Network Heartbeat Traffic Characterization

Mackenzie Haffey
Martin Arlitt
Carey Williamson
Department of Computer Science
University of Calgary

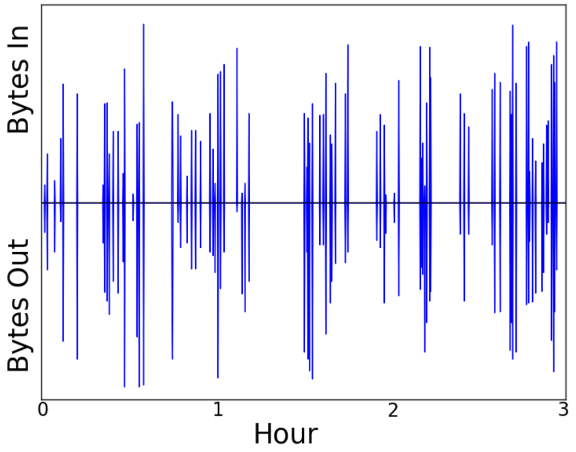
What is a Network Heartbeat?

- An event that occurs repeatedly at fairly regular intervals within a particular observational time frame
- In our case, the event is a connection initiated between two specific transport-level endpoints on a network (i.e., periodic network communications)

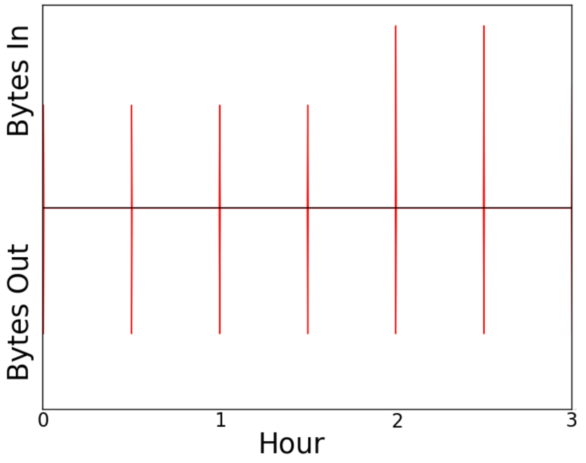
- Some heartbeats are <u>regular</u> (e.g., NTP)
- Some heartbeats are <u>irregular</u>, since they can be disrupted by user behaviour, NAT/DHCP, network outages, premature termination, or non-deterministic effects

Network Traffic Examples

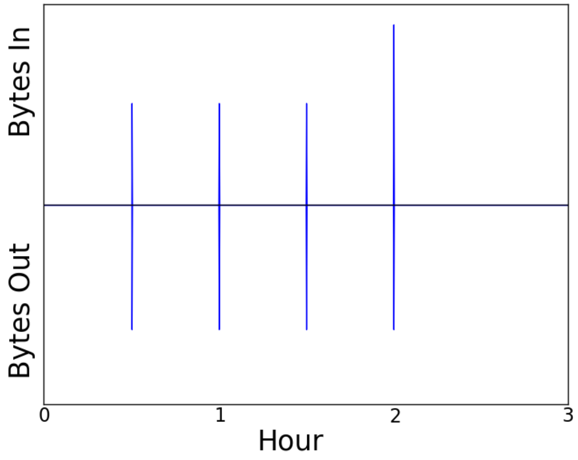
Random Traffic



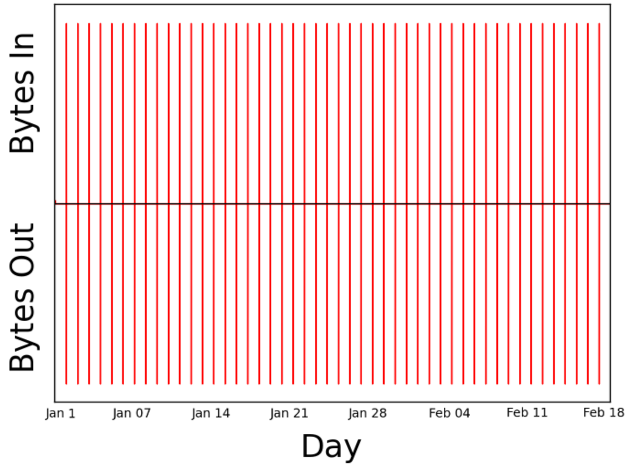
Regular Heartbeat



Irregular Heartbeat



NTP Heartbeat



Why Do We Care?

- Network heartbeats can be useful as an indicator of the operational health of an edge network:
  - Presence (or absence) of heartbeats for expected services
- Network heartbeats can indicate unexpected or undesired traffic on your network:
  - Peer-to-Peer (P2P) applications
  - Scanning
  - Malicious software (malware), such as botnets, which use periodic communications for command/control channels
- In general, there is a limited understanding of the use and characteristics of heartbeats in real networks, and how to leverage heartbeat information effectively

Background and Related Work

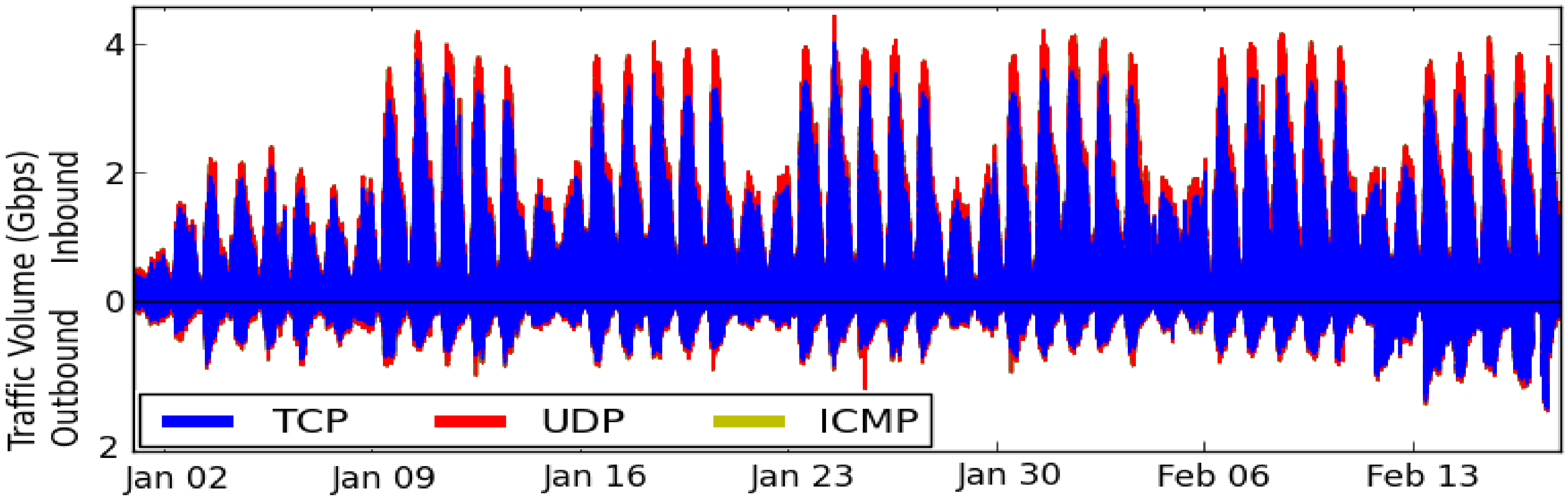
- Periodicity detection
  - -Statistical methods [Hubballi and Goyal 2013]
  - -Spectral methods [Assadhan et al. 2014] [Heard et al. 2014]
  - -Autocorrelation [Gu 2008] [Qiao 2013] [van Splunder 2015]
- Malware detection in Intrusion Detection System (IDS)
  - -Baywatch [Hu et al. 2016]
  - -Disclosure [Bilge et al. 2012]
  - -Stratosphere [Garcia 2015]
- Heartbeat identification [Bartlett 2011] [Heard 2014]

Talk Outline

- Introduction/Motivation/Related Work
- Our Campus Edge Network
- Heartbeat Detection Methodology
- Heartbeat Classification Taxonomy
- Heartbeat Characterization Study
- Discussion and Implications
- Conclusions

Our Campus Network

- University edge network with about 32,000 students and about 3,000 faculty and staff
- Includes both managed and unmanaged subnets
- Many unmanaged subnets are BYOD environment
- Strong diurnal usage pattern reflecting work days
- Peak inbound traffic near 4 Gbps; outbound 1 Gbps



Detecting Periodic Traffic (1 of 2)

