Chapter 2: Planning Data Collection

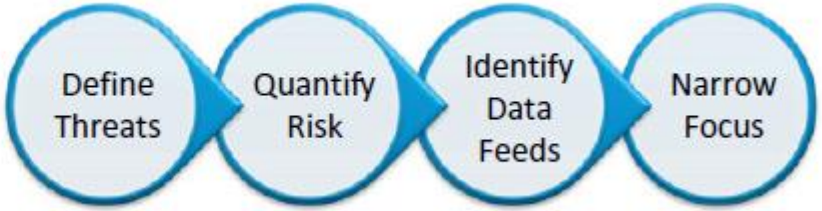
Gonzalo De La Torre Parra, Ph.D.

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

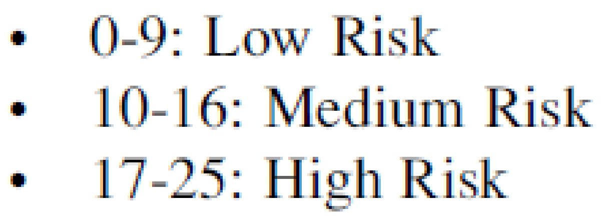
- Impact takes into consideration how a given threat, should it manifest itself, could affect the organization.
- This is measured on a scale of 1 to 5, with 1 meaning that the threat would have little impact, and 5 meaning that the threat would have a large impact.

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

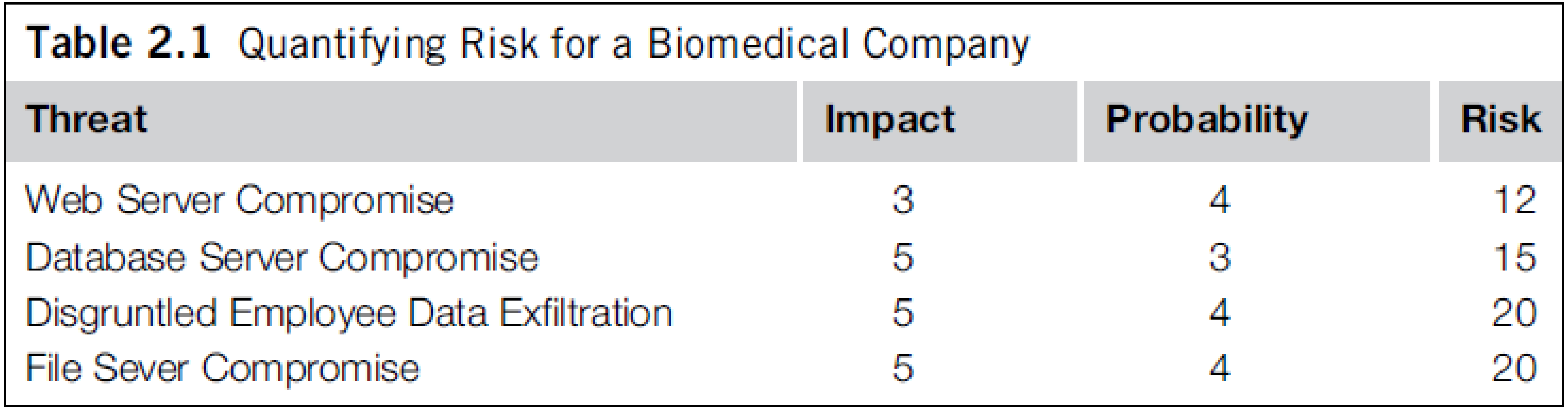


Table 2.1 Quantifying Risk for a Biomedical Company

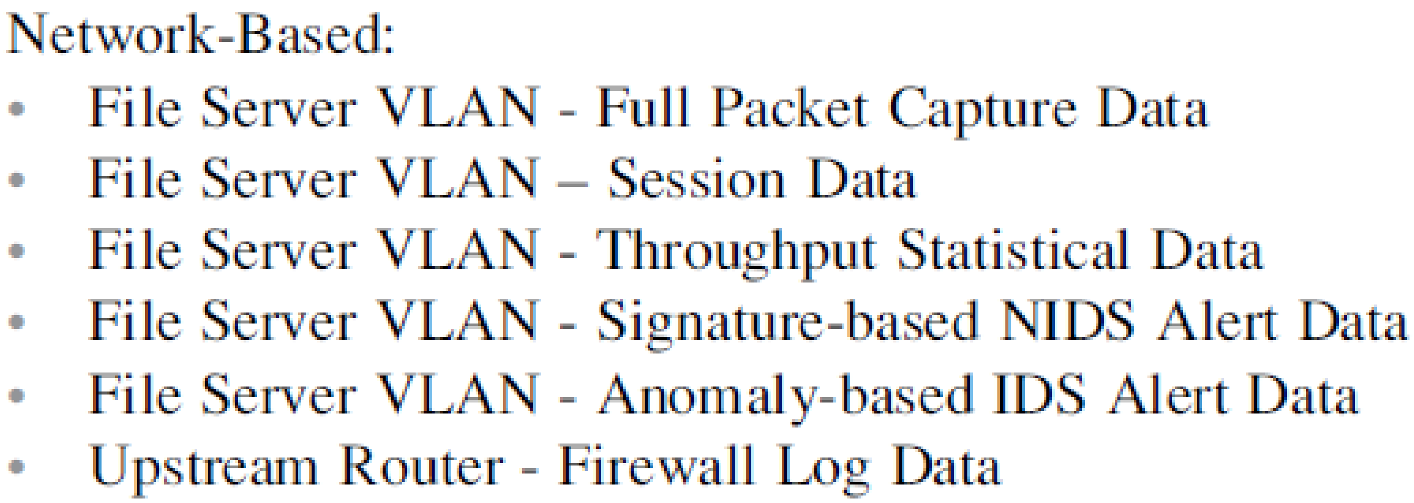
Threat	Impact	Probability	Risk
Web Server Compromise	3	4	12
Database Server Compromise	5	3	15
Disgruntled Employee Data Exfiltration	5	4	20
File Sever Compromise	5	4	20

- 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?

