The Uber Data Source: Holy Grail or Final Fantasy?

Josh Goldfarb

FloCon

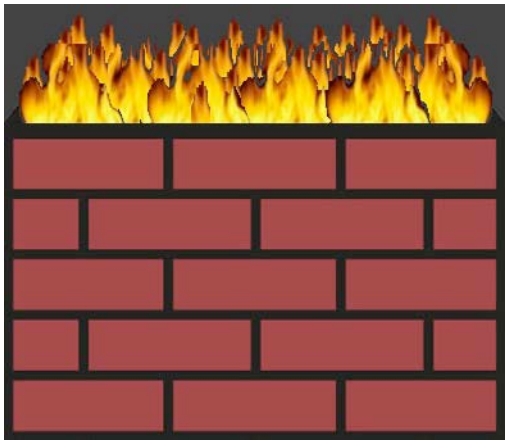
January 2012

Poignant Quote

"We are drowning in information, but starved for knowledge"

--John Naisbitt

Which Data Source?

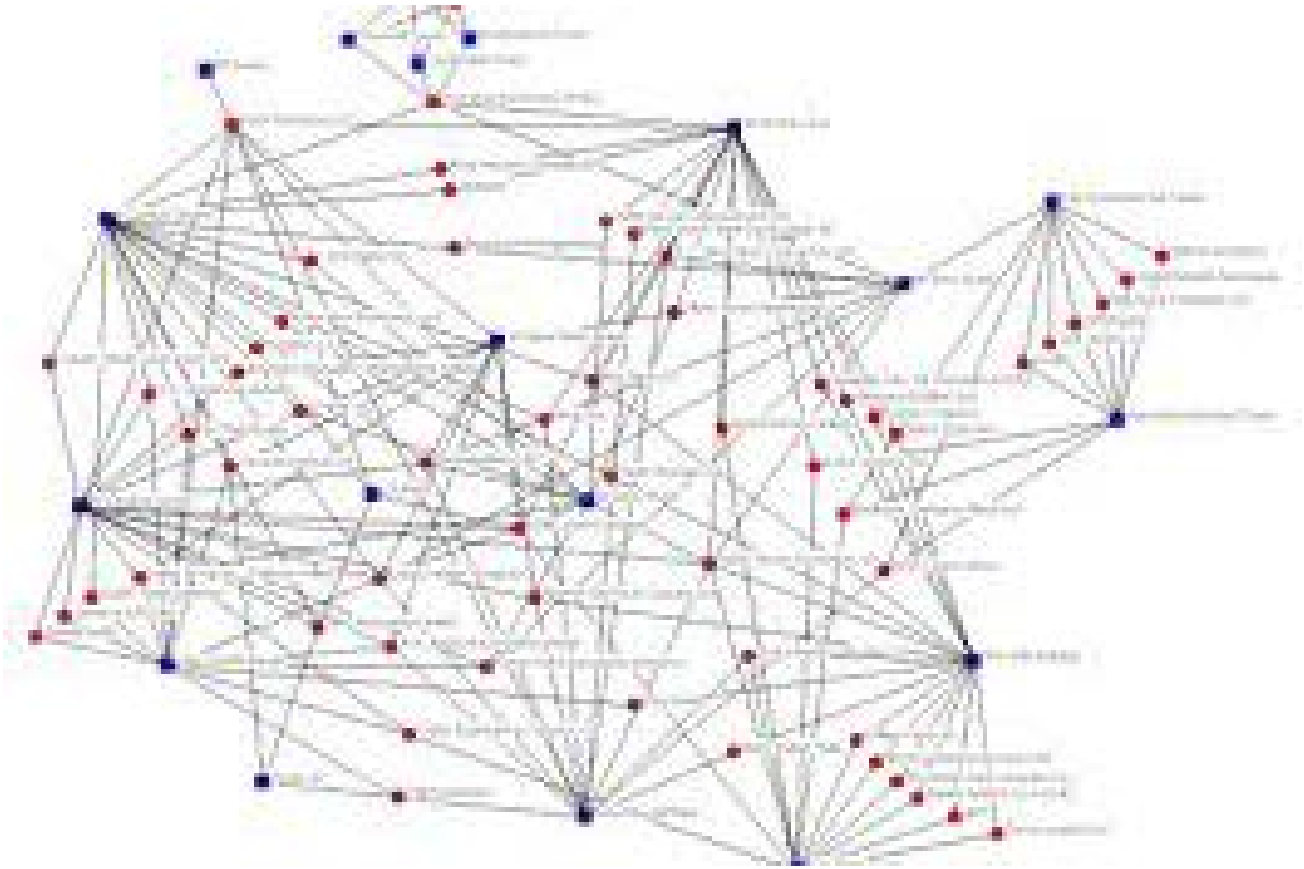


CERT | Software Engineering Institute
Carnegie Mellon

Unfortunate Reality

- No one data type gives organizations what they need analytically/forensically/legally
- There is great uncertainty of what data needs to be collected and stored to ensure adequate “network knowledge”
- To play it safe, organizations often collect everything they can
- Each data source has a different value for network monitoring