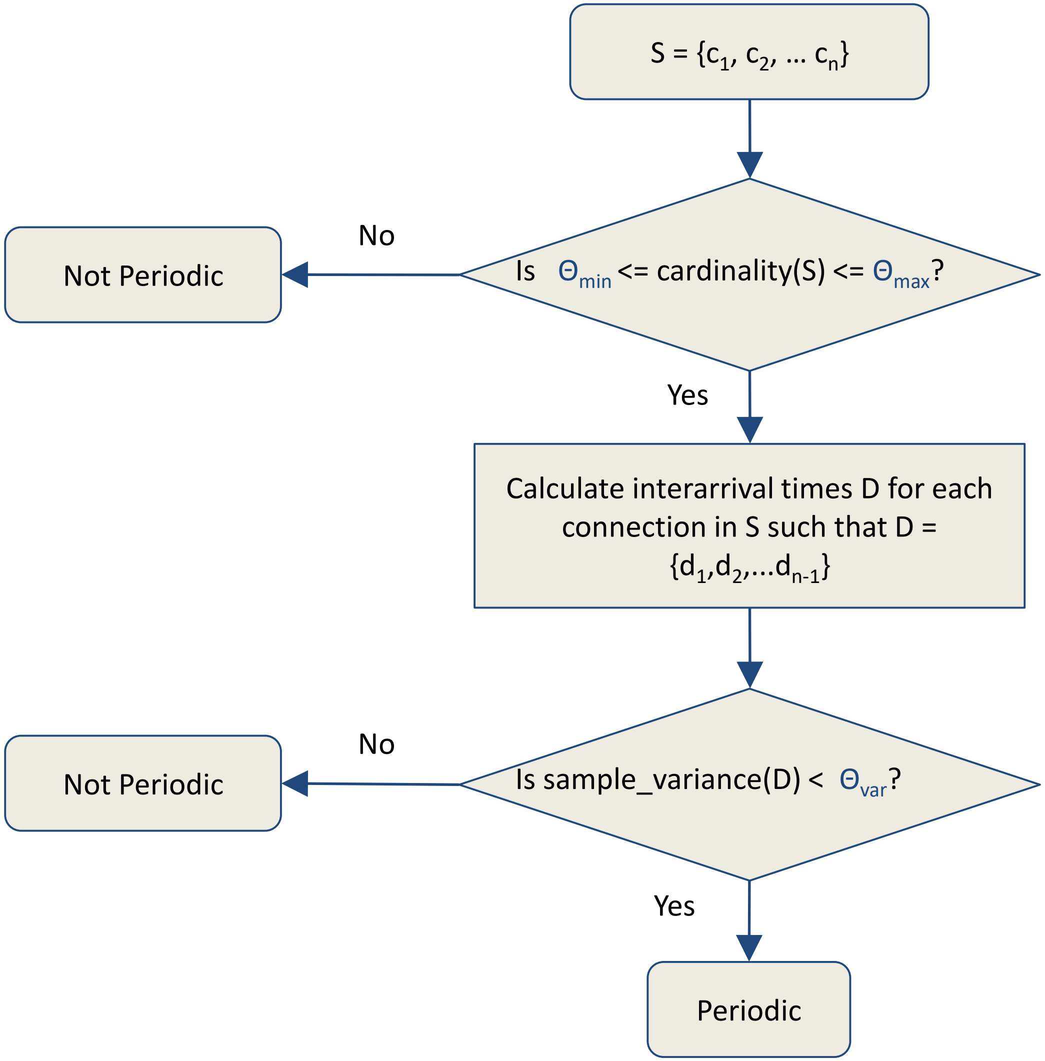
To detect periodicity, we consider connection 5-tuples:

- $c = (ts, h_s, h_r, dest_{port}, proto)$

- Construct "candidate connections sets" based on the same h_s, h_r, $dest_{port}$, and proto
- Prune candidate connection sets with too few or too many connections to manifest periodicity
- Compute inter-arrival times for connections in a set
- If the variance of inter-arrival times is below a specified threshold (Θ_{var}), then the candidate connection set is said to be periodic; otherwise, it is not periodic

Detecting Periodic Traffic (2 of 2)

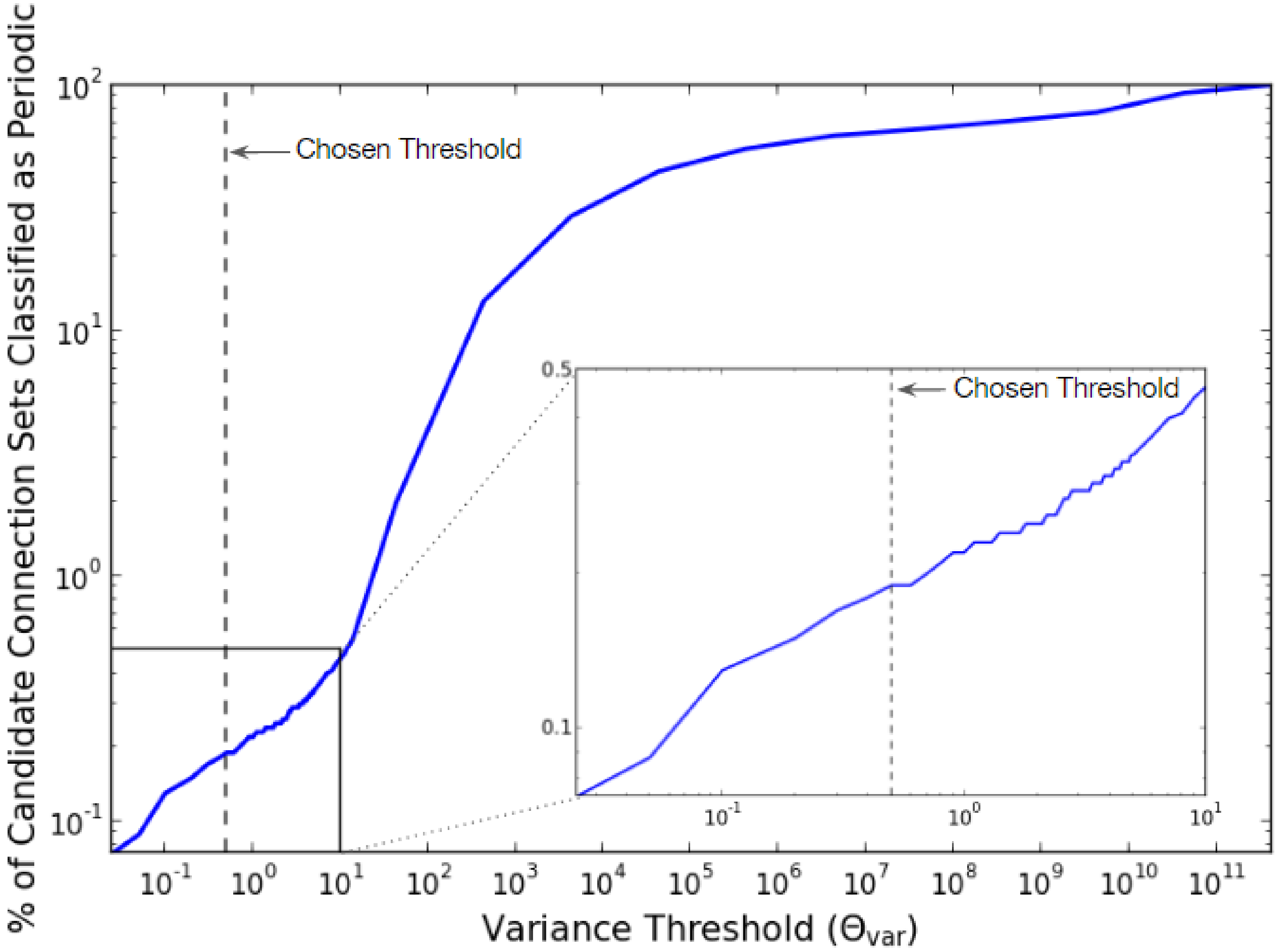
For every candidate connection set S:


$S = \{c_1, c_2, \ldots c_n\}$
Is Θ_{min} <= cardinality(S) <= Θ_{max}?
No
Not Periodic
Yes
Calculate interarrival times D for each connection in S such that D = $\{d_1, d_2, \ldots d_{n-1}\}$
Is sample_variance(D) < Θ_{var}?
No
Not Periodic
Yes
Periodic

* All done in SQL

* We conduct this process on the whole log, daily logs, and hourly logs, and then merge results

Sensitivity Analysis



Data Collection

- Data collection from Jan. 1, 2017 - Feb 18, 2017
- Data was collected from a mirrored stream of all network traffic entering/leaving U of C campus
- Data was processed and stored in Bro logs in real time
  - Records all TCP, UDP, and ICMP traffic “connections”
  - 15 billion connections during our observational period
  - 3.5 TB worth of data