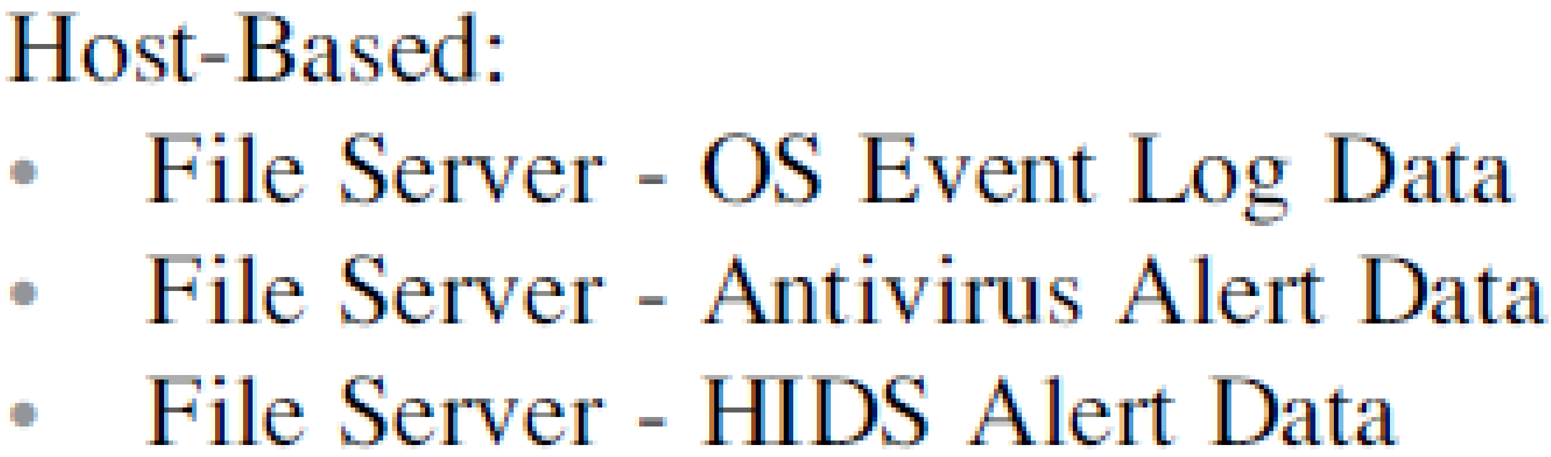
Potential Network-based Data Feeds



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

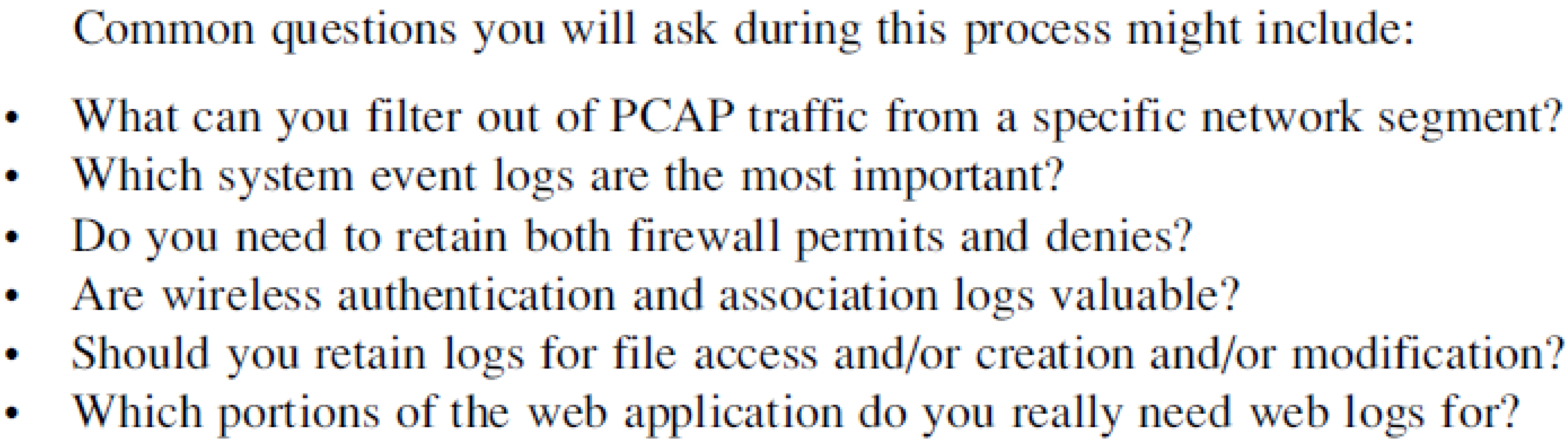
Potential Host-based Data Feeds



Host-Based:

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

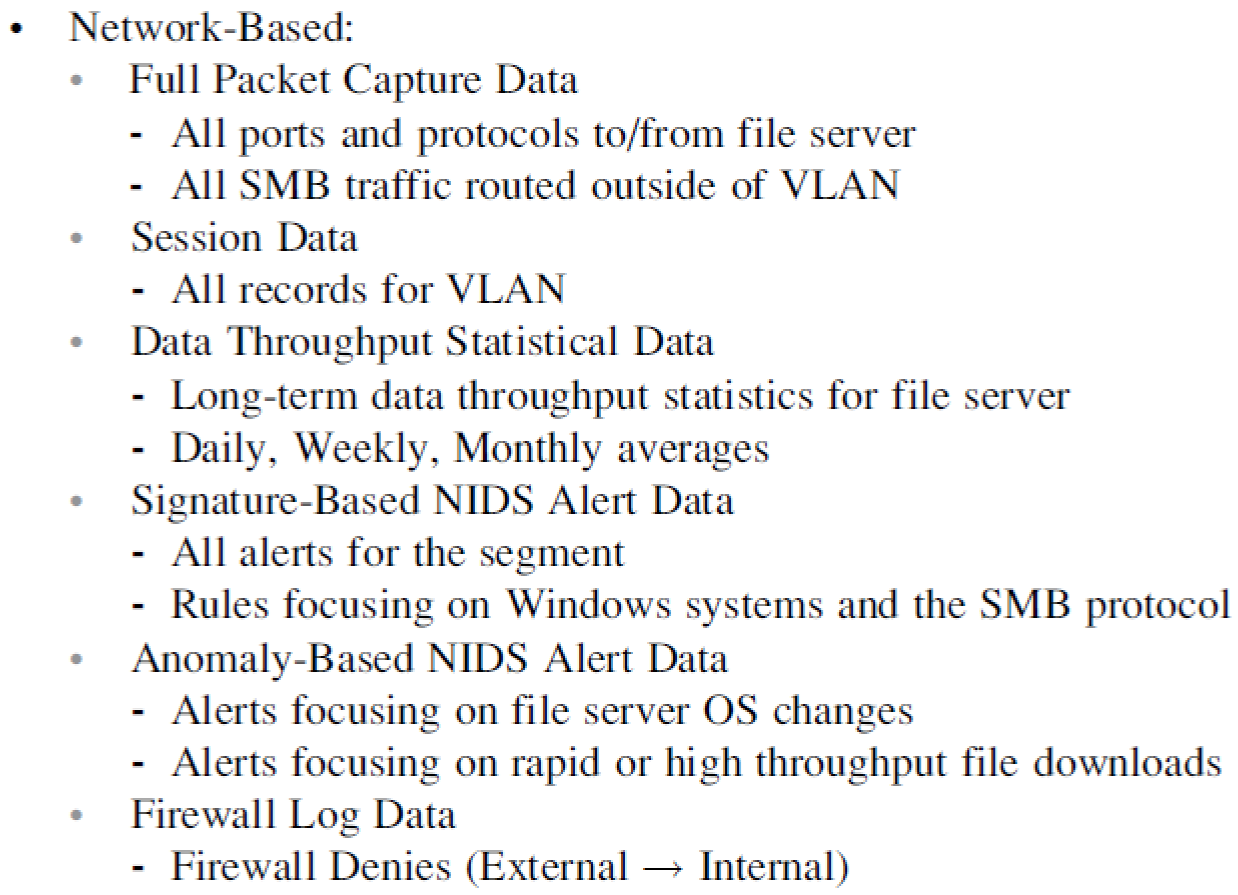
Narrow Focus



Common questions you will ask during this process might include:

