Chapter 2: Planning Data Collection

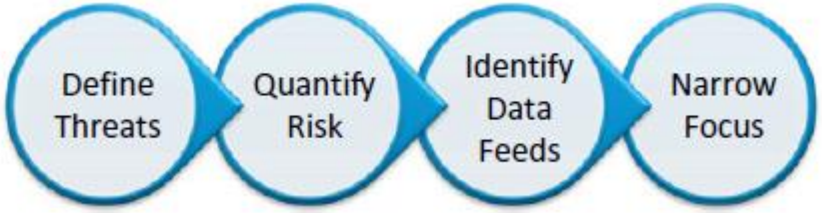
Gonzalo De La Torre Parra, Ph.D.

Fall 2021

The Applied Collection Framework

- Abraham Lincoln said, “If I had six hours to chop down a tree, I’d spend the first four hours sharpening my axe.”

- The Applied Collection Framework



How can we define potential Threats

- What devices generate raw research data, and how does that data traverse the network?
- From what devices do employees process raw research data?
- On what devices is processed research data stored?
- Who has access to raw and processed research data?
- Is raw or processed research data available from outside the network?
- What paths into the internal network are available externally?
- What level of access do temporary employees have to research data?

Quantifying Risks

- Once a list of potential technical threats has been identified, those threats must be prioritized.
- One way to achieve this is to calculate the risk posed by each potential threat by determining the product of impact and probability.
- This is represented by the equation Impact (I) Probability (P)¼Risk (R).

Impact

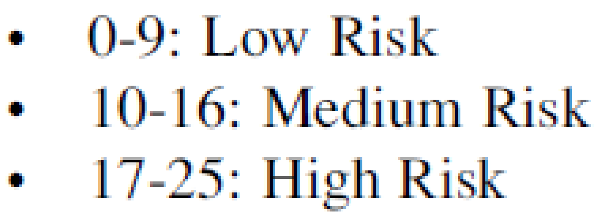
- Impact takes into consideration how a given threat, should it manifest itself, could affect the organization.
- This is measured on a scale of 1 to 5, with 1 meaning that the threat would have little impact, and 5 meaning that the threat would have a large impact.

Probability

- Probability represents the likelihood that a threat will manifest itself.
- This is also measured on a scale of 1 to 5, with 1 meaning that there is a low probability that the threat will manifest itself, and 5 meaning that the threat has a high probability of manifestation.
- The determination of probability can include consideration of an asset's exposure or attack surface visible to the threat, the level of intimacy with the network required to execute an attack, or even the likelihood that someone would be able to gain physical access to an asset.
- Over enough time, the probability of a vulnerability being exploited increases.
- When we create probability rankings they represent the moment in time in which they are created, which means that they should be revisited over time.

Risk

- The product of impact and probability is the level of risk, or the “risk weight” the threat poses to the security of the network in relation to the organization’s business goals. This is measured on a scale of 1 to 25.