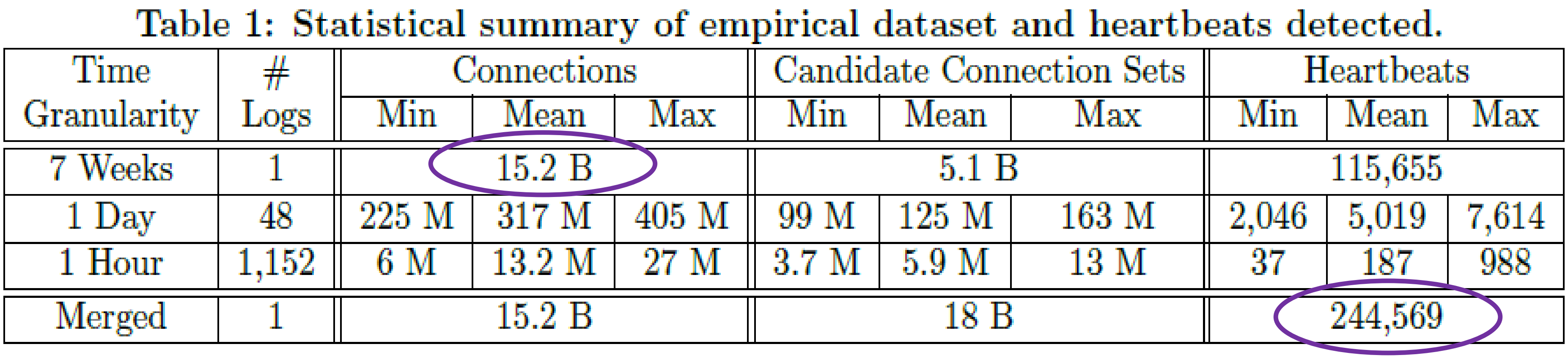
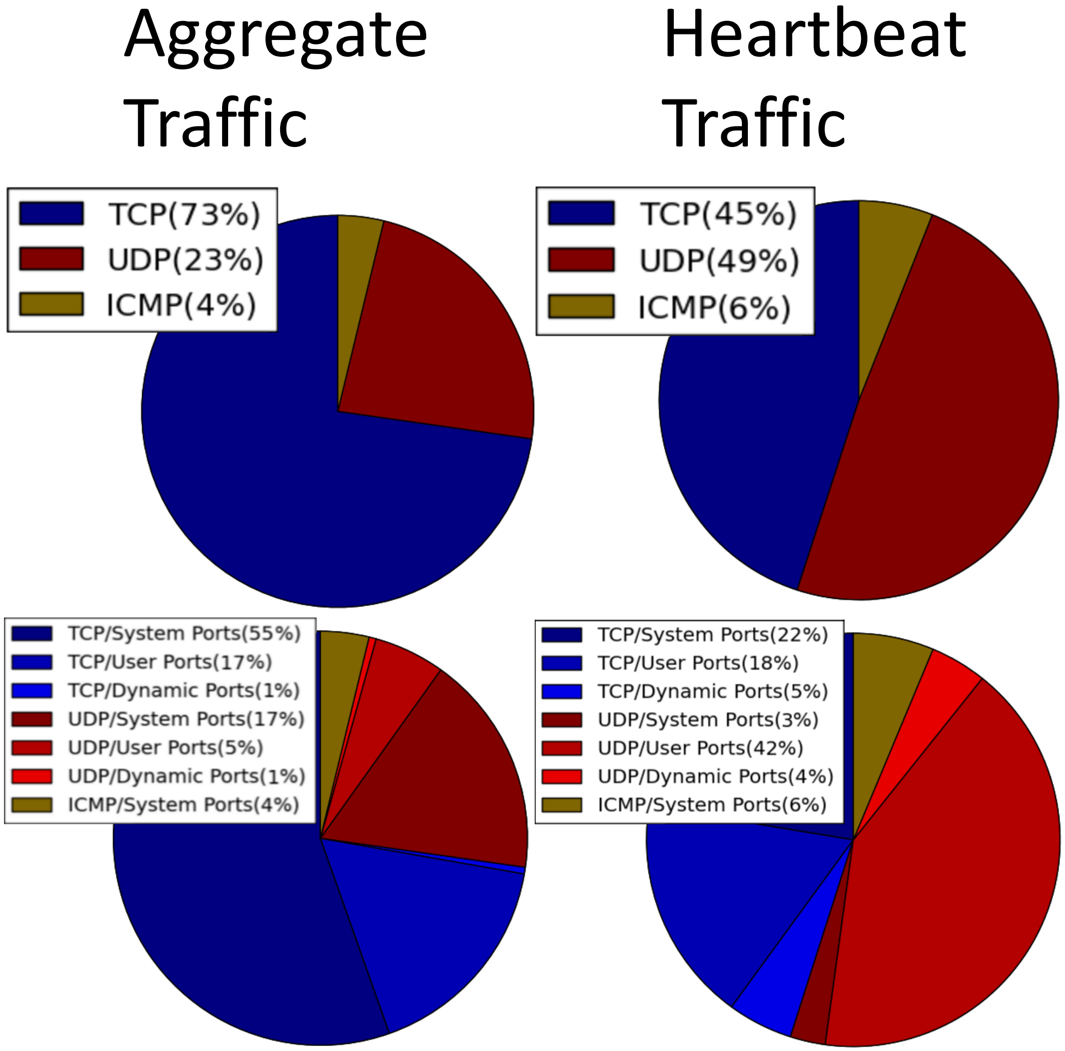


Table 1: Statistical summary of empirical dataset and heartbeats detected.

Time Granularity	# Logs	Connections			Candidate Connection Sets			Heartbeats		
		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
7 Weeks	1	15.2 B			5.1 B			115,655		
1 Day	48	225 M	317 M	405 M	99 M	125 M	163 M	2,046	5,019	7,614
1 Hour	1,152	6 M	13.2 M	27 M	3.7 M	5.9 M	13 M	37	187	988
Merged	1	15.2 B			18 B			244,569		

Results Overview

- Composition of heartbeat traffic differs a lot from aggregate traffic
- More UDP and User/Dynamic ports due to CDN, P2P, and botnets
- Most periodic ICMP traffic is scanning related