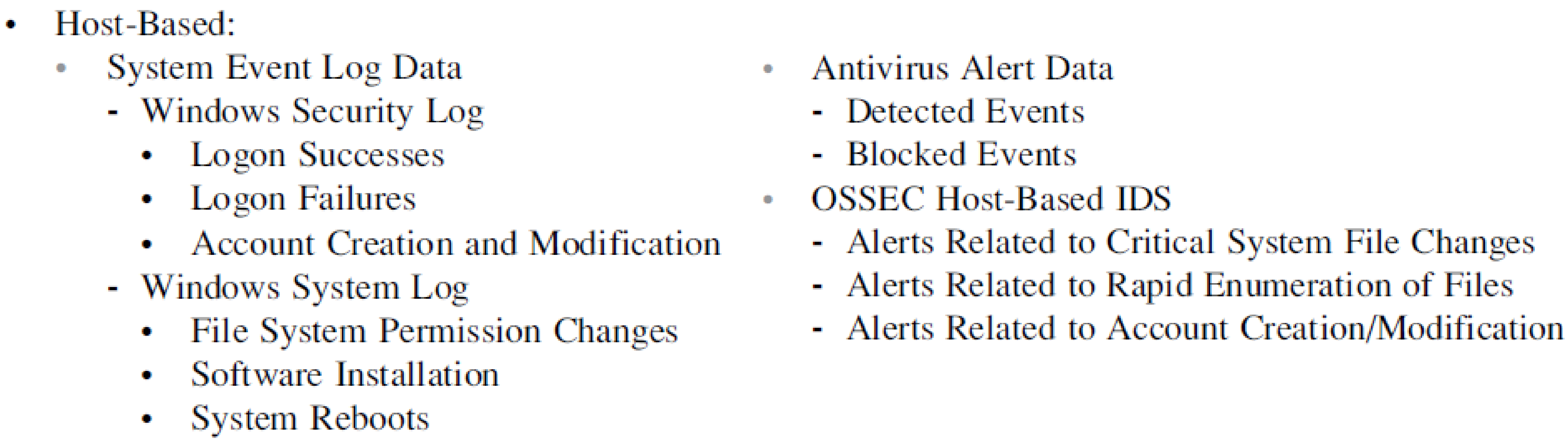
- What can you filter out of PCAP traffic from a specific network segment?
- Which system event logs are the most important?
- Do you need to retain both firewall permits and denies?
- Are wireless authentication and association logs valuable?
- Should you retain logs for file access and/or creation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

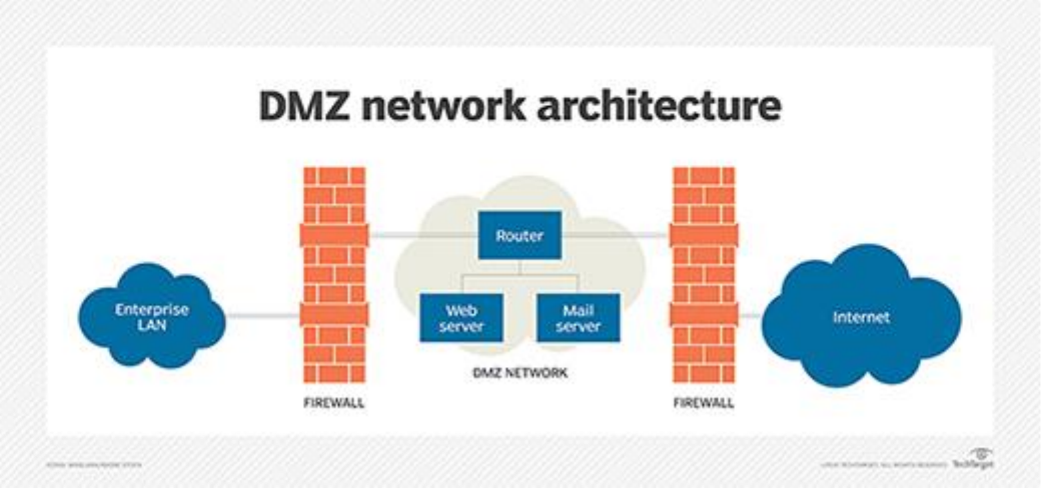
Host Granular Information



- 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

What is a DMZ Network?

- A demilitarized zone (DMZ) is a perimeter network that protects an organization’s internal local-area network (LAN) from untrusted traffic.
- A common DMZ meaning is a subnetwork that sits between the public internet and private networks. It exposes external-facing services to untrusted networks and adds an extra layer of security to protect the sensitive data stored on internal networks, using firewalls to filter traffic.
- The end goal of a DMZ is to allow an organization to access untrusted networks, such as the internet, while ensuring its private network or LAN remains secure. Organizations typically store external-facing services and resources, as well as servers for the Domain Name System (DNS), File Transfer Protocol (FTP), mail, proxy, Voice over Internet Protocol (VoIP), and web servers, in the DMZ.
- These servers and resources are isolated and given limited access to the LAN to ensure they can be accessed via the internet but the internal LAN cannot. As a result, a DMZ approach makes it more difficult for a hacker to gain direct access to an organization’s data and internal servers via the internet.



Why are DMZs important?

- DMZs provide a level of network segmentation that helps protect internal corporate networks. These subnetworks restrict remote access to internal servers and resources, making it difficult for attackers to access the internal network. This strategy is useful for both individual use and large organizations.

- Businesses place applications and servers that are exposed to the internet in a DMZ, separating them from the internal network. The DMZ isolates these resources so, if they are compromised, the attack is unlikely to cause exposure, damage or loss.

How does a DMZ work?

- DMZs function as a buffer **zone between the public internet and the private network**. The DMZ subnet is deployed between two firewalls.
- All **inbound network packets are then screened using a firewall** or other security appliance before they arrive at the servers hosted in the DMZ.
- If better-prepared **threat actors** pass through the first firewall, **they must then gain unauthorized access to the services in the DMZ before they can do any damage**. Those systems are likely to be hardened against such attacks.
- Finally, assuming well-resourced **threat actors** take over a system hosted in the DMZ, they **must still break through the internal firewall before they can reach sensitive enterprise resources**. Determined attackers can breach even the most secure DMZ architecture. However, a DMZ under attack will set off alarms, giving security professionals enough warning to avert a full breach of their organization.

What are the benefits of using a DMZ?