- 0-9: Low Risk
- 10-16: Medium Risk
- 17-25: High Risk

Example

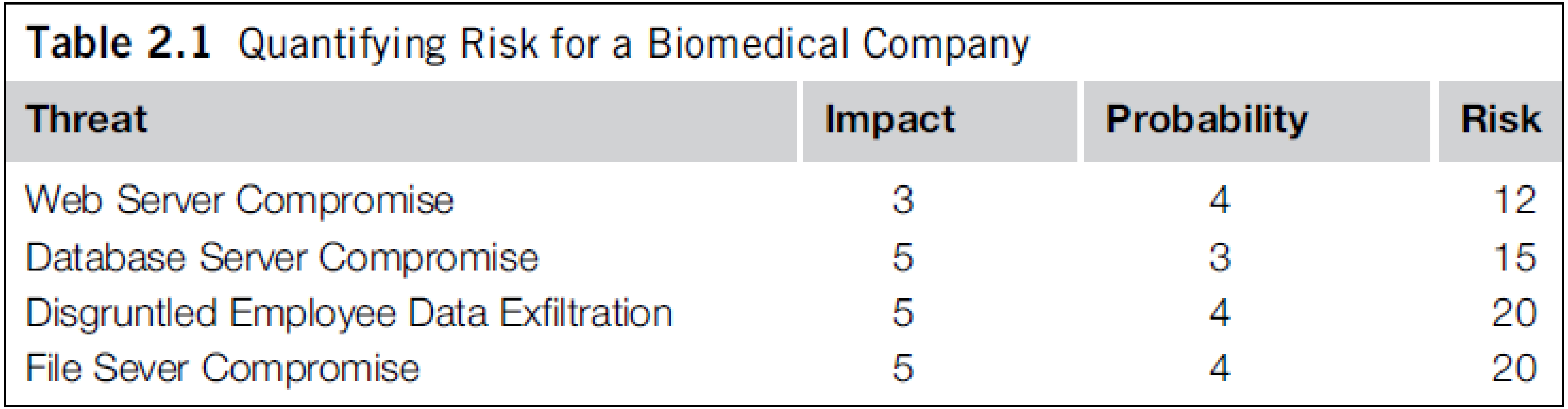


Table 2.1 Quantifying Risk for a Biomedical Company

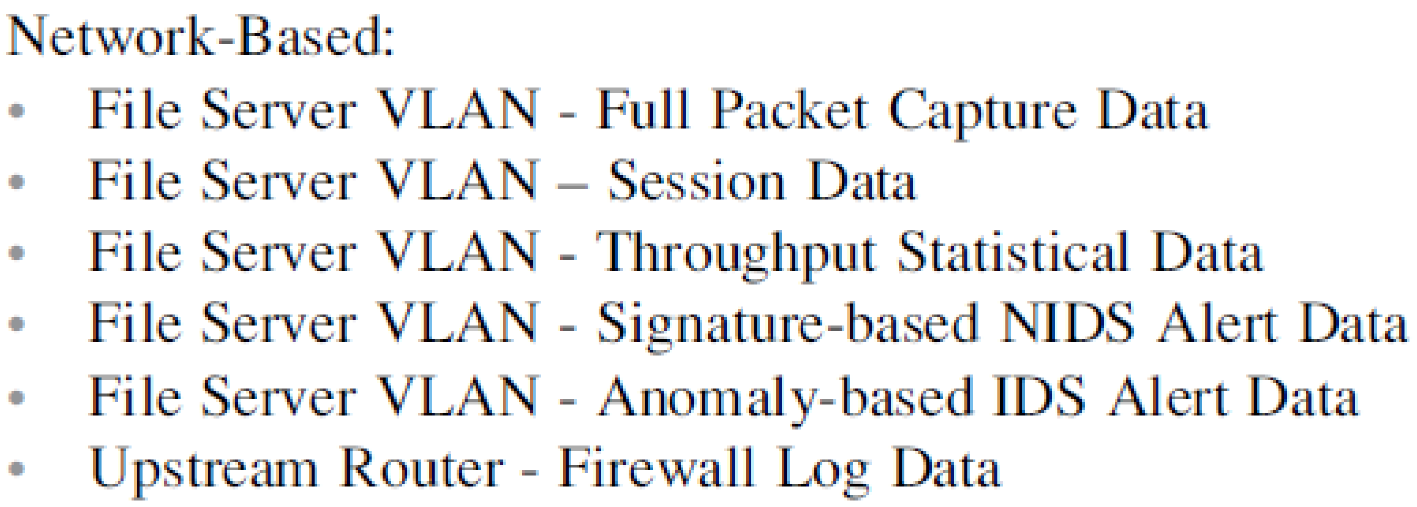
Threat	Impact	Probability	Risk
Web Server Compromise	3	4	12
Database Server Compromise	5	3	15
Disgruntled Employee Data Exfiltration	5	4	20
File Sever Compromise	5	4	20

- These numbers are still subjective.
- It is important that these numbers are generated by committee and that the same group of individuals participate in the ranking of all identified threats.

Identify Data Feeds

- Threat of File Server Compromise:
  - What is the server's architecture?
  - Location in the network
  - Who has access to this server?
  - What are the pathways the data can take to and from it?

