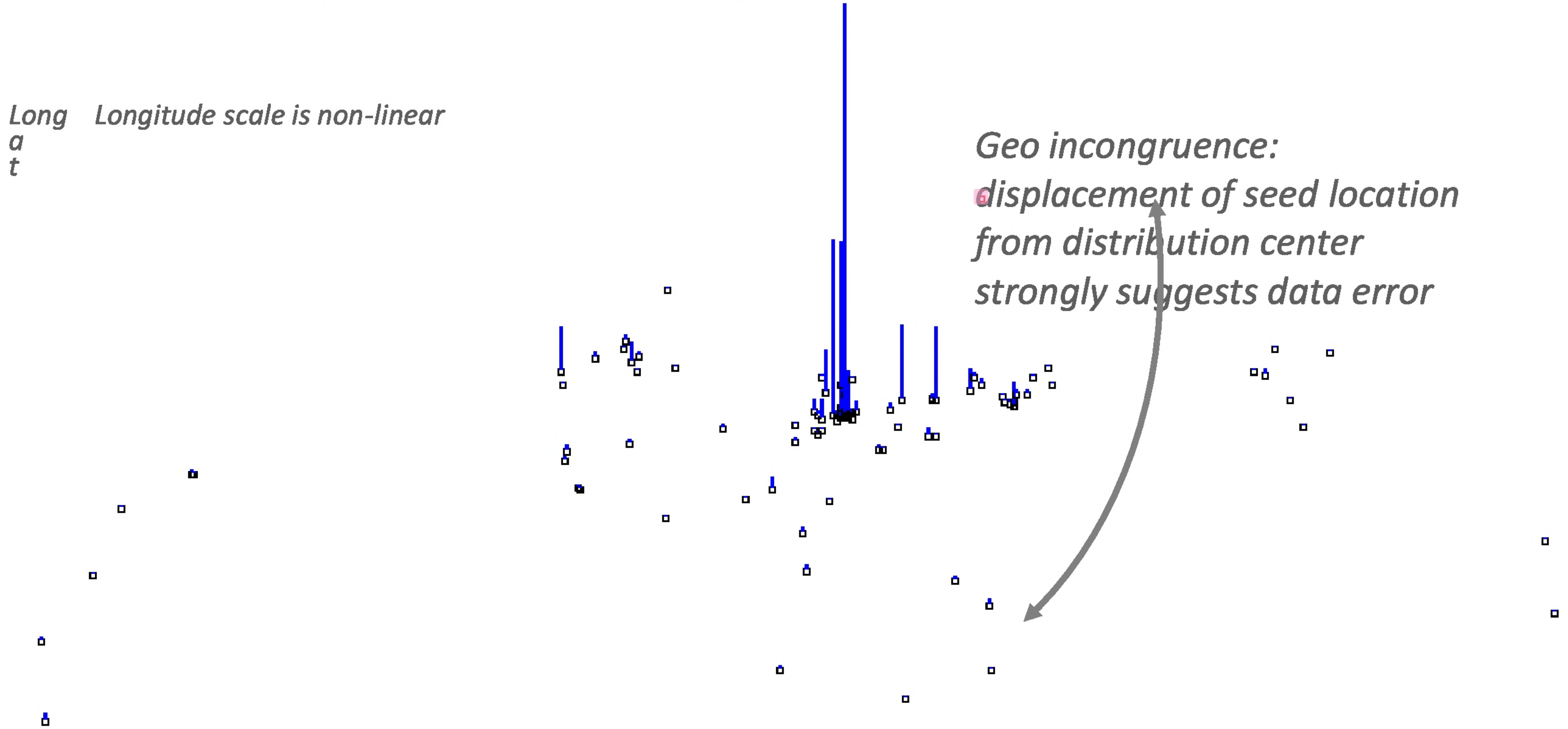
most far-flung sites are wireless gateways with many other wireless gateways in set

Profiled/seed IP location: Square = geographic location

Hopped-to IP location: Line height = numbers of unique hopped-to IPs at location

Plot of where else IDs seen at seed IP have been seen in two weeks
Plot shows most hopped to IPs are nearby - confirming reported seed geo data

Effect of Invalid Geo Information



Longitude scale is non-linear

Geo incongruence: displacement of seed location from distribution center strongly suggests data error

Profiled/seed IP location: □ Square = geographic location

Hopped-to IP location: | Line height = numbers of unique hopped-to IPs at location