- **Access control.** A DMZ network provides access control to services outside an organization's network perimeters that are accessed from the internet.
- **Network reconnaissance prevention.** A DMZ also prevents an attacker from being able to scope out potential targets within the network. Even if a system within the DMZ is compromised, the internal firewall still protects the private network, separating it from the DMZ.
- **Protection against Internet Protocol (IP) spoofing.** In some cases, attackers attempt to bypass access control restrictions by spoofing an authorized IP address to impersonate another device on the network. A DMZ can stall potential IP spoofers, while another service on the network verifies the IP address's legitimacy by testing whether it is reachable.

Examples of DMZs

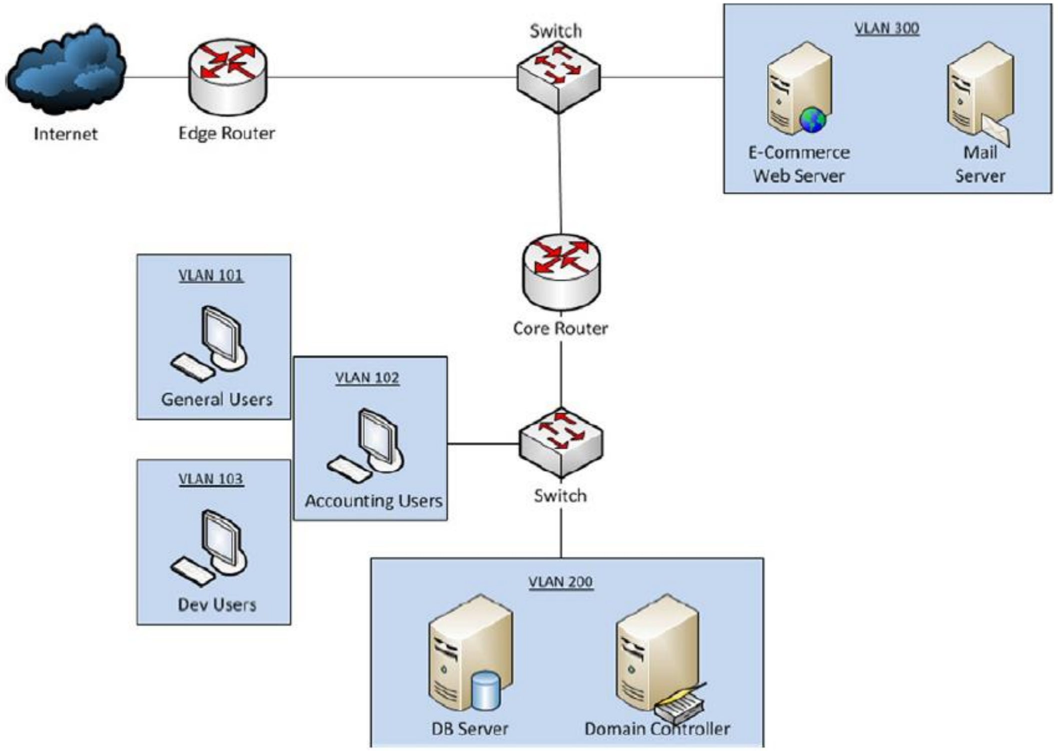
- Cloud services. Some cloud services, such as Microsoft Azure, use a hybrid security approach in which a DMZ is implemented between an organization's on-premises network and the virtual network.
- Home networks. A DMZ can also be useful in a home network in which computers and other devices are connected to the internet using a broadband router and configured into a LAN. Some home routers include a DMZ host feature. This can be contrasted with the DMZ subnetwork used in organizations with many more devices than would be found in a home.
- Industrial control systems (ICS). DMZs provide a potential solution to the security risks of ICSes. Industrial equipment, such as turbine engines, or ICSes are being merged with information technology (IT), which makes production environments smarter and more efficient, but it also creates a larger threat surface.

CASE SCENARIO: ONLINE RETAILER

- An online retailer is establishing an NSM capability for the first time. Our fictitious company, Purple Dog Inc. (PDI), uses their website to market and sell crafts and knick-knacks produced by other suppliers. They have no traditional brick-and-mortar stores, so their entire revenue stream depends upon their ability to make sales from their website.

Online Retailer Cont.

- This is a fairly typical network design with publicly accessible servers in a DMZ behind an edge router. Users and internal network servers reside in various VLANs behind a core router.



Identify Organizational Threats (Executive Perspective)

- Fear 1: “All of our customers credit card information getting stolen. We will have to pay huge fines, our customers won’t trust us anymore, and business will suffer.”
- Fear 2: “Something bad happens to our website causing it to be inaccessible for an extended time. At a certain point, this might threaten the continuity of the business.”
- Fear 3: “An individual finds a bug that allows them to place orders on the website without paying for them. This could result in lost revenues.”

Actual Threats

- Theft of Customer PII (Confidentiality)
- Disruption of E-Commerce Service (Availability)
- Unintended Use of E-Commerce Service (Integrity)