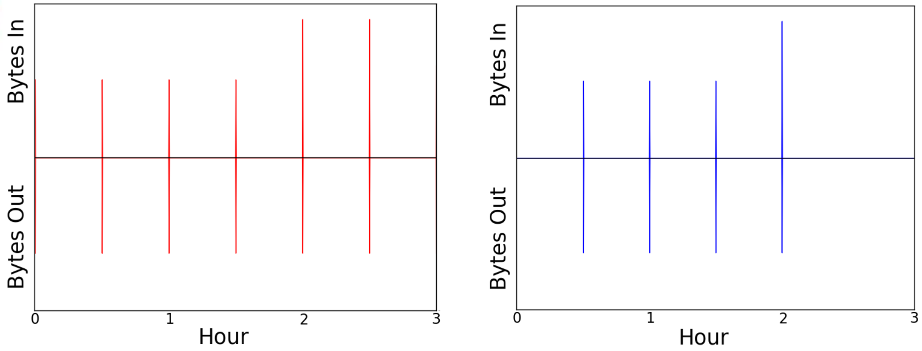

Aggregate Traffic

Heartbeat Traffic

Heartbeat Classification Attributes

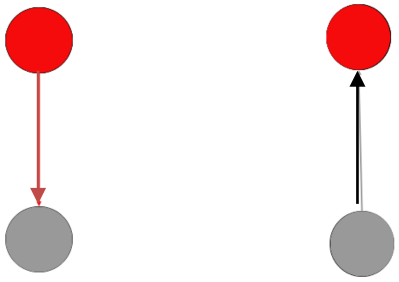
Heartbeat Regularity

Regular vs. Irregular



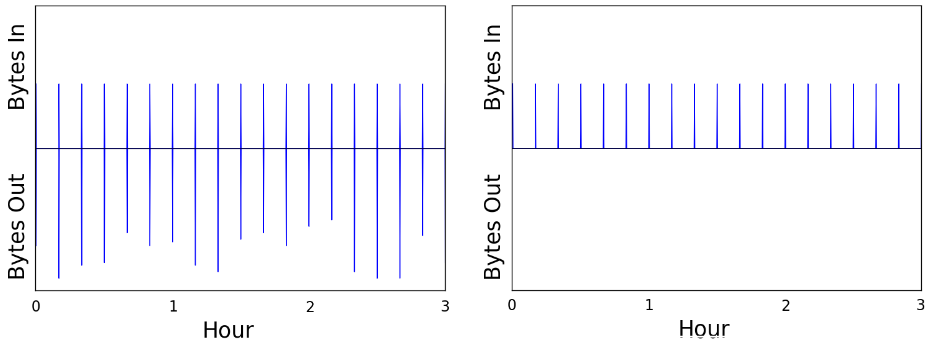
Heartbeat Direction

Inbound vs. Outbound



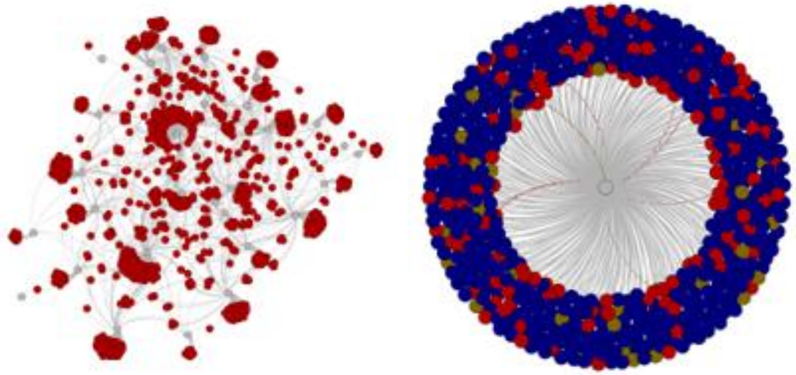
Heartbeat Liveness

Alive vs. Dead

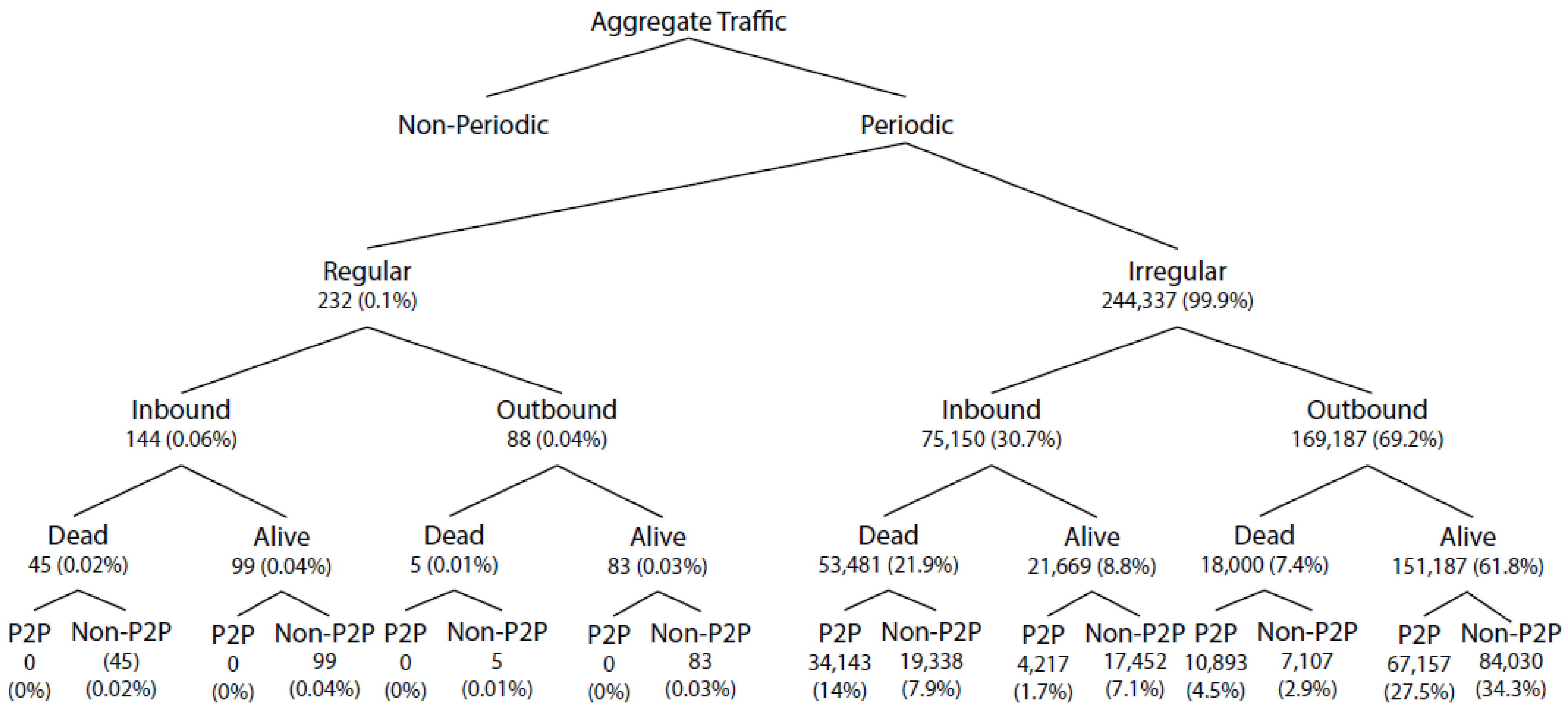


Application Architecture

P2P vs. Non-P2P



Heartbeat Classification Taxonomy



Aggregate Traffic

Non-Periodic

Periodic

Regular
232 (0.1%)

Irregular
244,337 (99.9%)

Inbound
144 (0.06%)

Outbound
88 (0.04%)

Inbound
75,150 (30.7%)

Outbound
169,187 (69.2%)

Dead
45 (0.02%)

Alive
99 (0.04%)

Dead
5 (0.01%)

Alive
83 (0.03%)

Dead
53,481 (21.9%)

Alive
21,669 (8.8%)

Dead
18,000 (7.4%)

Alive
151,187 (61.8%)

P2P
0
(0%)

Non-P2P
(45)
(0.02%)

P2P
0
(0%)

Non-P2P
99
(0.04%)

P2P
0
(0%)

Non-P2P
5
(0.01%)

P2P
0
(0%)

Non-P2P
83
(0.03%)

P2P
34,143
(14%)

Non-P2P
19,338
(7.9%)

P2P
4,217
(1.7%)

Non-P2P
17,452
(7.1%)

P2P
10,893
(4.5%)

Non-P2P
7,107
(2.9%)

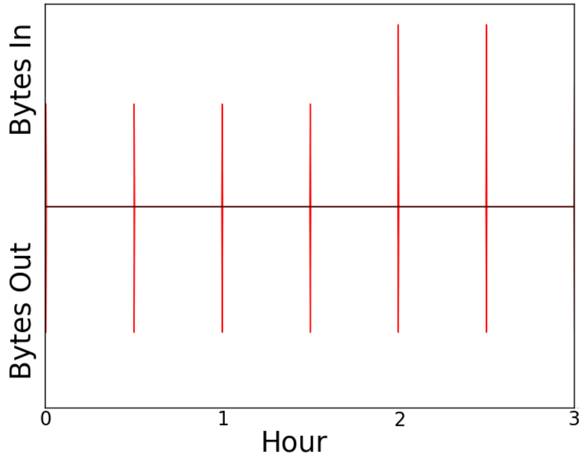
P2P
67,157
(27.5%)

Non-P2P
84,030
(34.3%)

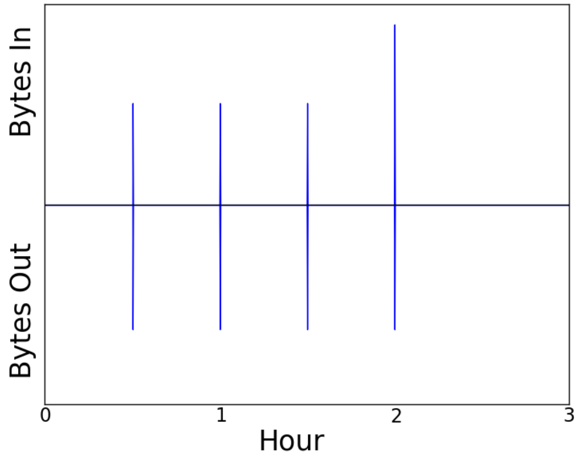
Heartbeat Regularity

- Regular heartbeats are persistent and continuous, occurring at regular intervals throughout the entire duration of the observation
- Regular
  - Regular heartbeats are intuitive, but make up less than 0.01% of heartbeats
  - Typically daily or weekly patterns
  - Primarily on managed portions
  - Primarily related to well-known protocols: NTP, HTTP, and DNS
- Irregular
  - Should be considered normal too!
  - Irregularity from DHCP churn, NAT, powering off, BYOD environment

Regular



Irregular



Heartbeat Direction

- Heartbeats can be inbound or outbound

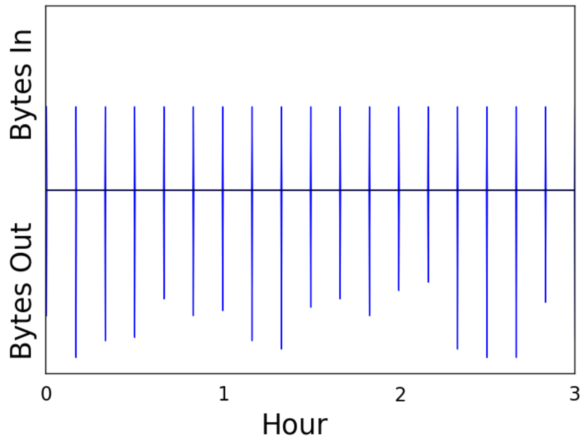
- Inbound heartbeats
  - Originate from outside our campus edge network
  - University-hosted services (e.g., Linux OS mirror site)
  - Periodic scanning (some malicious, some benign)
  - Services interacting with users on our network
  - Some P2P and CDN-related traffic

- Outbound heartbeats
  - Originate from within our campus edge network
  - Primarily generated by users interacting with services
  - Other significant contributors were CDN node and P2P

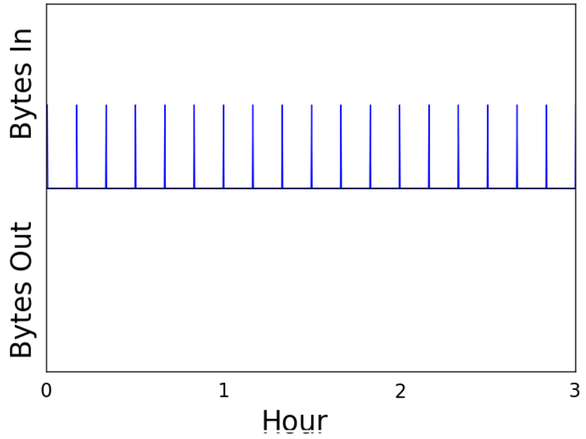
Heartbeat Liveness

- Heartbeats can be alive or dead
- Alive
  - Heartbeats that elicit a response from the recipient
  - These make up the majority of heartbeats
  - Usage pattern is similar to overall periodic
  - traffic pattern
  - A larger proportion of outbound heartbeats were alive than inbound heartbeats
- Dead
  - Heartbeats that do not elicit a response from the recipient
  - Surprisingly large number of heartbeats were dead (29%)
  - Scanning for hosts and services
  - Service vendors attempting to talk to hosts on our network
  - For the P2P traffic, this is likely caused by churn