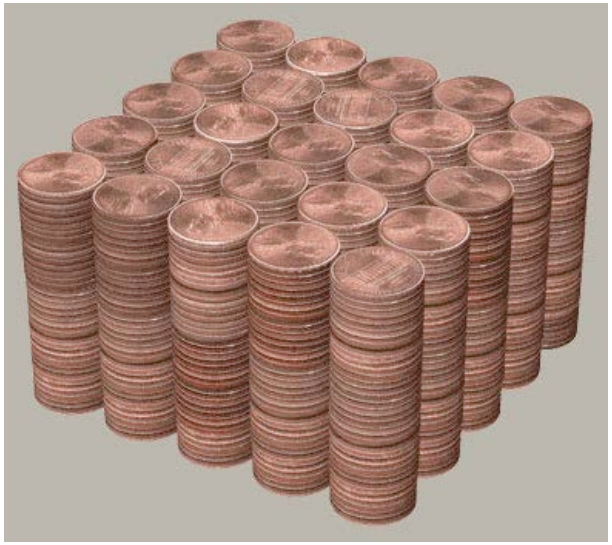
Unfortunate Results



Creates Inefficiencies

- Causes confusion and inhibits incident response/forensics
- Complicates analytical/operational workflow and obstructs proper network monitoring
- Wastes precious skilled labor (analyst/technical/professional) cycles on data munging/data organization rather than monitoring
- Utilizes extra storage space that could be used instead to increase the length of retention rather than the variety of data stored

Value Over Volume



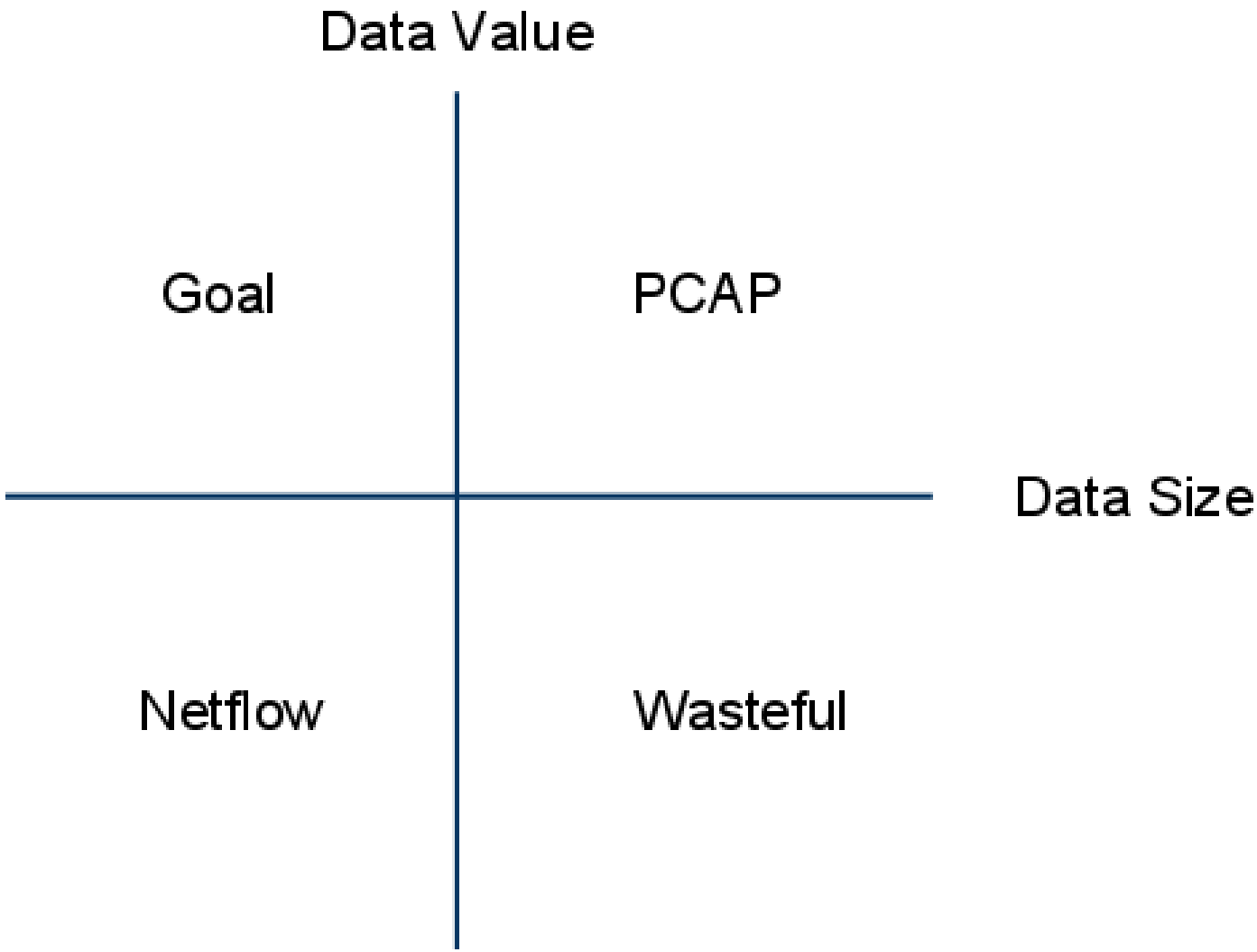
Challenge

- Organized, well-structured approach necessary for network monitoring success
- Volume and variety of network data make this a challenge
- Is there a better way?

Uber Data Source?

Concept

- Enrich layer 4 meta-data (e.g., netflow) with layer 7 (application layer) data
- Focus on data value instead of data volume
- Identify layer 7 fields that add the greatest value
- Tune the dial appropriately between extremely compact size, but no context and full context, but extremely large size
- For certain protocols, this is already standard practice!
- Generalize to all protocols


Data Value
Goal
PCAP
Data Size
Netflow
Wasteful

Contact Information

Josh Goldfarb

Freelance Security Analyst

josh@yourcyberanalyst.com

http://www.yourcyberanalyst.com/

http://ananalyticalapproach.blogspot.com/