Quantify Risk

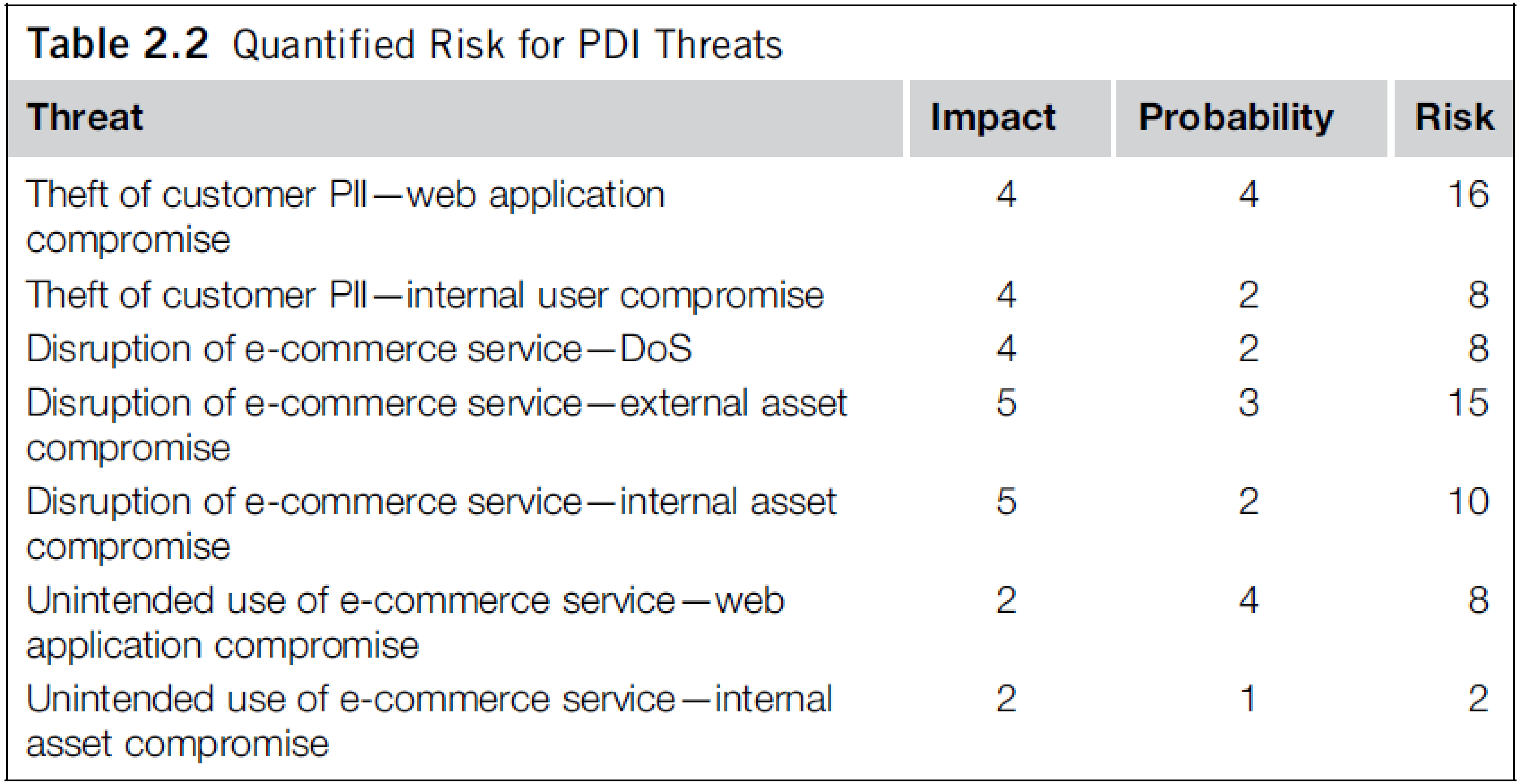


Table 2.2 Quantified Risk for PDI Threats

Threat	Impact	Probability	Risk
Theft of customer PII—web application compromise	4	4	16
Theft of customer PII—internal user compromise	4	2	8
Disruption of e-commerce service—DoS	4	2	8
Disruption of e-commerce service—external asset compromise	5	3	15
Disruption of e-commerce service—internal asset compromise	5	2	10
Unintended use of e-commerce service—web application compromise	2	4	8
Unintended use of e-commerce service—internal asset compromise	2	1	2

Quantify Risk Cont.

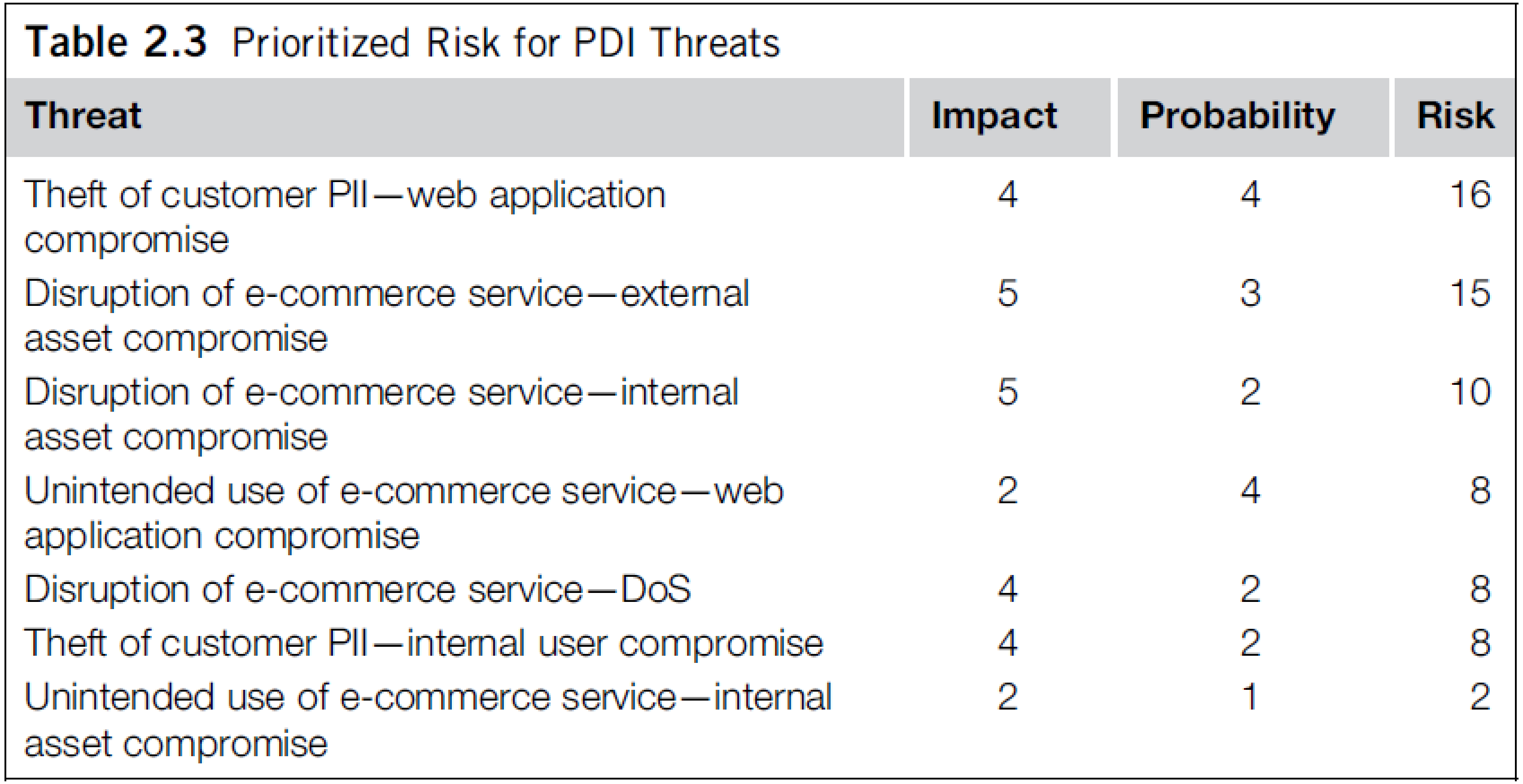


Table 2.3 Prioritized Risk for PDI Threats

Threat	Impact	Probability	Risk
Theft of customer PII—web application compromise	4	4	16
Disruption of e-commerce service—external asset compromise	5	3	15
Disruption of e-commerce service—internal asset compromise	5	2	10
Unintended use of e-commerce service—web application compromise	2	4	8
Disruption of e-commerce service—DoS	4	2	8
Theft of customer PII—internal user compromise	4	2	8
Unintended use of e-commerce service—internal asset compromise	2	1	2