Alive



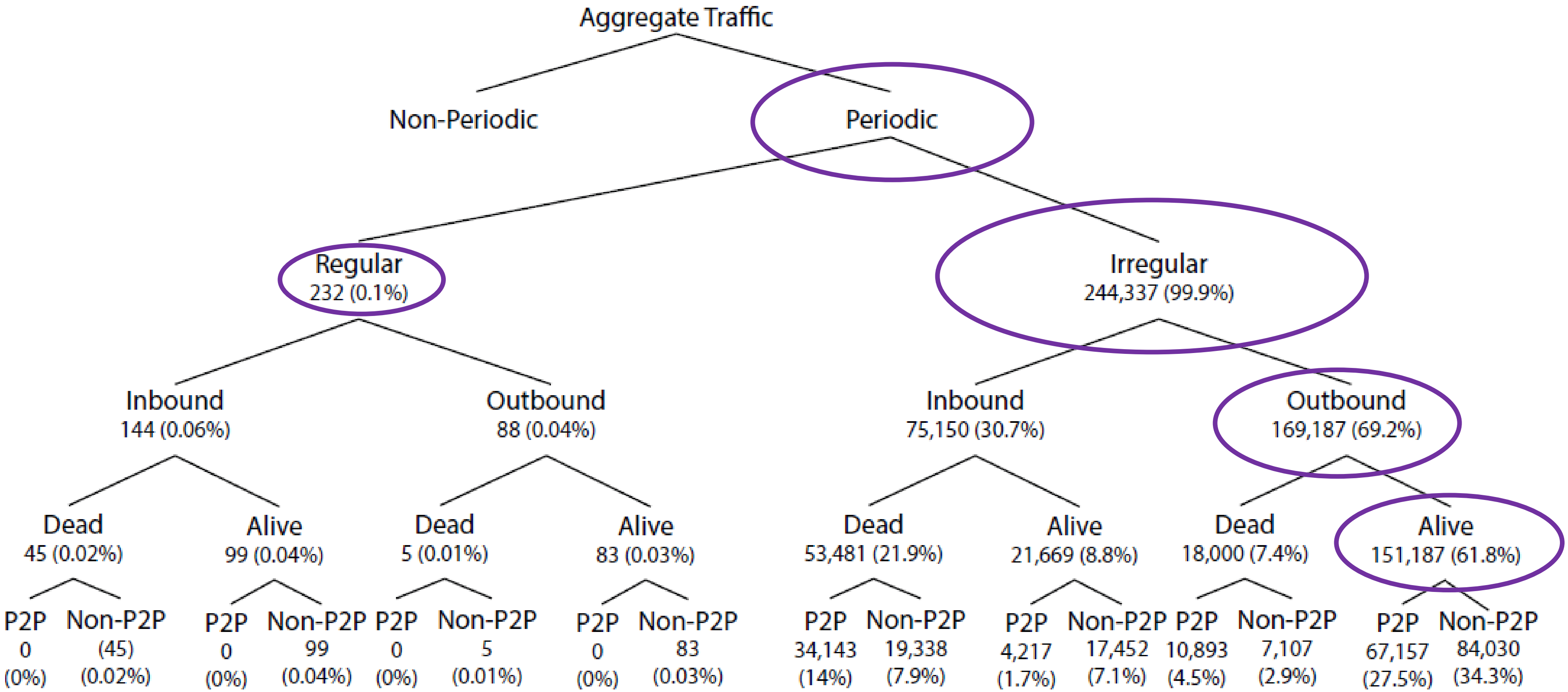
Dead



Peer-to-Peer (P2P) Heartbeats

- P2P heartbeats make up a large proportion of all the heartbeats observed (48%)
- These applications included BitTorrent, PPStream, ZeroAccess botnet, and Sality botnet
- P2P
  - Make up the most of the heartbeat traffic observed
  - Each peer sends periodic updates to other known peers, which generated a high number of heartbeats
  - Almost all done over UDP
  - Contributed greatly to the number of dead heartbeats, likely due to churn of P2P applications
- Non-P2P
  - Similar protocol/port usage to aggregate traffic: primarily TCP, concentrated in system port range
  - All regular heartbeats detected were from non-P2P apps

Heartbeat Classification Taxonomy



Aggregate Traffic

Non-Periodic

Periodic

Regular
232 (0.1%)

Irregular
244,337 (99.9%)

Inbound
144 (0.06%)

Outbound
88 (0.04%)

Inbound
75,150 (30.7%)

Outbound
169,187 (69.2%)

Dead
45 (0.02%)

Alive
99 (0.04%)

Dead
5 (0.01%)

Alive
83 (0.03%)

Dead
53,481 (21.9%)

Alive
21,669 (8.8%)

Dead
18,000 (7.4%)

Alive
151,187 (61.8%)

P2P
0
(0%)

Non-P2P
(45)
(0.02%)

P2P
0
(0%)

Non-P2P
99
(0.04%)

P2P
0
(0%)

Non-P2P
5
(0.01%)

P2P
0
(0%)

Non-P2P
83
(0.03%)

P2P
34,143
(14%)

Non-P2P
19,338
(7.9%)

P2P
4,217
(1.7%)

Non-P2P
17,452
(7.1%)

P2P
10,893
(4.5%)

Non-P2P
7,107
(2.9%)

P2P
67,157
(27.5%)

Non-P2P
84,030
(34.3%)

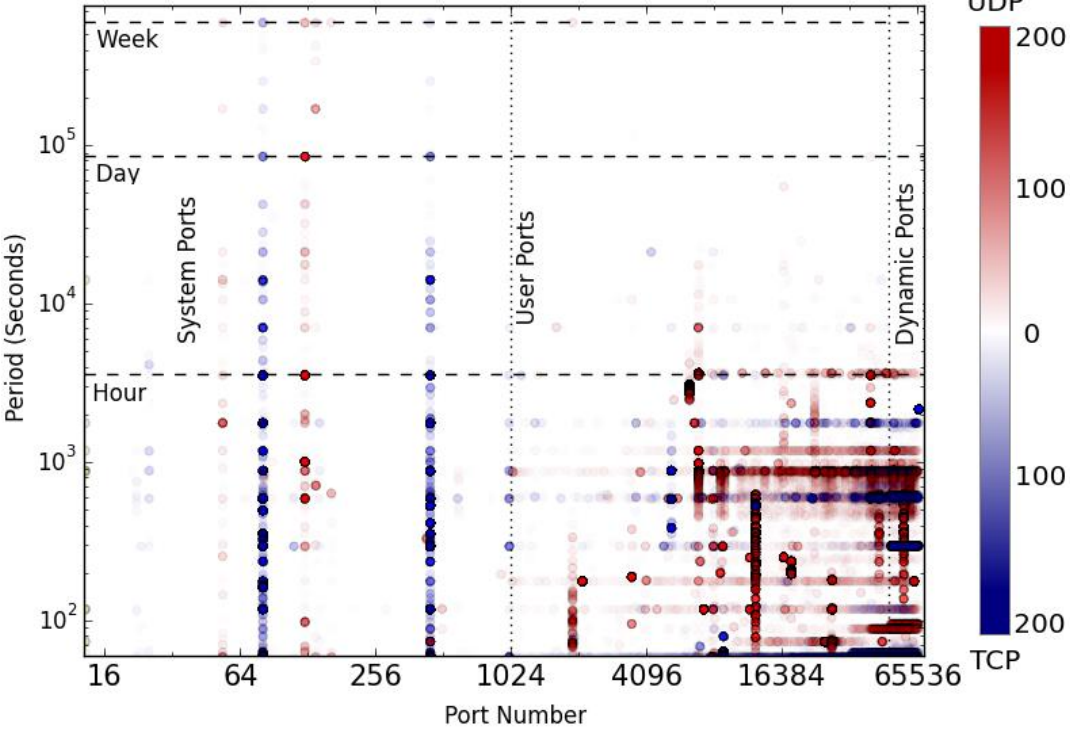
Heartbeat Characterization Study

- During our work, we identified several interesting characteristics of the heartbeat ecosystem:
  - **Structural Characteristics** - Characteristics related to the defining properties of a heartbeat – Period and Port
  - **Temporal Characteristics** - Characteristics related to the period and lifespan (longevity) of heartbeats
  - **Subnet Characteristics** - Characteristics related to how heartbeats manifest on different types of subnets
  - **Application Characteristics** - Characteristics related to how different application architectures, services, or vendors make use of heartbeats

Structural Analysis (1 of 2)

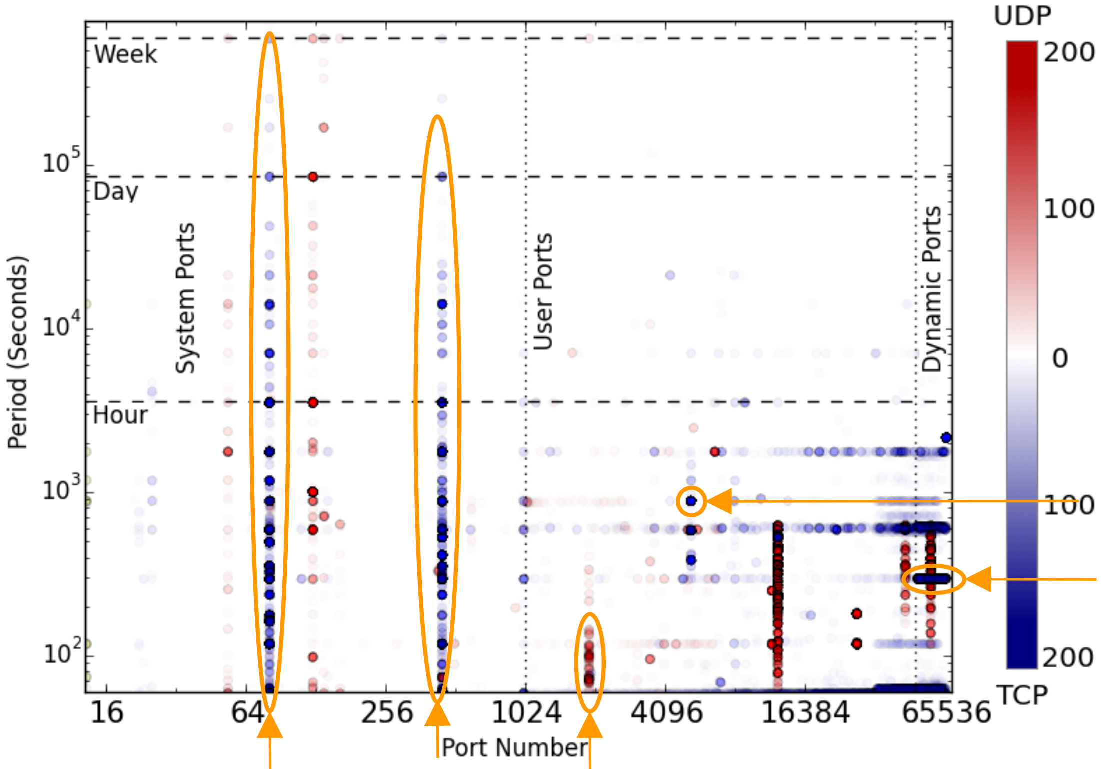
- Clustering patterns reveal points, horizontal bands, and vertical bands
- Prominent services or applications can be identified by analyzing clusters

Scatter Plot of All Heartbeats



Structural Analysis (2 of 2)