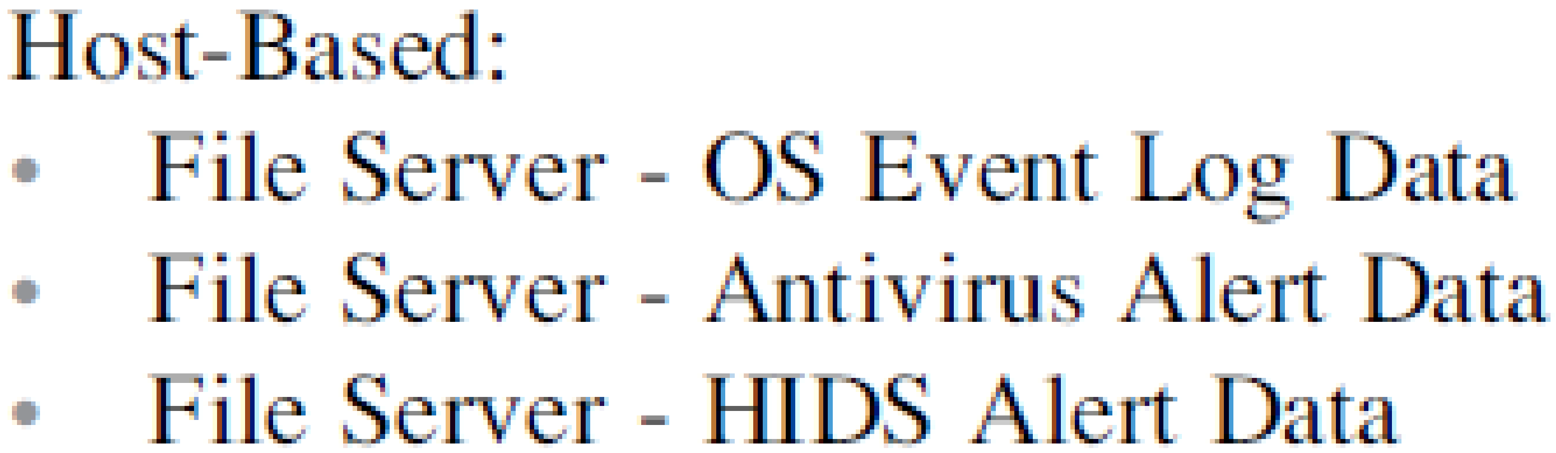
Potential Network-based Data Feeds



Network-Based:

- File Server VLAN - Full Packet Capture Data
- File Server VLAN – Session Data
- File Server VLAN - Throughput Statistical Data
- File Server VLAN - Signature-based NIDS Alert Data
- File Server VLAN - Anomaly-based IDS Alert Data
- Upstream Router - Firewall Log Data