Effect of invalid seed geo information readily apparent

Hop-Out Destinations Seen

Other domestic airports

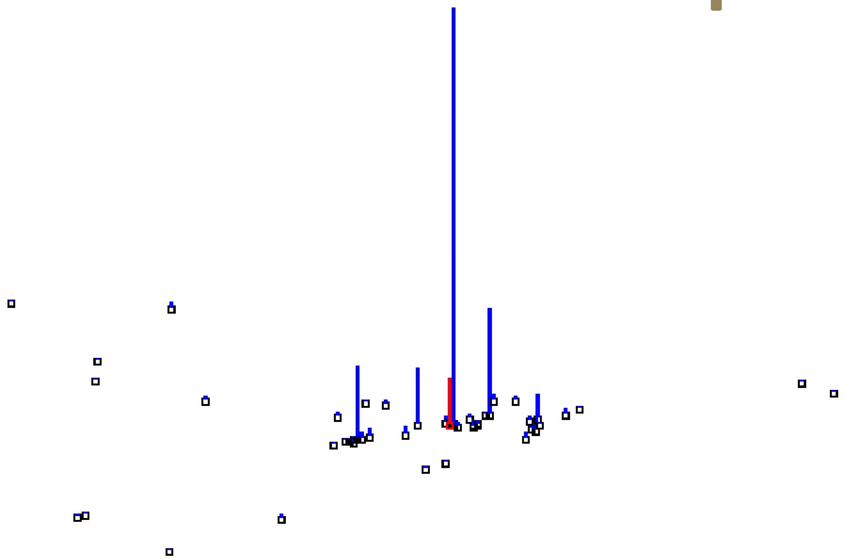
Other terminals, lounges, transport hubs

Hotels in many cities

Mobile gateways in many cities

Etc.

