# Beacon detection in PCAP files

Supervisor: Sjoerd PeerIkamp[ShelI]

Leendert van Duijn

Universiteit van Amsterdam - System and Network Engineering

*Leendert.vanduijn@os3.nl*

July 3, 2014

# Beacons

## What are they?

Reoccurring automated messages

## What can they do?

Reveal your location, Leak data, Get Bots configured to do damage

# What sends beacons?

- Malware polling for instructions
- Botnet membership maintenance
- Periodic service checks, Nagios
- Periodic updates, your favorite software
- Visual feedback from network services

# Related work

- ProVeX, deep packet inspection of Malware traffic, Rossow and Dietrich
- Detecting P2P Malware traffic based on regional periodicity, Qiao et al.
- Jackstraws, executable code analysis using behavior graphs, Jacob et al.
- Using host level intrusion detection to detect advanced persistent threats, Liang et al.

# Data sets

- Sinkholing, reducing Malware impact by redirecting it
- Multiple days of traffic dumps available
- Diverse hosts, protocols and realistic data
- Not truly the native behaviour

# Research questions

- Can traffic dumps be used to detect beacons produced by Malware?
- Can detection performance be improved by early classification?
- Is it possible to differentiate Malware in the presence of legitimate beacons?

How can this be used in practice?

# Detecting beacons

- Obtain a traffic dump with suspected beacon activity
- Separate packets into several classes of similar or related traffic
- Identify/prioritize suspect classes using prior knowledge or experience
- Look for local patterns within individual classes
- Export traffic per class to investigate with Wireshark

# Classes

- Focus on relevance
- Capture anomalies
- Adjustable

# Classes

- Focus on relevance
- Capture anomalies
- Adjustable

- Clustering using K means
- Clustering by tree building
- *Rule based, user configurable classes*

# Classifiers

- Source IP address
- TCP Destination port
- Source and destination IP, protocol, length, entropy

# Patterns

- Localized
- Generic
- Performance

# Patterns

- Localized
- Generic
- Performance

- Histogram, activity over time
- Frequency analysis
- Auto correlation

# Sinkhole, what hosts are beaconing?

| | |
|---|---|
| Classifier | Source IP |
| Packets | 2.4M over 2 days |
| Found classes | 421 |
| X axis | Auto correlation over 1 window |
| Y axis | Sliding window in time from top to bottom |
| Selection | From the top 10 in number of packets |