Scatter Plot of Non-P2P Heartbeats



Apple's Push Notification Service

Google Hangouts

HTTP

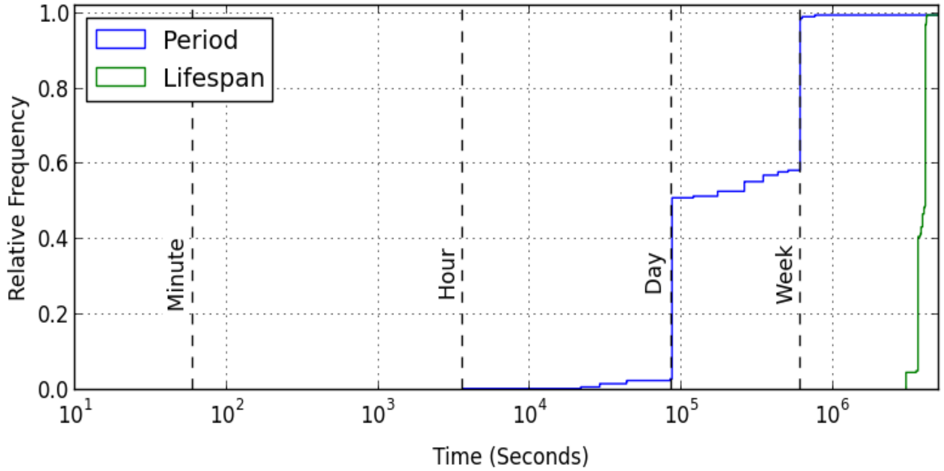
HTTPS

SSDP

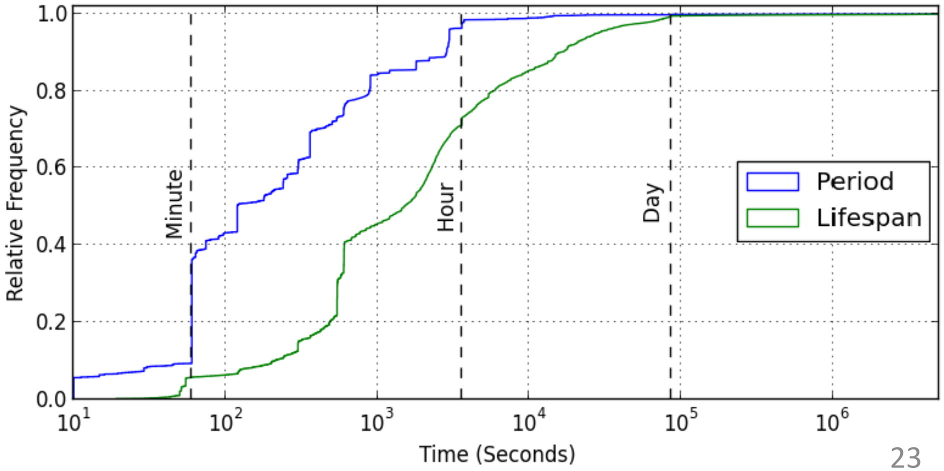
Temporal Analysis (1 of 2)

- Two pertinent properties:
  - **Period:** The time between successive connections
  - **Lifespan:** Elapsed time between first and last conn
- Regular heartbeats typically fall into very structured periods
- Irregular heartbeats are much less structured
- Irregular heartbeats typically have shorter periods and lifespans

Regular Heartbeats



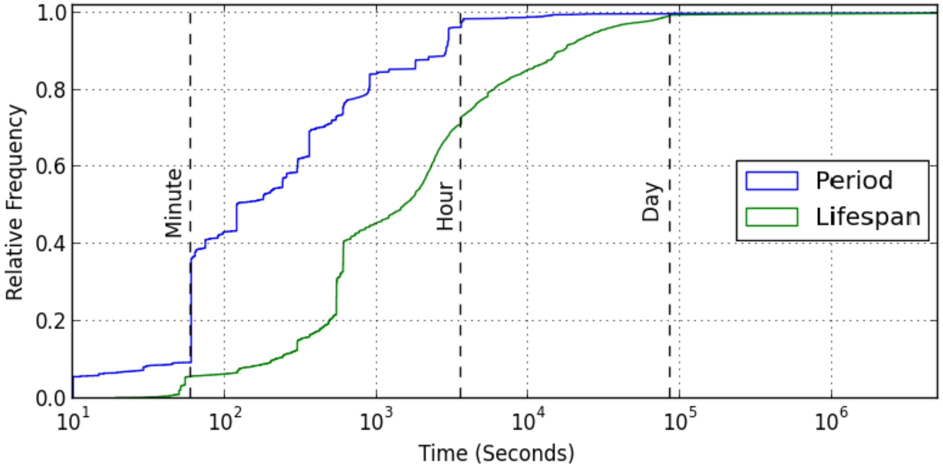
Irregular Heartbeats



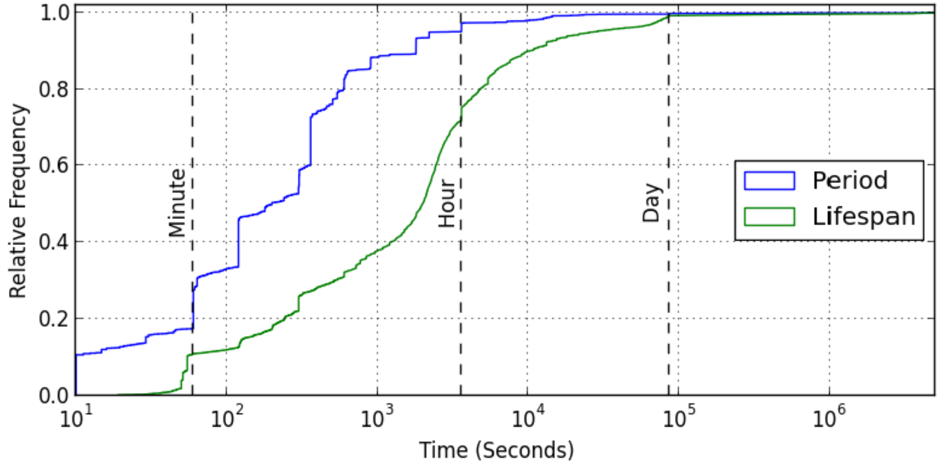
Temporal Analysis (2 of 2)

- For irregular heartbeats, the periods and lifespans tend to be relatively short
- Wide range – periods from 10 s to 8.8 days; lifespans from 30 s to 47.9 days
- Moderate positive correlation between period and lifespan (+0.73)

All Heartbeats

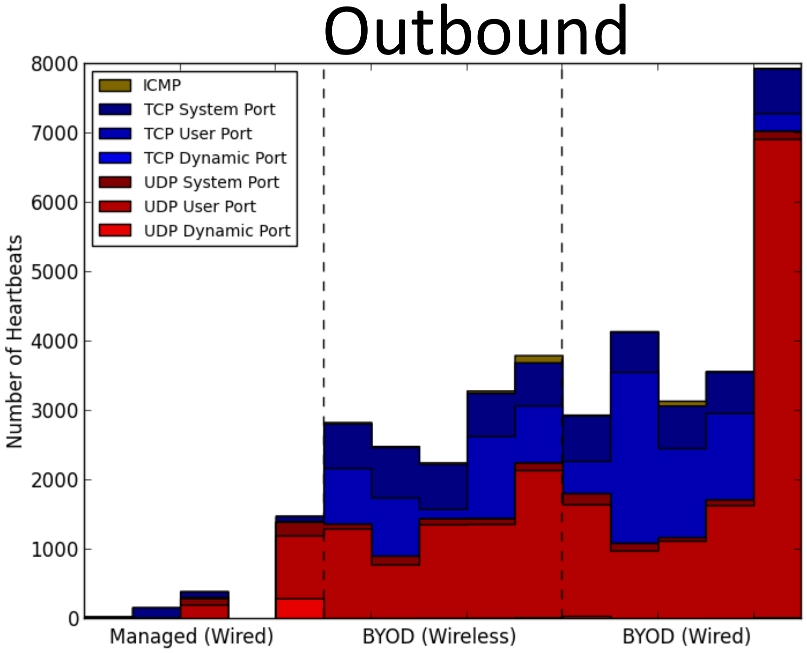


Non-P2P Heartbeats



Subnet Analysis – Outbound Heartbeats

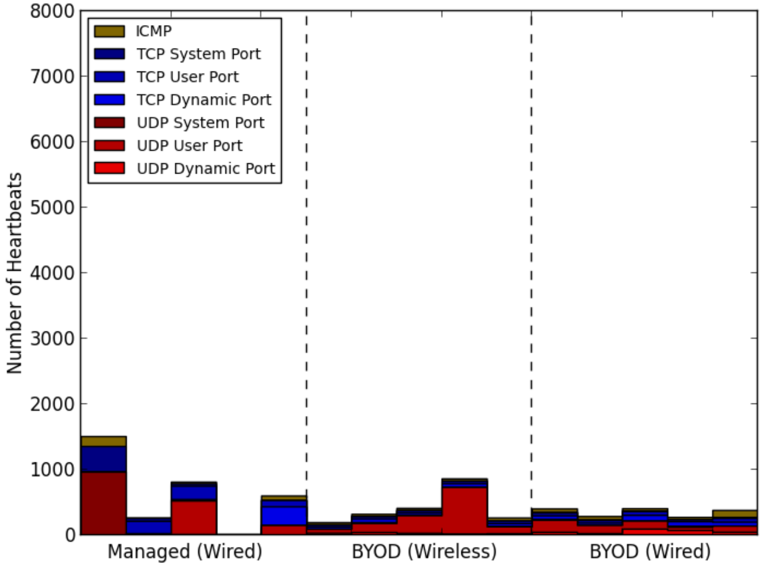
- Different subnet types have different types of heartbeats
- Managed subnets tend to generate fewer outbound heartbeats (except for NAT)
- Managed subnets produce different types of heartbeats depending on their purpose
- BYOD subnets generate many outbound heartbeats from P2P, services, and end user applications