"Discovered" Other CDN Airport IP

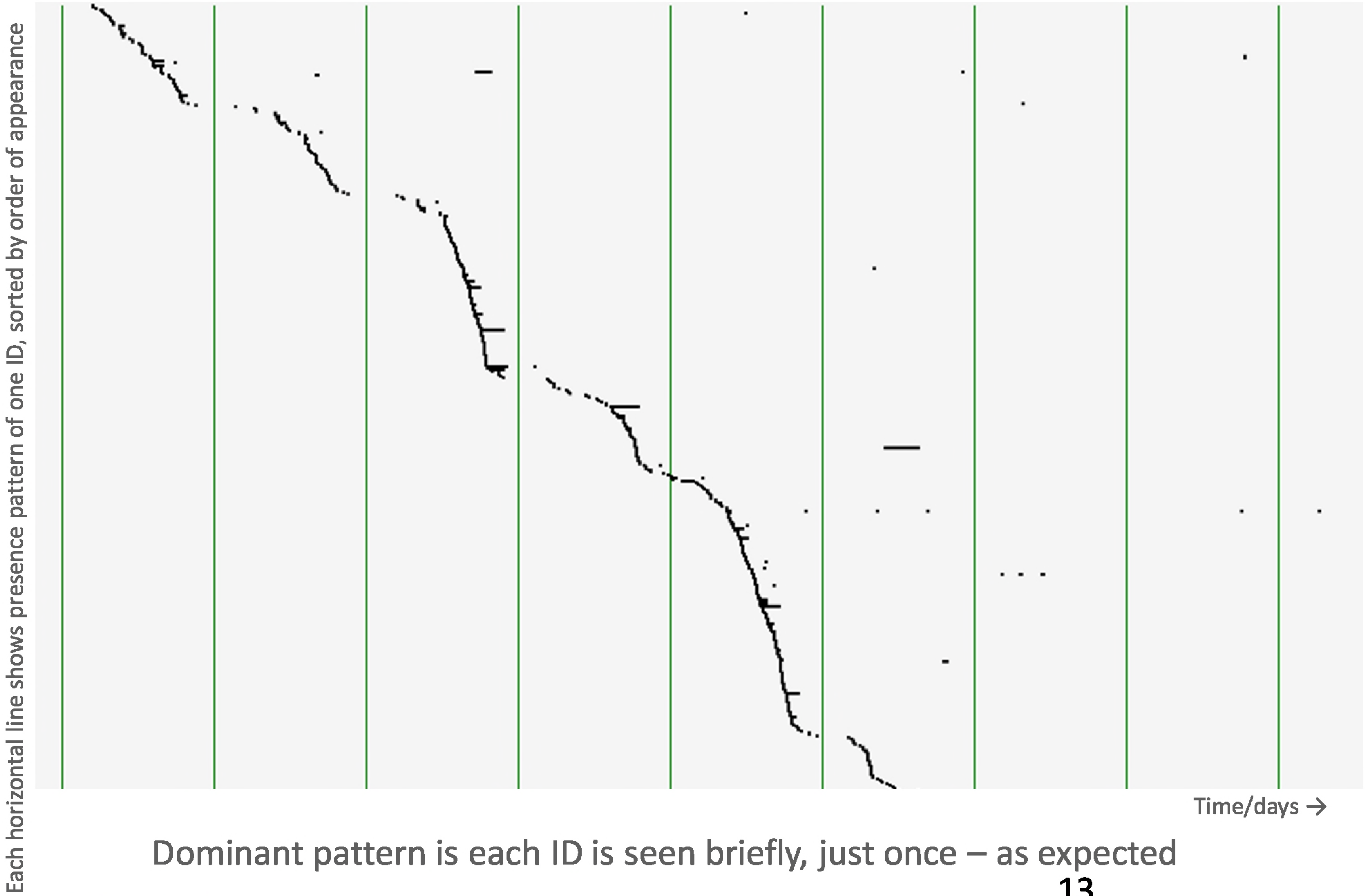


Domestic terminal

Closeness of majority of hopped-to