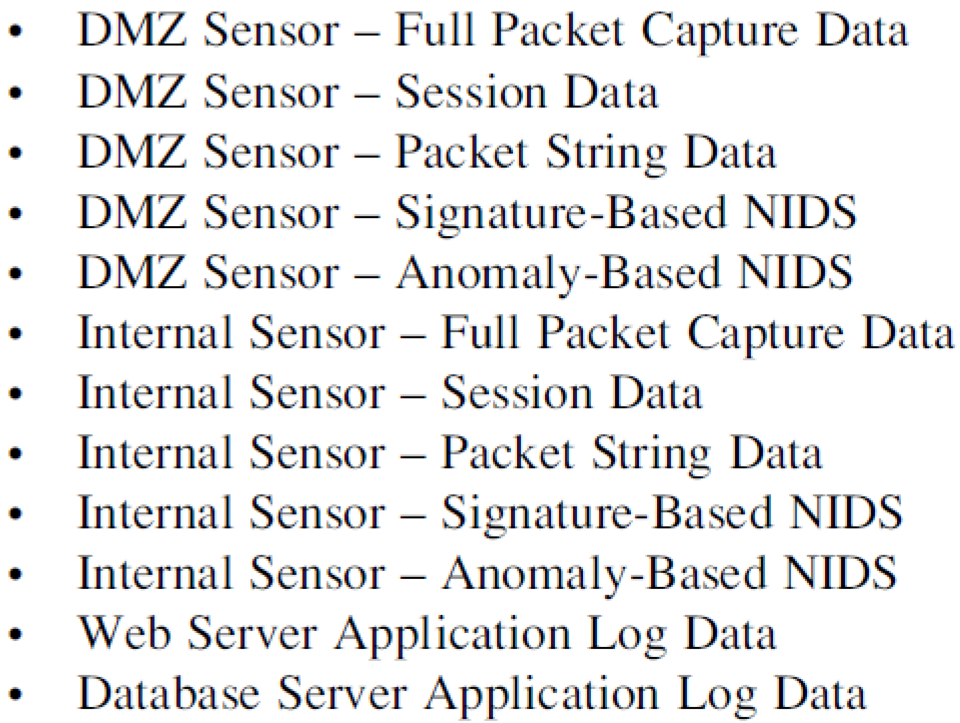
Theft of Customer PII – Web Application Compromise

- The threat presenting the most risk to the organization is customer PII being stolen as a result of a web application compromise.
- This presents a potentially large attack surface from the perspective of the web application, but a rather small attack surface from the perspective of network assets.

What should we monitor?

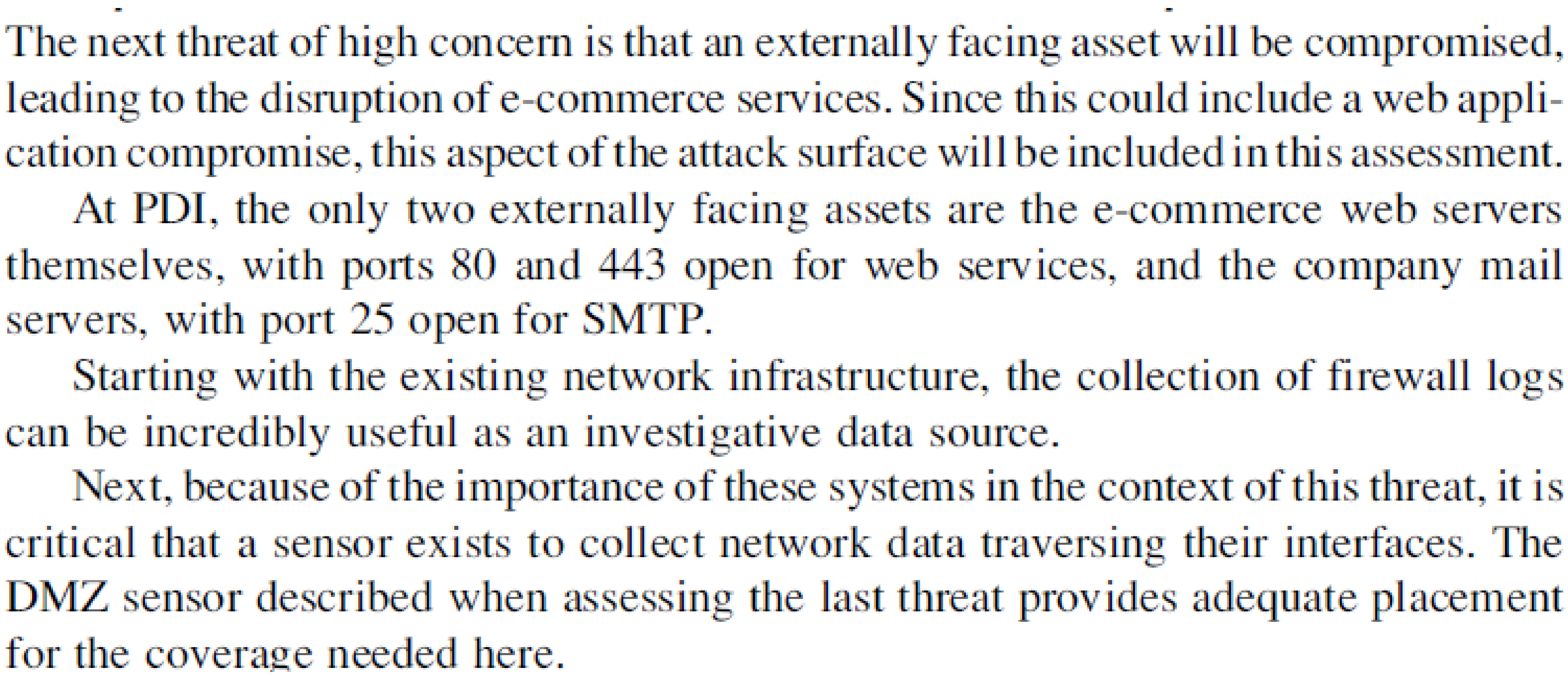
- A sensor can be placed at the network edge to collect full packet capture data, session data, or packet string data. This will also allow for the use of signature and anomaly-based NIDS.
- Actions of the web server by collecting its application-specific log data.
- The database server resides in the internal network, so this will require a second sensor placed so that it has visibility here.
- Again, this provides for collection of full packet capture data, session data, and packet string data, and allows the use of signature and anomaly-based NIDS.
- Finally, the database server will likely generate its own application-specific logs that can provide visibility into its actions.

Data sources based on this plan?