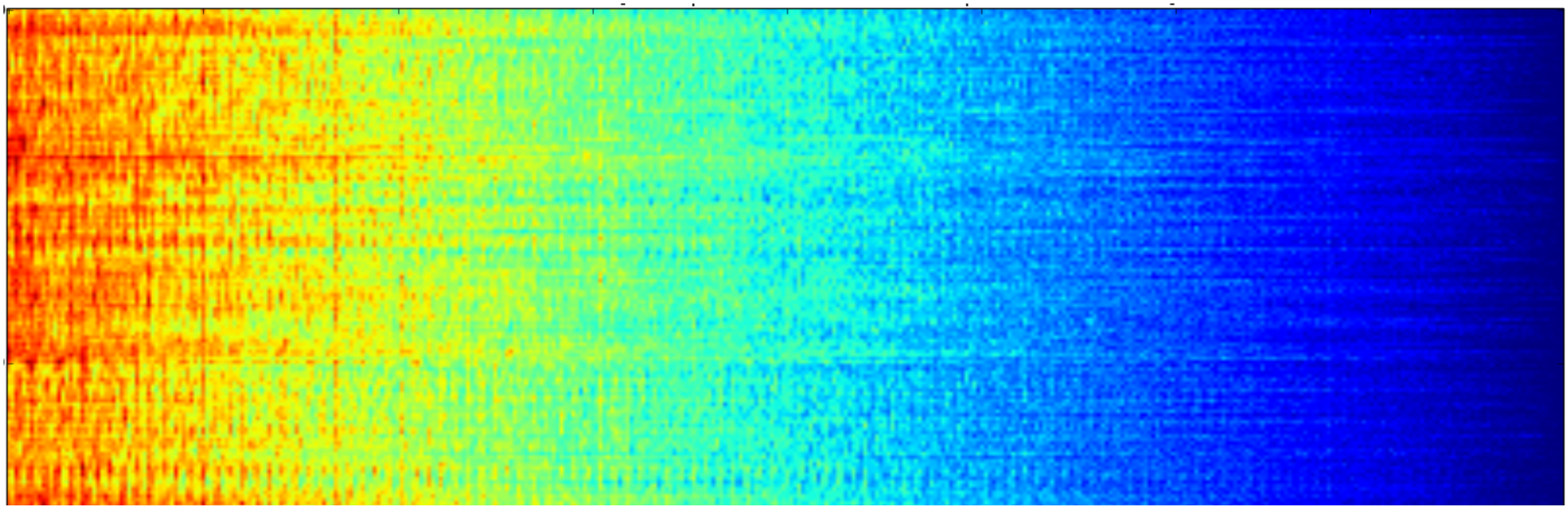

Figure: Outgoing packets sinkhole, all protocols and destinations

# Sinkhole, results

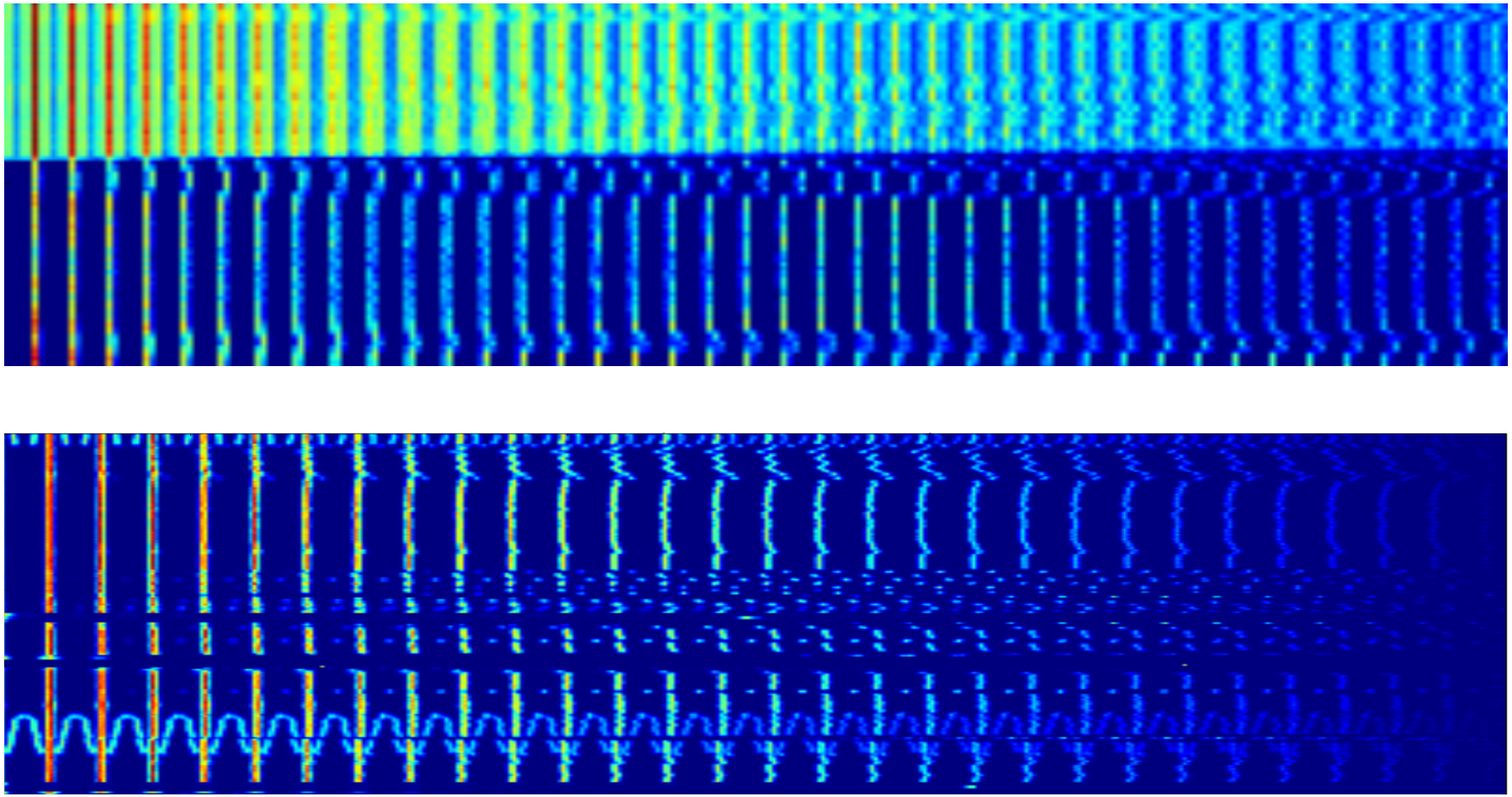

Figure: TCP outgoing every 26 and 3 seconds respectively. Uniform traffic

# Non Malware beacons

- Legitimate beacons can occur, what do they look like?
- Don't websites autorefresh all the time?
- Does encryption hide beacons?