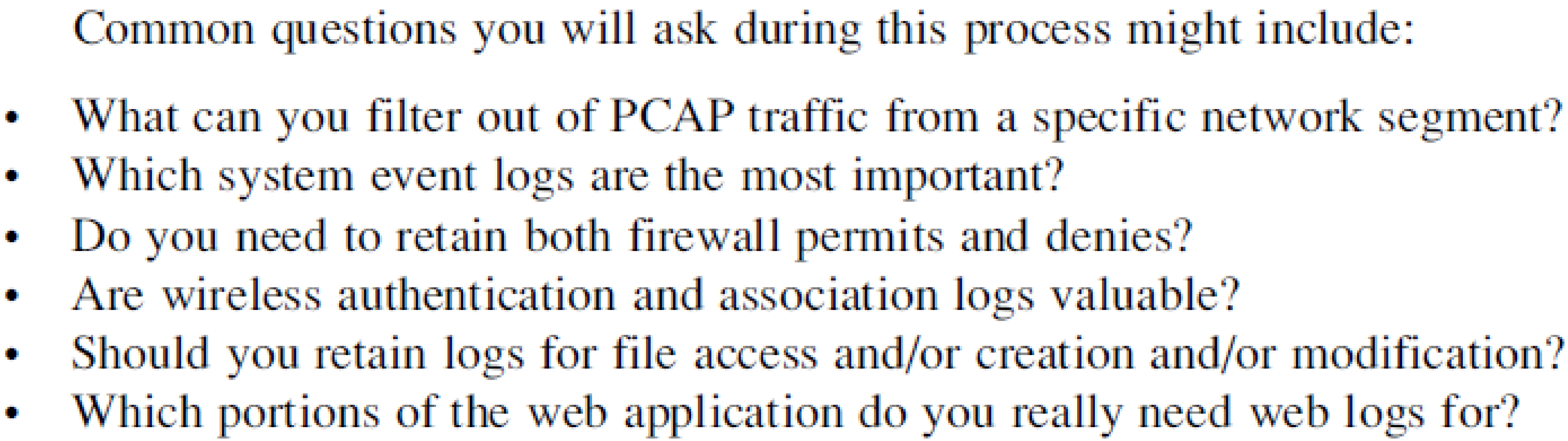
Potential Host-based Data Feeds



Host-Based:

- File Server - OS Event Log Data
- File Server - Antivirus Alert Data
- File Server - HIDS Alert Data

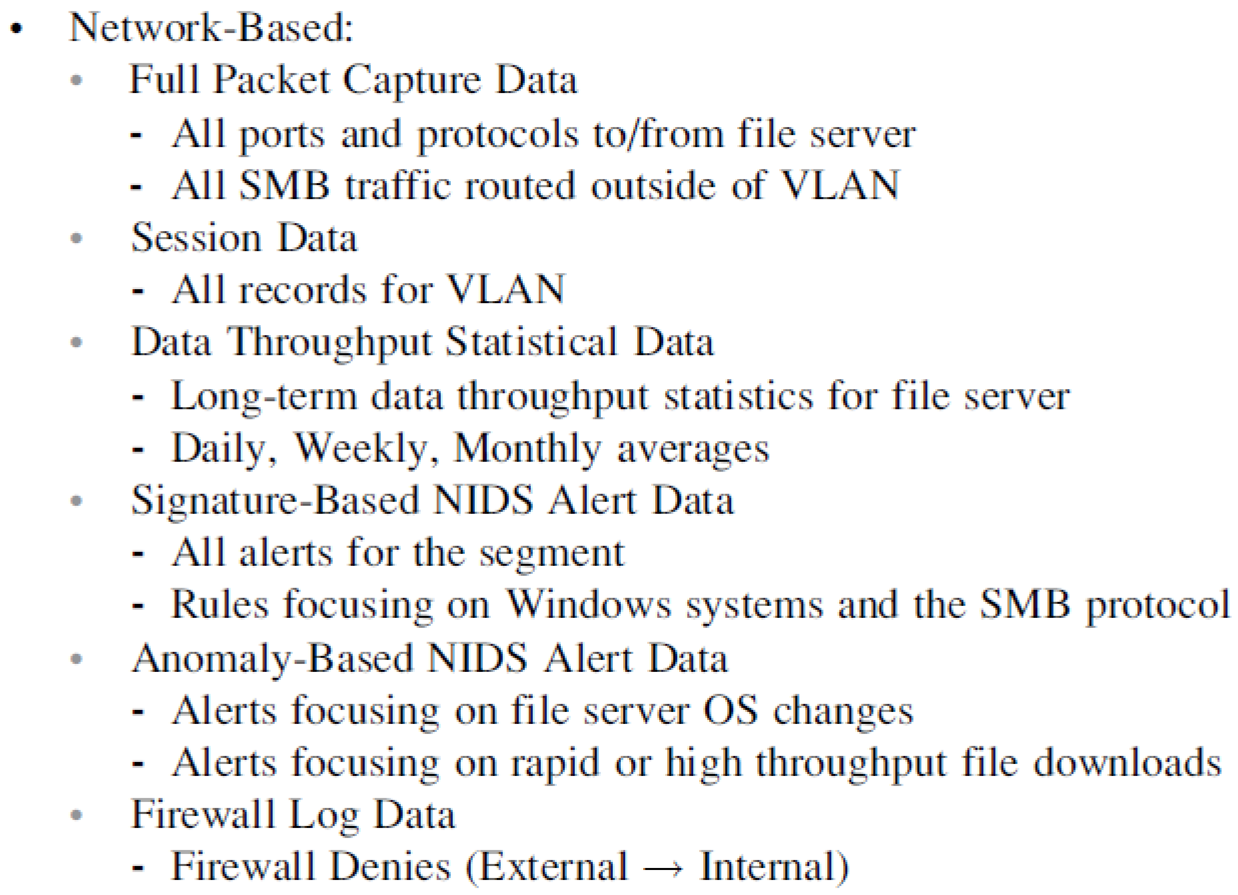
Narrow Focus



Common questions you will ask during this process might include:

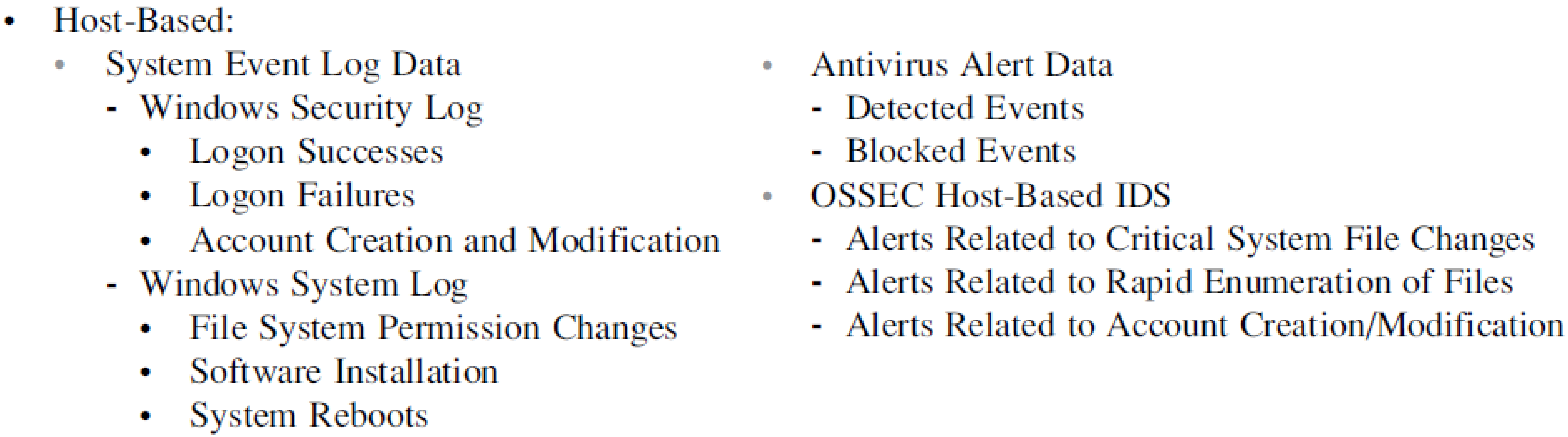
- What can you filter out of PCAP traffic from a specific network segment?
- Which system event logs are the most important?
- Do you need to retain both firewall permits and denies?
- Are wireless authentication and association logs valuable?
- Should you retain logs for file access and/or creation and/or modification?
- Which portions of the web application do you really need web logs for?

Network Granular Information



- Network-Based:
  - Full Packet Capture Data
    - All ports and protocols to/from file server
    - All SMB traffic routed outside of VLAN
  - Session Data
    - All records for VLAN
  - Data Throughput Statistical Data
    - Long-term data throughput statistics for file server
    - Daily, Weekly, Monthly averages
  - Signature-Based NIDS Alert Data
    - All alerts for the segment
    - Rules focusing on Windows systems and the SMB protocol
  - Anomaly-Based NIDS Alert Data
    - Alerts focusing on file server OS changes
    - Alerts focusing on rapid or high throughput file downloads
  - Firewall Log Data
    - Firewall Denies (External → Internal)

Host Granular Information



- Host-Based:
  - System Event Log Data
    - Windows Security Log
      - Logon Successes
      - Logon Failures
      - Account Creation and Modification
    - Windows System Log
      - File System Permission Changes
      - Software Installation
      - System Reboots
  - Antivirus Alert Data
    - Detected Events
    - Blocked Events
  - OSSEC Host-Based IDS
    - Alerts Related to Critical System File Changes
    - Alerts Related to Rapid Enumeration of Files
    - Alerts Related to Account Creation/Modification