- DMZ Sensor – Full Packet Capture Data
- DMZ Sensor – Session Data
- DMZ Sensor – Packet String Data
- DMZ Sensor – Signature-Based NIDS
- DMZ Sensor – Anomaly-Based NIDS
- Internal Sensor – Full Packet Capture Data
- Internal Sensor – Session Data
- Internal Sensor – Packet String Data
- Internal Sensor – Signature-Based NIDS
- Internal Sensor – Anomaly-Based NIDS
- Web Server Application Log Data
- Database Server Application Log Data

Disruption of E-Commerce Server – External Asset Compromise



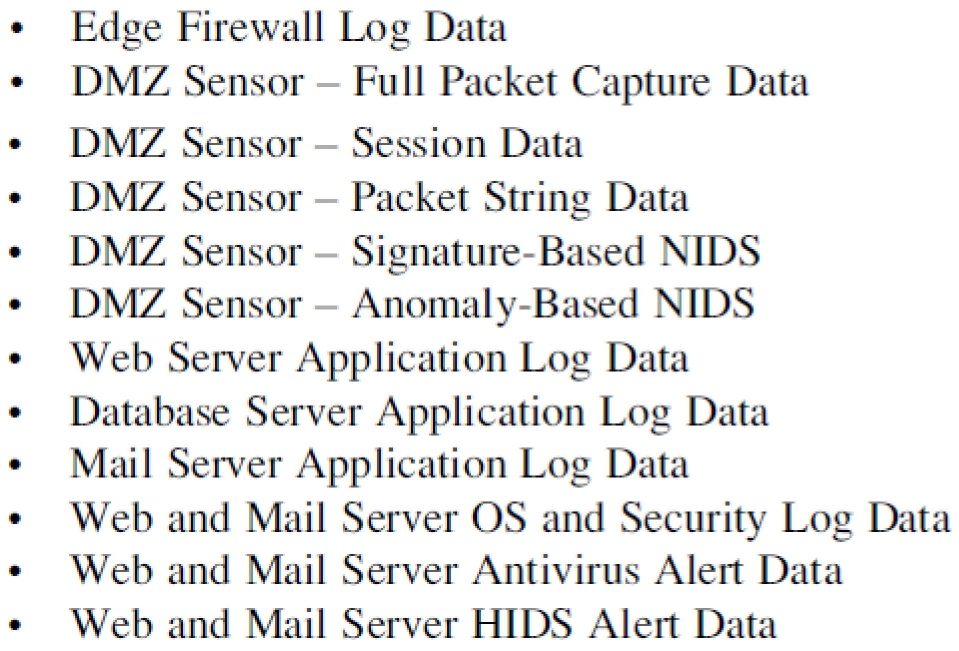
The next threat of high concern is that an externally facing asset will be compromised, leading to the disruption of e-commerce services. Since this could include a web application compromise, this aspect of the attack surface will be included in this assessment.

At PDI, the only two externally facing assets are the e-commerce web servers themselves, with ports 80 and 443 open for web services, and the company mail servers, with port 25 open for SMTP.

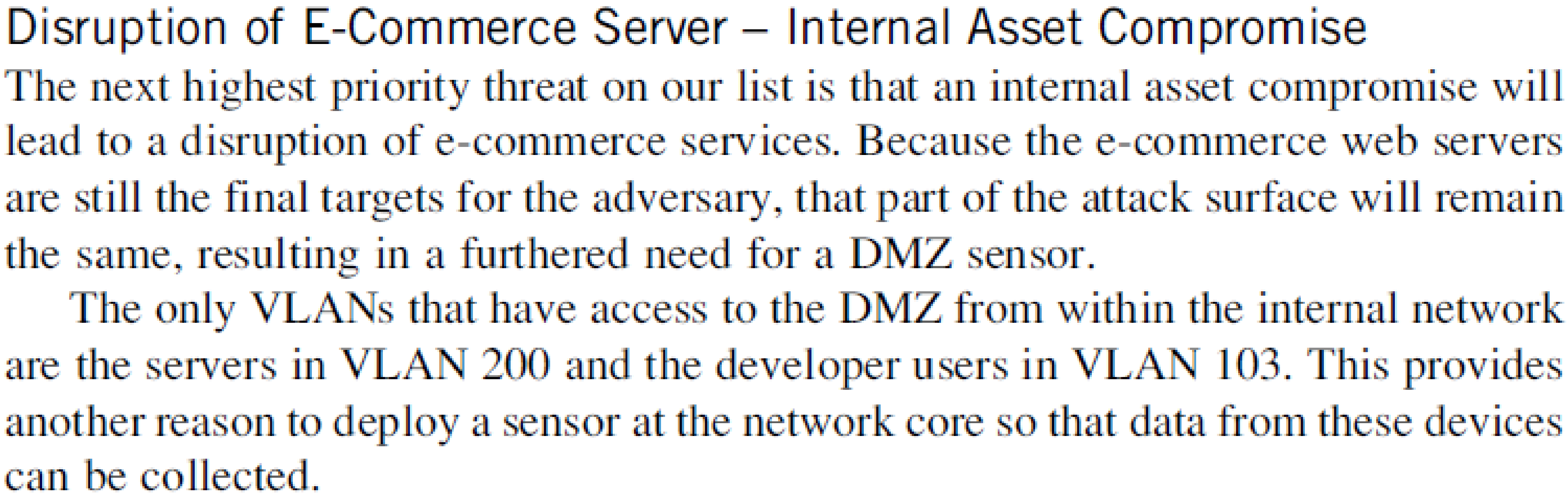
Starting with the existing network infrastructure, the collection of firewall logs can be incredibly useful as an investigative data source.

Next, because of the importance of these systems in the context of this threat, it is critical that a sensor exists to collect network data traversing their interfaces. The DMZ sensor described when assessing the last threat provides adequate placement for the coverage needed here.

Potential Data Source



- Edge Firewall Log Data
- DMZ Sensor – Full Packet Capture Data
- DMZ Sensor – Session Data
- DMZ Sensor – Packet String Data
- DMZ Sensor – Signature-Based NIDS
- DMZ Sensor – Anomaly-Based NIDS
- Web Server Application Log Data
- Database Server Application Log Data
- Mail Server Application Log Data
- Web and Mail Server OS and Security Log Data
- Web and Mail Server Antivirus Alert Data
- Web and Mail Server HIDS Alert Data



Disruption of E-Commerce Server – Internal Asset Compromise

The next highest priority threat on our list is that an internal asset compromise will lead to a disruption of e-commerce services. Because the e-commerce web servers are still the final targets for the adversary, that part of the attack surface will remain the same, resulting in a furthered need for a DMZ sensor.

The only VLANs that have access to the DMZ from within the internal network are the servers in VLAN 200 and the developer users in VLAN 103. This provides another reason to deploy a sensor at the network core so that data from these devices can be collected.