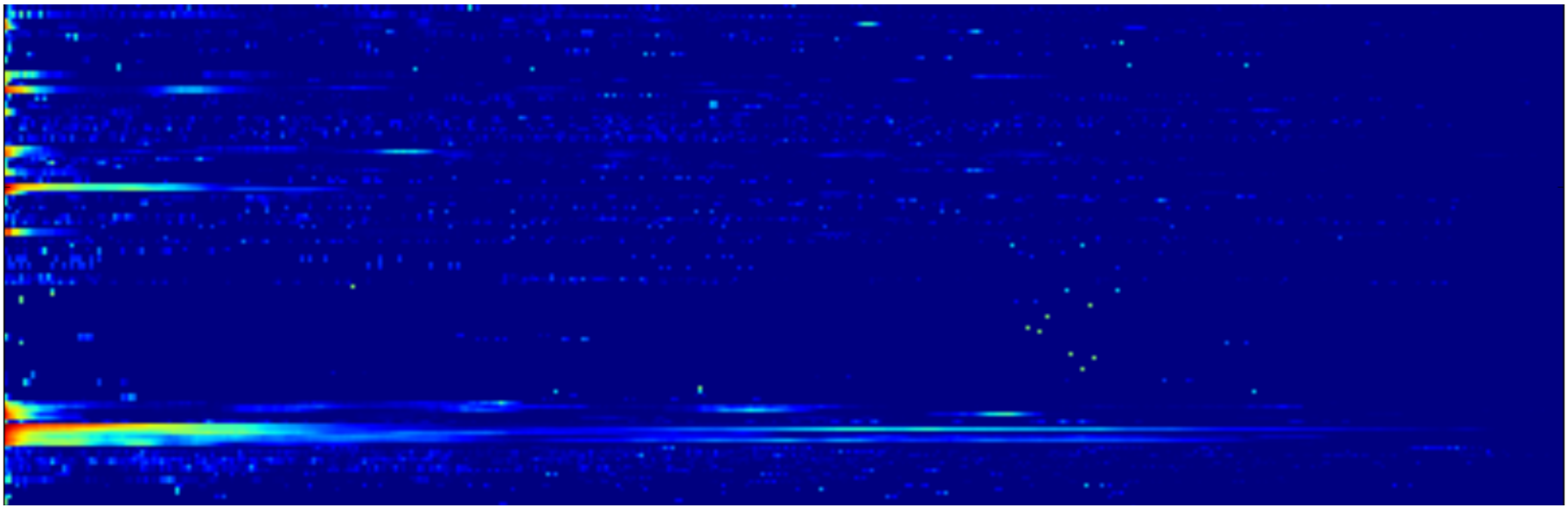

Figure: An hour of HTTPS traffic while writing and browsing

# Non Malware beacons - HTTP, SSH

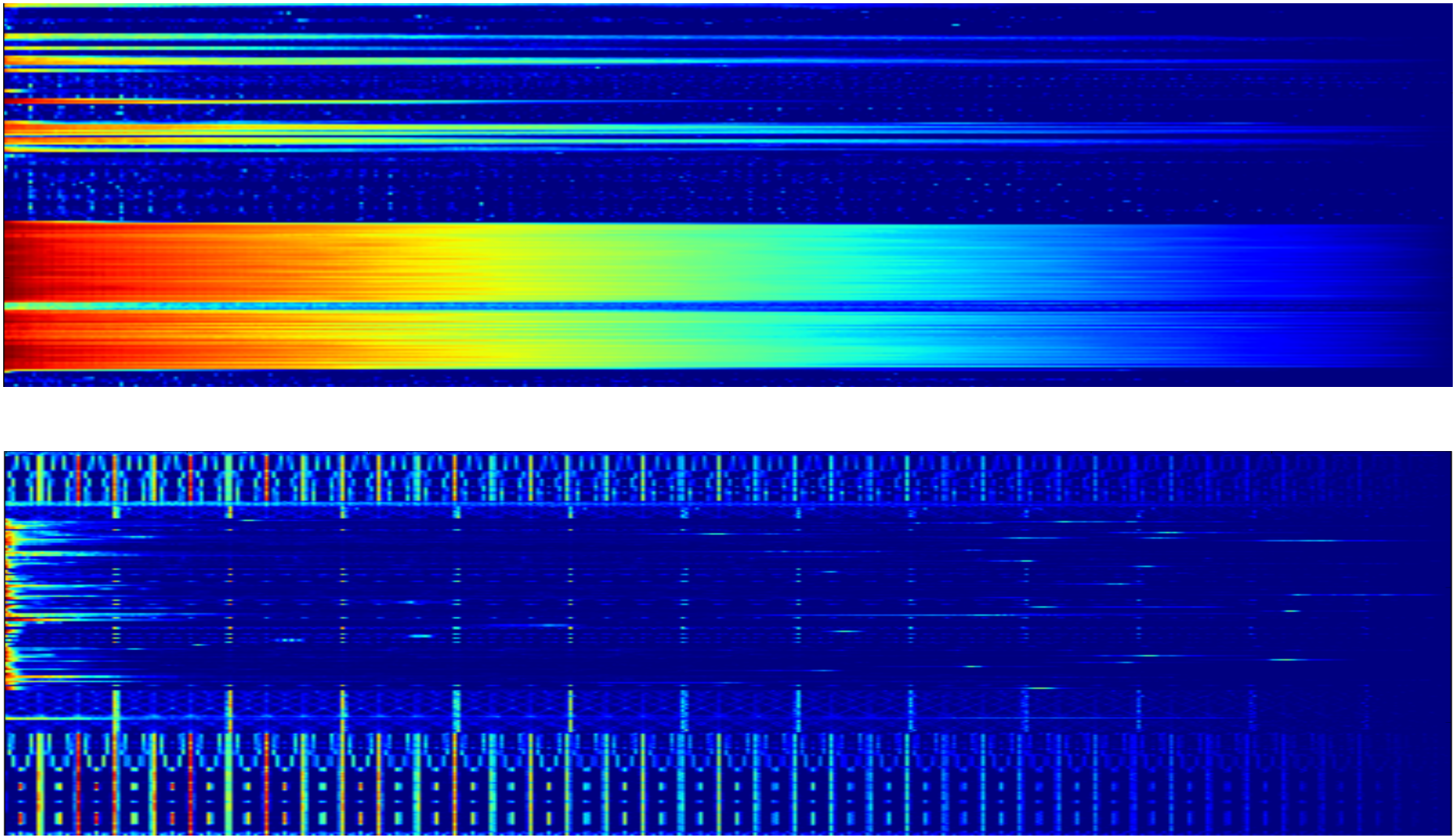

Figure: An hour of traffic, watching a movie, listing some files

# Conclusion

- *Can traffic dumps be used to detect beacons produced by Malware?*
  It is possible to detect beacon traffic in packet dumps using auto correlation of the packet rate over time.
- *Can detection performance be improved by early classification?*
  Using classification or clustering can help in isolating streams/types of traffic, increasing the number of data sets to analyze in exchange for signal clarity.
- *Is it possible to detect Malware in the presence of legitimate beacons?*
  There are features which can be used to distinguish beacons from each other, packet rate, packet uniformity and presence in time.

# Future work

- Define a scoring method for the Auto correlation waterfalls to automate potential hits
- Investigate parameter automation
- Go from audits to live analysis
- Investigate sparse data, methods of combining/splitting data with significant gaps.
- How does it handle noisy data, cloaked Malware and App traffic?

# Questions?

# References


Gregoire Jacob, Ralf Hund, Christopher Kruegel, and Thorsten Holz. Jackstraws: Picking command and control connections from bot traffic. In *Proceedings of the 20th USENIX Conference on Security*, SEC'11, pages 29–29, Berkeley, CA, USA, 2011. USENIX Association. URL `http://dl.acm.org/citation.cfm?id=2028067.2028096`.

Yu Liang, Guojun Peng, Huanguo Zhang, and Ying Wang. An unknown trojan detection method based on software network behavior. *Wuhan University Journal of Natural Sciences*, 18(5):369–376, 2013. ISSN 1007-1202. doi: 10.1007/s11859-013-0944-6. URL `http://dx.doi.org/10.1007/s11859-013-0944-6`.

Yong Qiao, Yuexiang Yang, Jie He, Chuan Tang, and Yingzhi Zeng. Detecting p2p bots by mining the regional periodicity. *Journal of Zhejiang University - Science C*, 14(9): 682–700, 2013.

Christian Rossow and Christian J. Dietrich. ProVeX: Detecting Botnets with Encrypted Command and Control Channels. In *Proceedings of the 10th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA)*, July 2013.