IPs confirms geo data

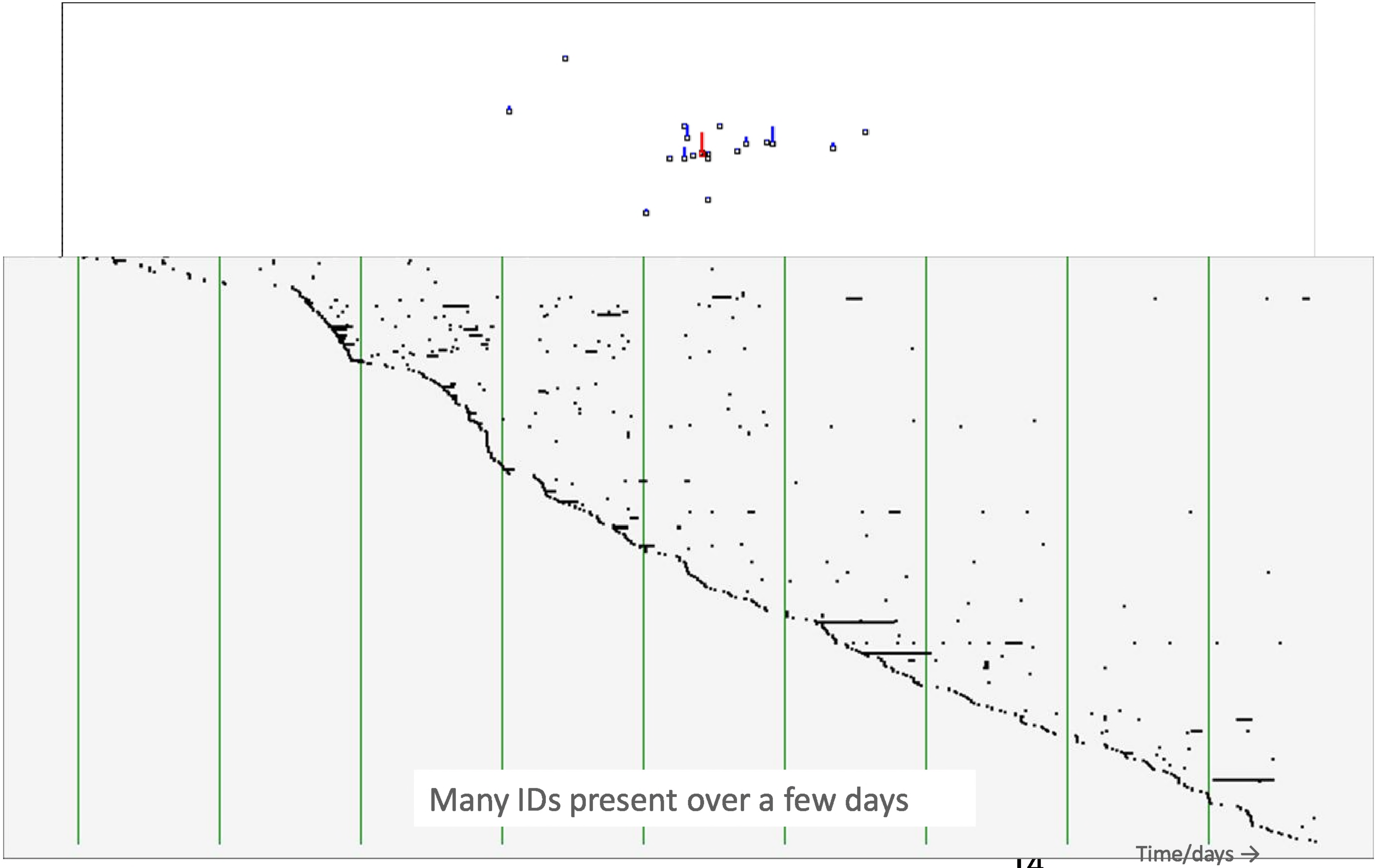
But, domestic airport can also look like a busy hotel ...