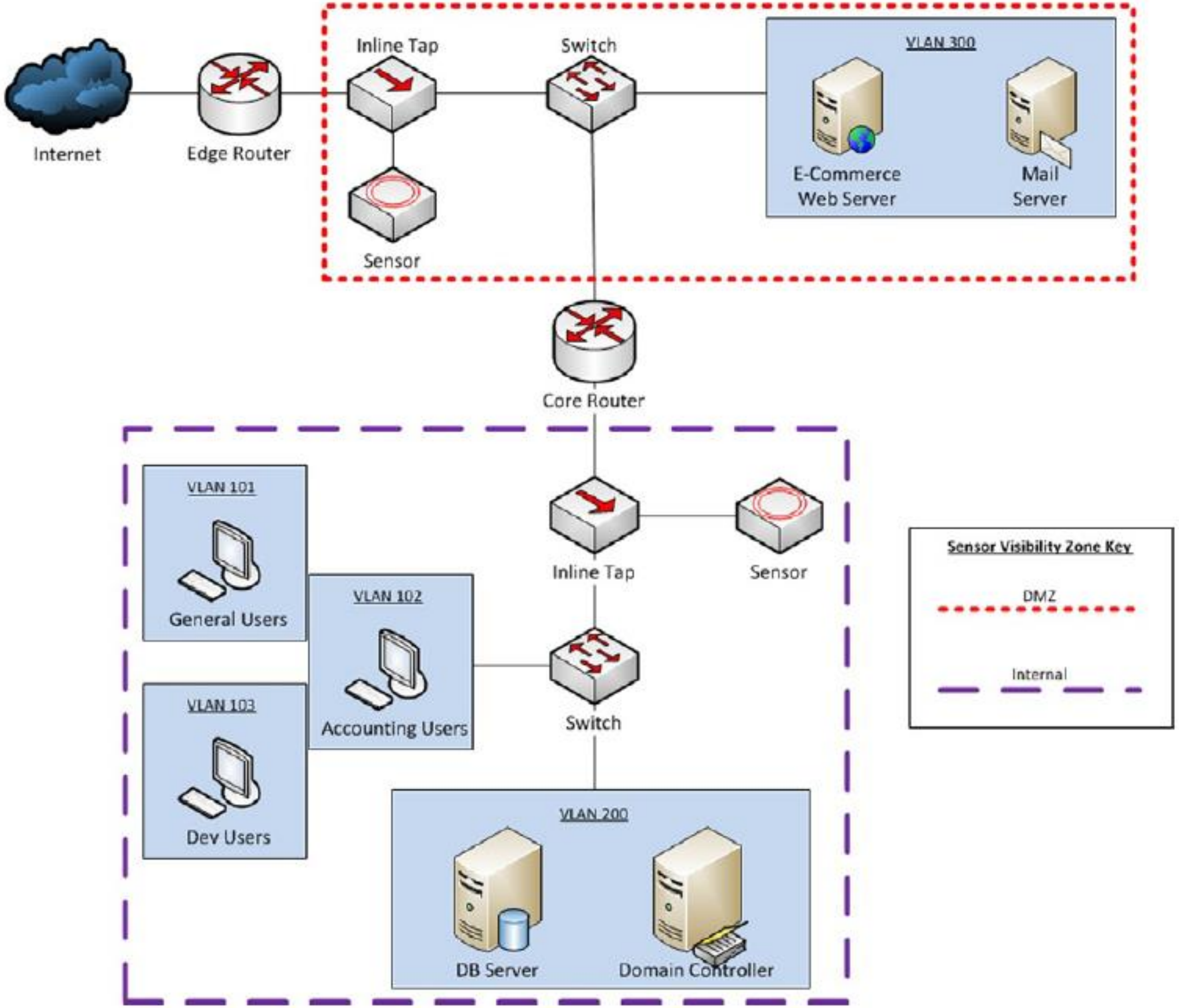
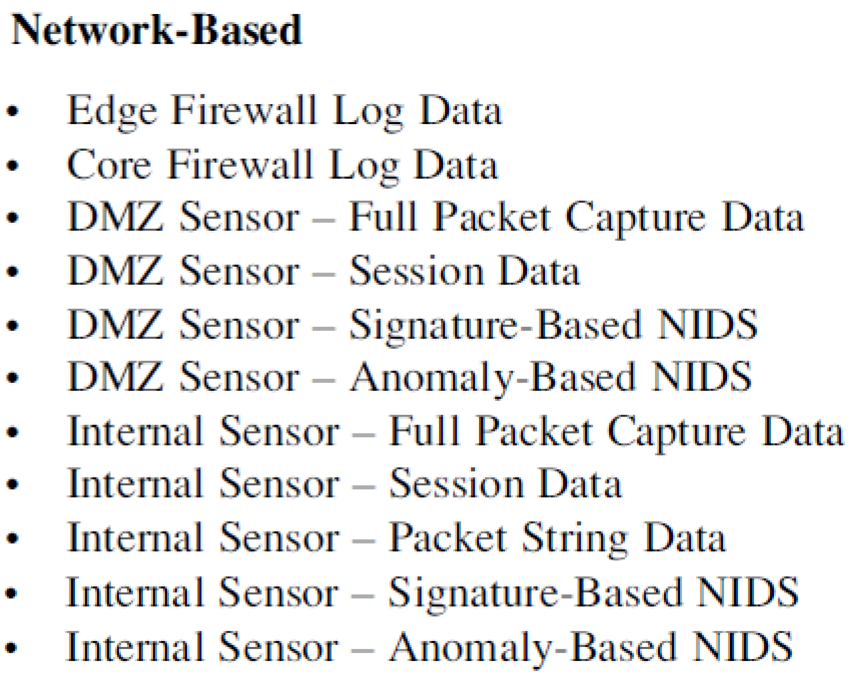


FIGURE 2.3

Updated Network Diagram Including Sensor Placement

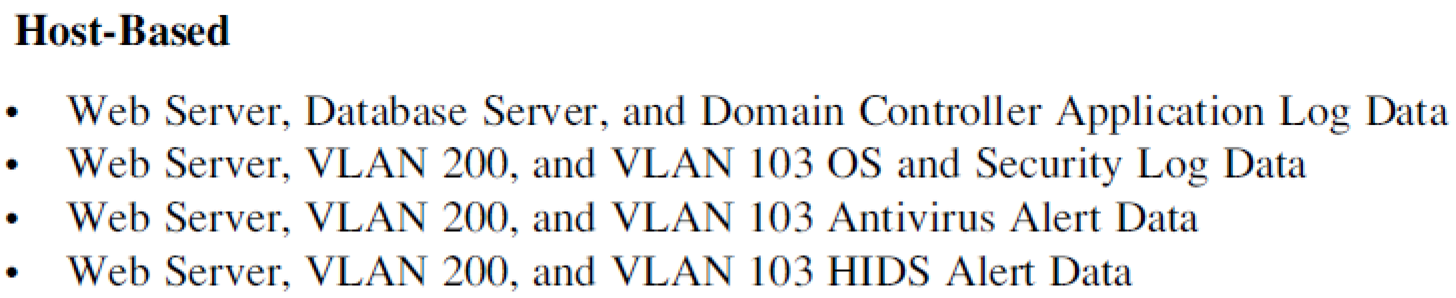
Data Sources



Network-Based

- Edge Firewall Log Data
- Core Firewall Log Data
- DMZ Sensor – Full Packet Capture Data
- DMZ Sensor – Session Data
- DMZ Sensor – Signature-Based NIDS
- DMZ Sensor – Anomaly-Based NIDS
- Internal Sensor – Full Packet Capture Data
- Internal Sensor – Session Data
- Internal Sensor – Packet String Data
- Internal Sensor – Signature-Based NIDS
- Internal Sensor – Anomaly-Based NIDS

Data Sources



Host-Based

- Web Server, Database Server, and Domain Controller Application Log Data
- Web Server, VLAN 200, and VLAN 103 OS and Security Log Data
- Web Server, VLAN 200, and VLAN 103 Antivirus Alert Data
- Web Server, VLAN 200, and VLAN 103 HIDS Alert Data