Subnet Analysis – Inbound Heartbeats

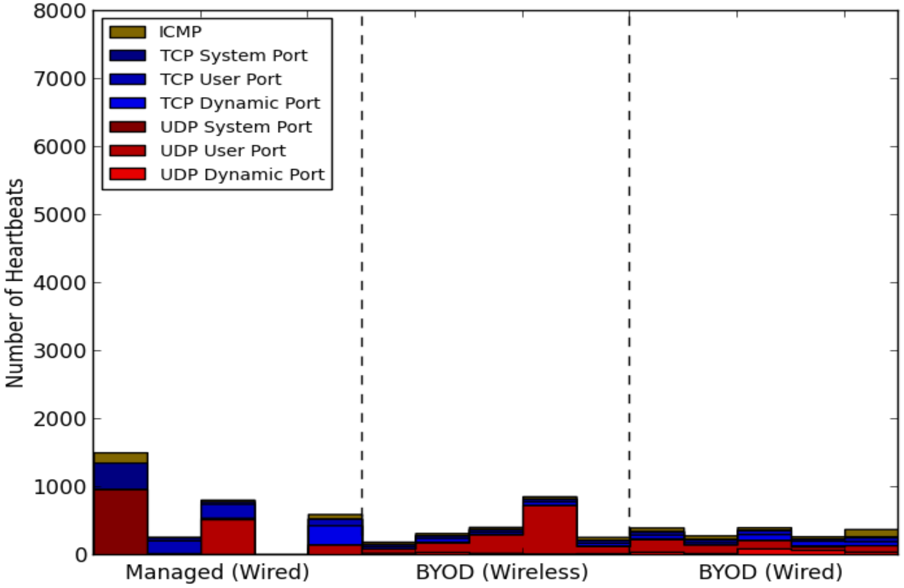
- Managed subnets receive lots of inbound heartbeats from scanning and university hosted services
- Specific types of heartbeats differ depending on the purpose of each subnet
- BYOD subnets have fewer inbound than outbound
- Inbound BYOD heartbeats have similar composition to outbound heartbeats, but also include scanning

Inbound

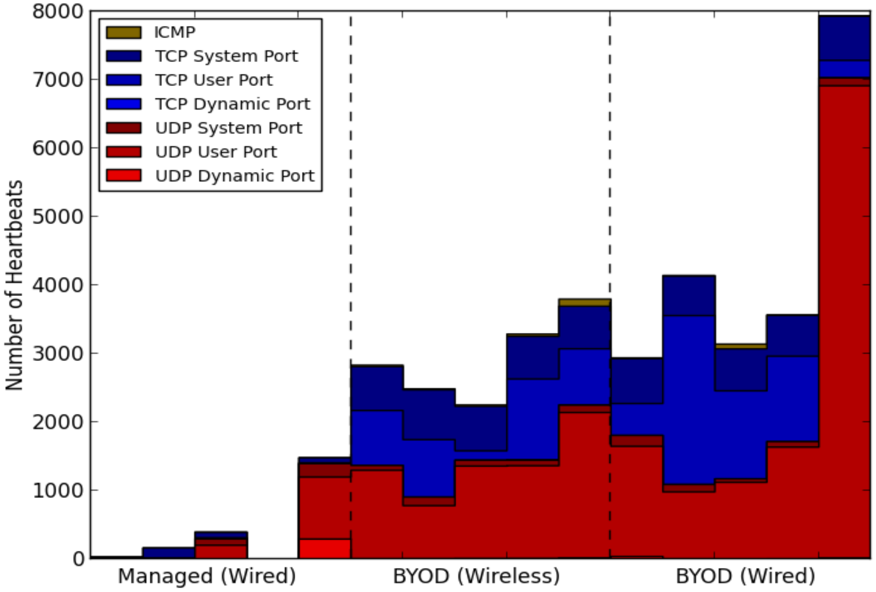


Subnet Analysis: Inbound versus Outbound

Inbound



Outbound



P2P Analysis (1 of 2)

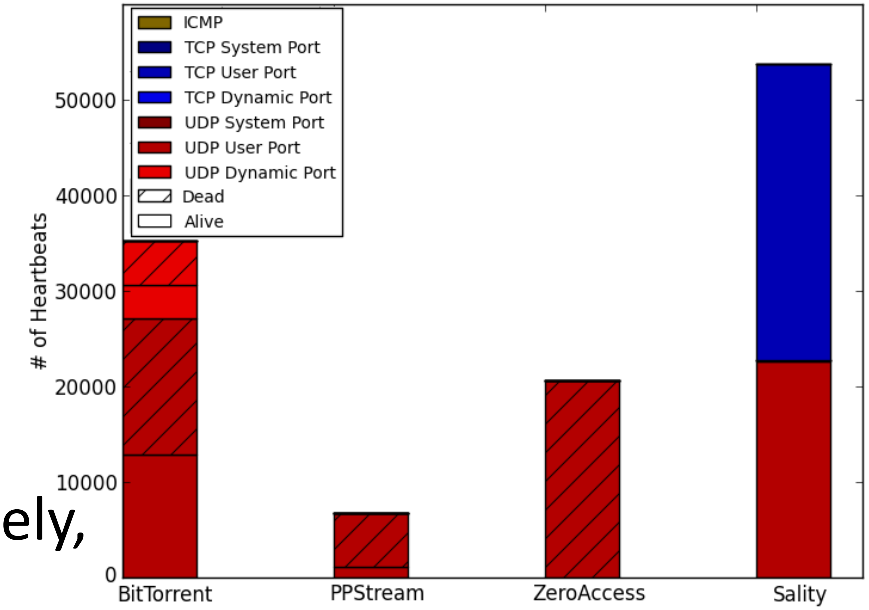
- Different P2P applications often produce heartbeats with similar characteristics, but not always
- BitTorrent:

  - A popular P2P file sharing application, uses UDP exclusively, many dead heartbeats
- PPStream:

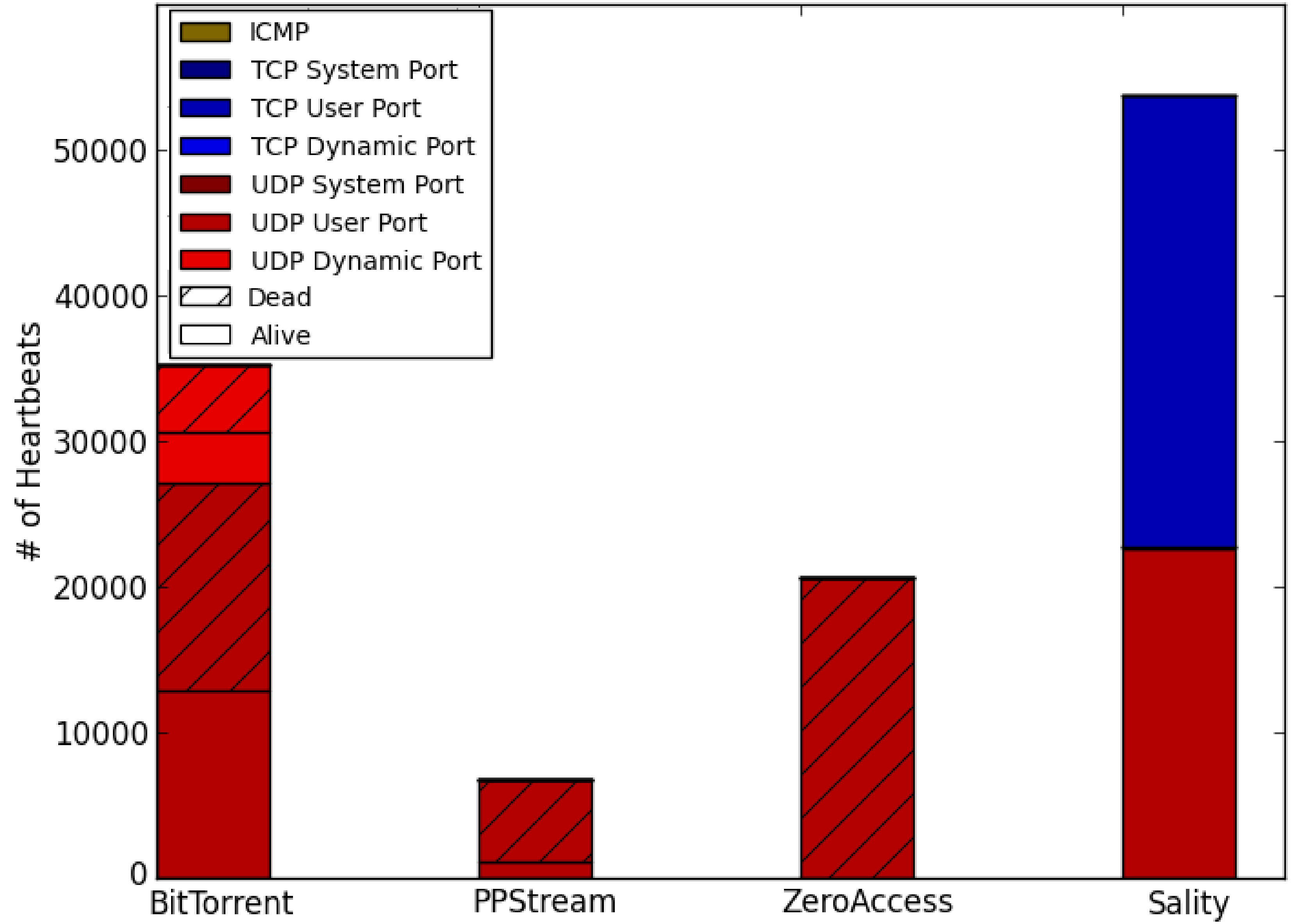
  - An East Asian P2P streaming application, uses UDP exclusively, mostly dead heartbeats
- ZeroAccess:

  - A P2P botnet, uses UDP exclusively, all dead heartbeats
- Sality:

  - Another P2P botnet, uses UDP and TCP, almost all heartbeats are alive



P2P Analysis (2 of 2)



Vendor Analysis (1 of 2)

- Different vendors use heartbeats in different ways
- Microsoft:

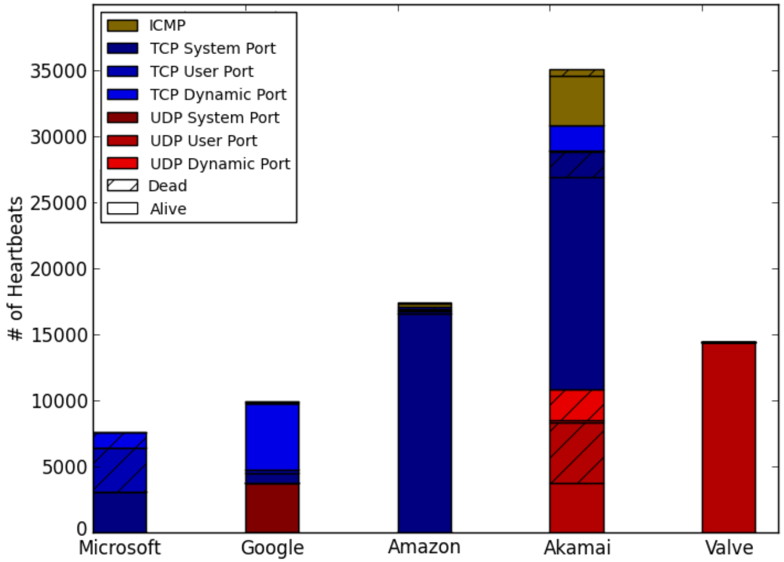
  - Software updates, information gathering, services provided by Microsoft
- Google:

  - Services provided by Google to users
- Amazon:

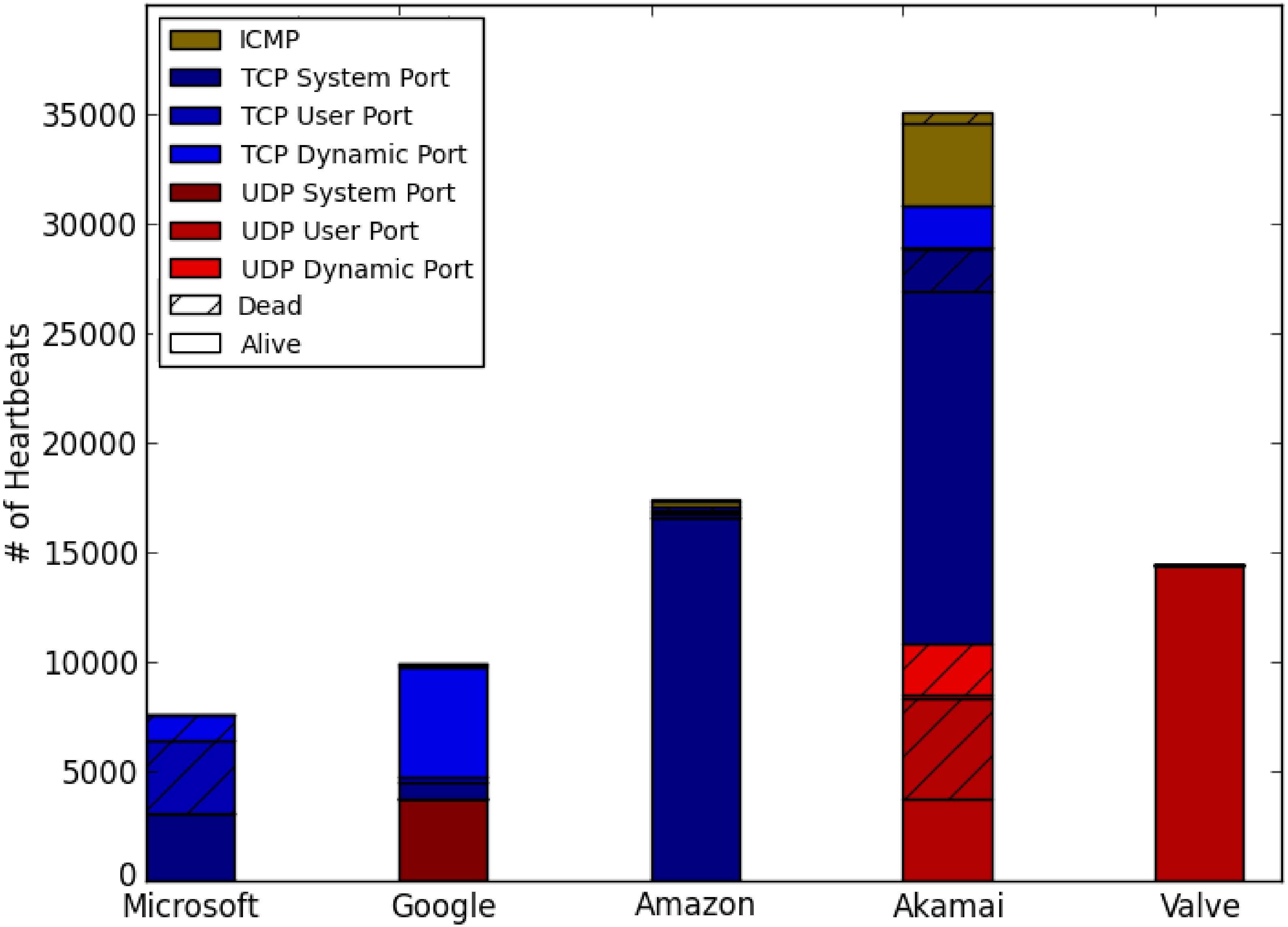
  - Third party services hosted on Amazon's servers
- Akamai:

  - Internal testing and reporting, Akamai NetSession Interface
- Valve:

  - Video games and in home streaming



Vendor Analysis (2 of 2)



Summary of Key Observations

- Heartbeat Characteristics

  -Heartbeats are generally transient and ephemeral
  -Heartbeats are typically short in both period and lifespan
  -Applications tend to generate heartbeats with similar periods and port numbers
  -Heartbeats are typically produced by end-user applications

- Heartbeat Trends

  -P2P applications generate a large number of heartbeats
  -Almost all heartbeats are irregular
  -Heartbeats are mostly outbound (on edge networks)
  -A large number of heartbeats are dead – due to churn and stale connection information

Discussion

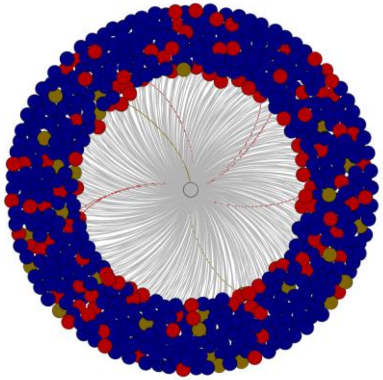
- System administrators of managed infrastructure can use heartbeat information to determine if any (critical) systems have heartbeats to unexpected places

- Security analysts could use heartbeat information to detect unusual applications running on a BYOD network that might pose a risk to the organization

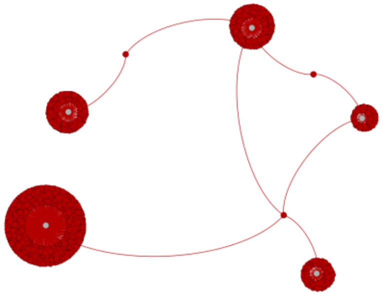
- We need effective ways to make heartbeat information accessible for these purposes, as well as for network operators (visualization can help!)

Example Heartbeat Visualizations using Gephi

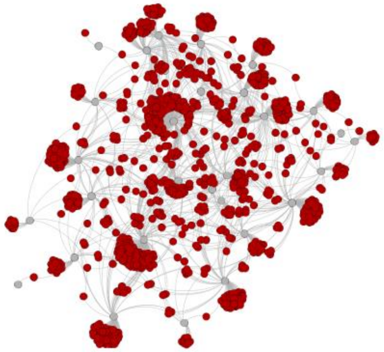
Akamai Node



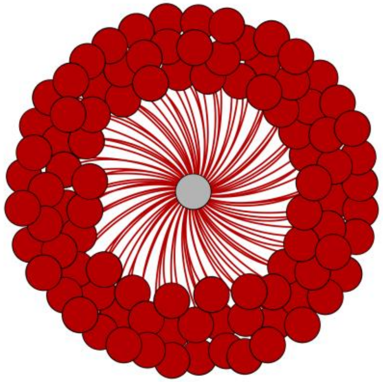
BitTorrent



Sality Botnet



ZeroAccess Botnet



Conclusions

- Network heartbeats: simple but powerful mechanism
  - Network monitoring
  - Security monitoring
  - Effective mechanism for detection of P2P, scanning, malware, and botnet traffic, as well as odd/stale system configurations
- Provides a means to assess the operational health of a campus edge network

- Future work:
  - Sensitivity to parameters used for heartbeat detection
  - In-depth analysis of heartbeats for NAT and DHCP
  - Coping with obfuscation of heartbeats for malware
  - Automating the analysis and interpretation of heartbeats

Questions?