IDs Presence Profile at "Discovered" Airport

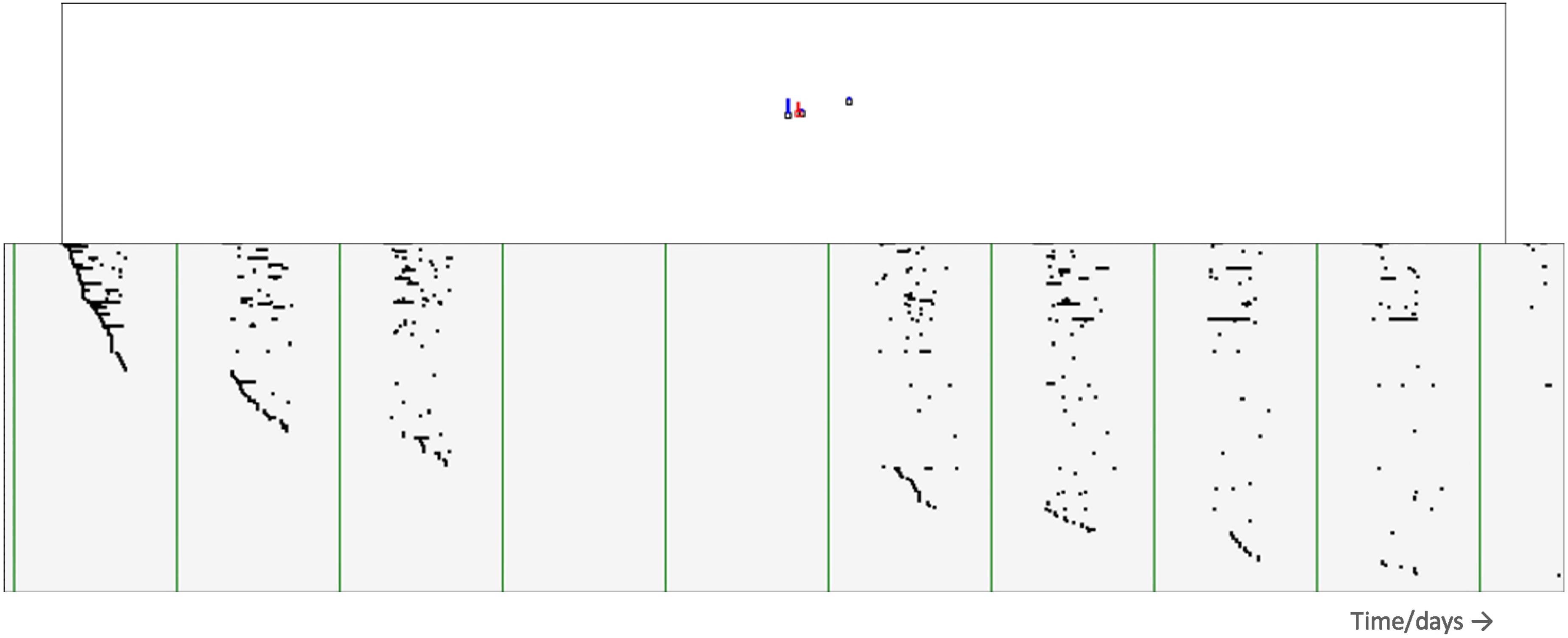


Dominant pattern is each ID is seen briefly, just once – as expected

Profiles of Discovered Hotel

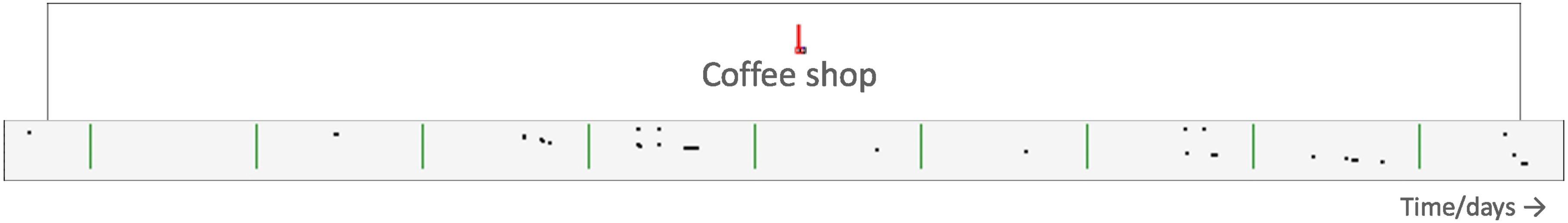

Many IDs present over a few days

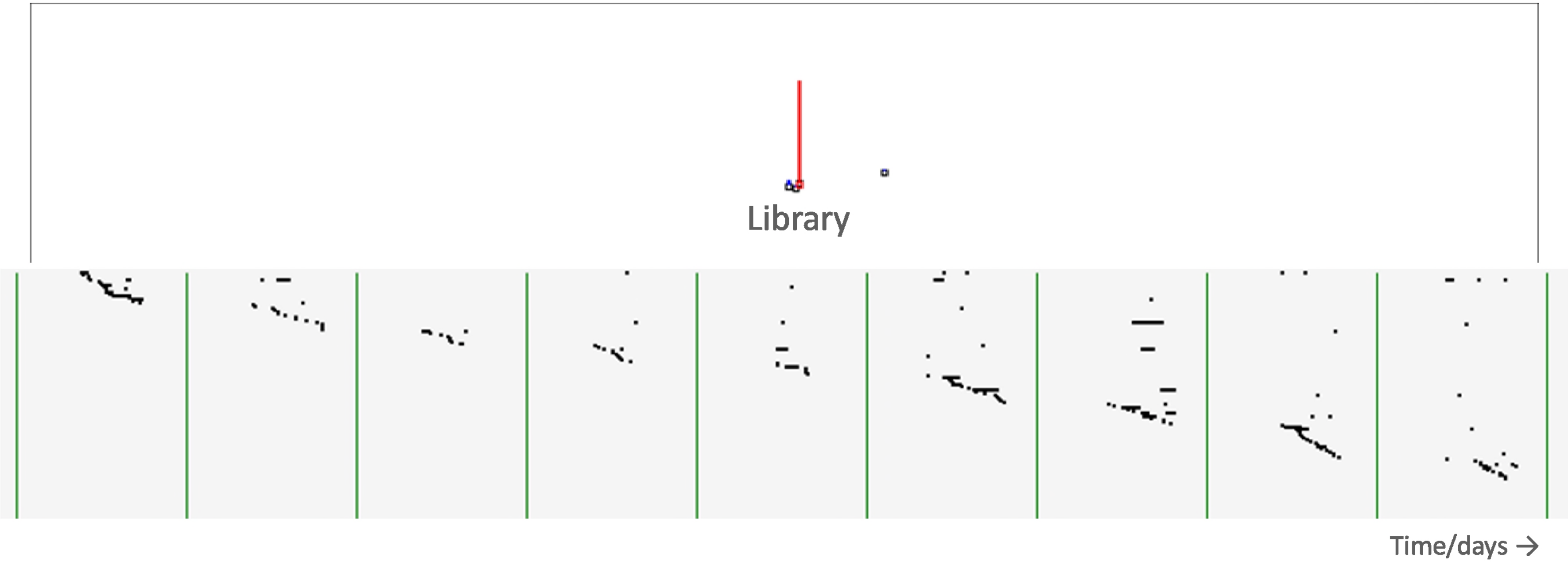
Profiles of Discovered Enterprise



Regular temporal presence (M-F) with local geographic span
Contrasts well against travel/roaming nodes

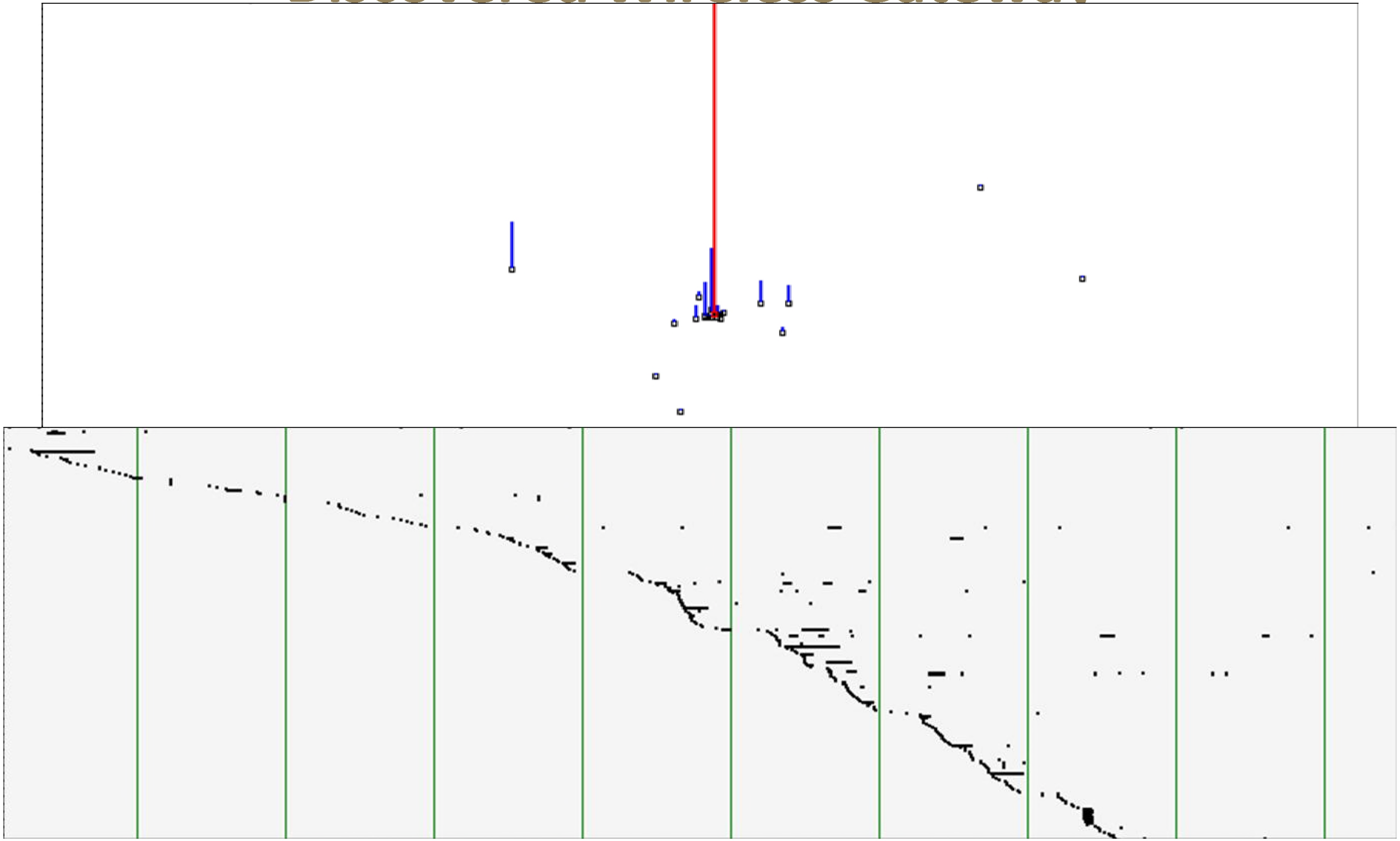
Discovered Coffee Shop, Library





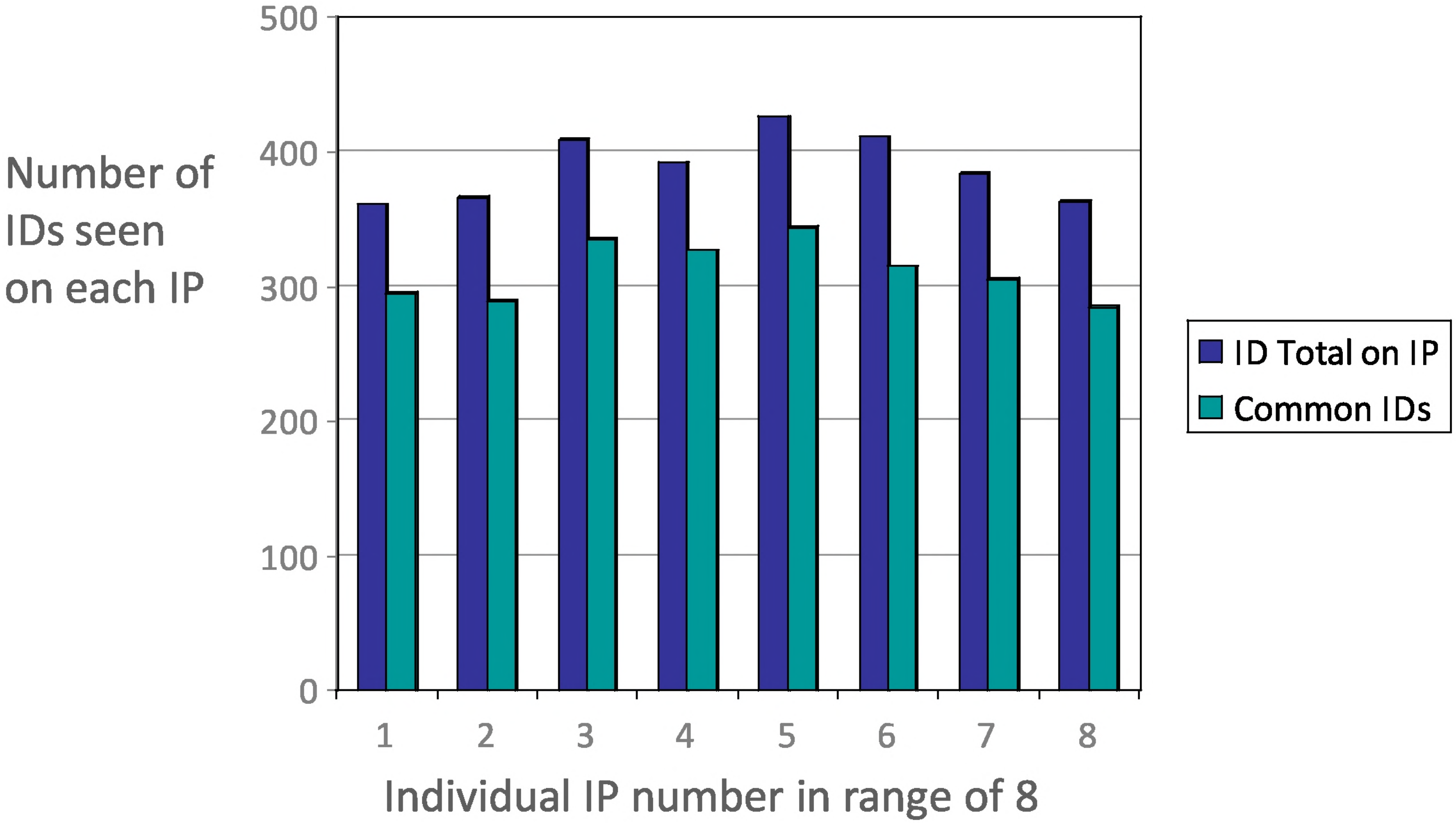
Similar patterns of mixed temporal & local geographic presence

Discovered Wireless Gateway



Wireless gateway not unlike a hotel, except ...

Partial Range Profile of Wireless Gateway



For wireless gateway, range behaviour is revealing
Most IDs seen on an IP are also scattered across entire range
ID totals & traffic across full range is very high

Mission Impact of IP Profiling

Tipping and Cueing Task Force (TCTF)

- a 5-Eyes effort to enable the SIGINT system to provide real-time alerts of events of interest
- alert to: target country location changes, webmail logins with time-limited cookies etc.

Targets/Enemies still target air travel and hotels

- airlines: shoe/underwear/printer bombs …
- hotels: Mumbai, Kabul, Jakarta, Amman, Islamabad, Egyptian Sinai …

Analytic can hop-sweep through IP address space to identify set of IP addresses for hotels and airports

- *detecting target presence within set will trigger an urgent alert*
- aim to productize analytics to reliably produce set of IPs for alerting

IP Profiling Summary

Different categories of IP ownership/use show distinct characteristics

- airports, hotels, coffee shops, enterprises, wireless gateways etc.
- clear characteristics enable formal modeling developments
- clear identification of hotels and airports enables critical Tipping & Cueing tradecraft

Geo-hop profile can confirm/refute IP geo look-up information

- later could fold-in time deltas for enhanced modeling

Can “sweep” a region/city for roaming access points to IP networks

- *leads to a new needle-in-a-haystack analytic ...*

Tradecraft Problem Statement

A kidnapper based in a rural area travels to an urban area to make ransom calls

- can't risk bringing attention to low-population rural area
- won't use phone for any other comms (or uses payphones ...)

Assumption: He has another device that accesses IP networks from public access points

- having a device isn't necessary, could use internet cafes, libraries etc.
- he is also assumed to use IP access around the time of ransom calls

Question: Knowing the time of the ransom calls can we discover the kidnapper's IP ID/device

- "contact chain" across air-gap (not a correlation of selectors)

Solution Outline

With earlier IP profiling analytics, we can “sweep” a city/region to discover and determine public accesses

We can then select which IP network IDs are seen as active in all times surrounding the known ransom calls

reduce set to a shortlist

Then we examine the reduced set of IP network IDs and eliminate baseline heavy users in the area that fall into the set intersection just because they are always active

that is, eliminate those that are highly active outside the times of the ransom calls

hopefully leaves only the one needle from the haystack

First Proof-of-Concept

Swept a modest size city and discovered two high traffic public access ranges with >300,000 active IDs over 2 weeks

used for initial expediency due to computational intensity

Presumed that there were 3 ransom calls, each 50 hours apart during daytime, looked for IDs within 1 Hr of calls

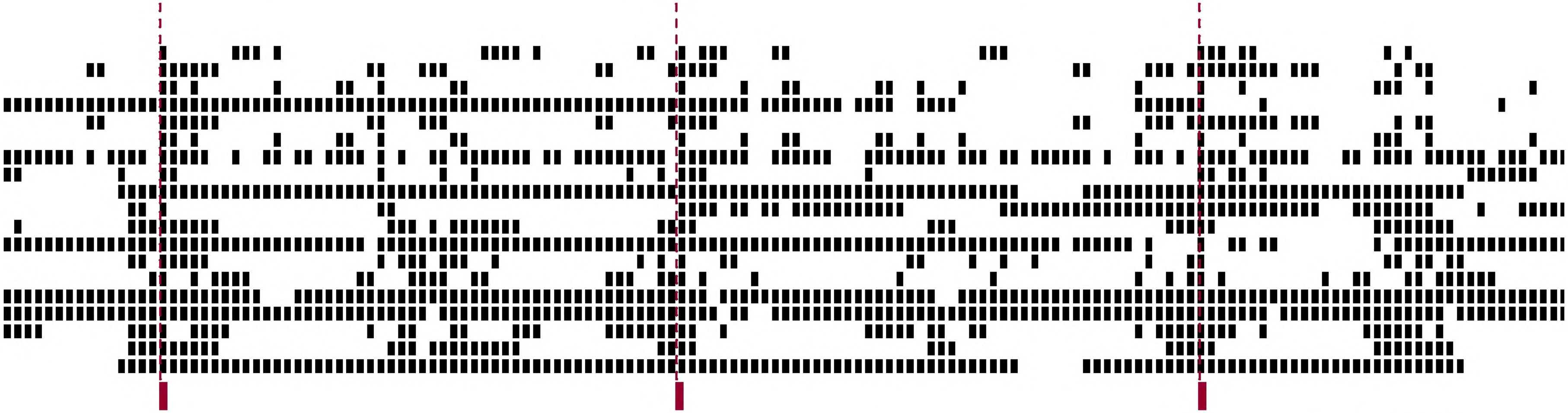
reduce large set to a shortlist of just 19 IP network IDs

Examined activity level of 19 IP network IDs – how many presences each had in 1 Hr slots over two weeks

main worry as the computation was running: there would be a lot of IDs that showed just a handful of appearances: e.g. 3, 4, 5 instances

ID Presence of Shortlist

Each horizontal line shows presence of ID over time/hour-slots



Time/hour-slots →

Postulated presence of kidnapper/target

Happy result: least active ID had appearances in 40 hour-slots!
Thus could eliminate all, leaving just the kidnapper (if he was there)

Big-Data Computational Challenge

All the previous analytics, while successful experimentally, ran much too slowly to allow for practical productization

CARE: Collaborative Analytics Research Environment

- a big-data system being trialed at CSEC (with NSA launch assist)
- non-extraordinary hardware
- minimal impedance between memory, storage and processors
- highly optimized, in-memory database capabilities
- columnar storage, high performance vector functional runtime
- powerful but challenging programming language (derived from APL)

Result of first experiments with CARE: game-changing

- run-time for hop-profiles reduced from 2+ Hrs to several seconds
- *allows for tradecraft to be profitably productized*

Overall Summary

- IP profiling showing terrific value
  - significant analytic asset for IP networks and target mobility
  - enables critical capability within Tipping & Cueing Task force
  - working to productize on powerful new computational platform
  - broader SSO accesses/apertures coming online at CSEC
  - look to formalize models & fold-in timing deltas

- A new needle-in-a-haystack analytic is viable: contact chaining across air-gaps
  - enabled by sweep capability of IP profiling
  - should test further to understand robustness with respect to loosening assumptions of target behaviour
  - beyond kidnapping, tradecraft could also be used for any target that makes occasional forays into other cities/regions

[redacted] Tradecraft Studio Example

[redacted]

Possible route for productizing analytics