Chapter 2. Planning data collection

The Applied Collection Framework

> The ACF is a loose set of steps that help an organization evaluate what data sources should be the focus of their collection efforts.

It is a series of steps:

1. Define threats
2. Quantify risk
3. Identify data feeds
4. Narrow focus

Define threats

It is important to identify threats specific to the organization being analyzed.

It helps to frame these threats by whether they negatively impact confidentiality, integrity, or availability.

Once threats have been identified, it is up to security personnel to identify specific technologies possibly affected by the threat.

Finally, you should be able to identify which assets are the most critical.

Quantify risk

Once a list of potential threats has been identified, those threats must be prioritized.

```
Impact (I) X Probability (P) = Risk (R)
```

Impact takes into consideration how a given threat, should it manifest itself, could affect the organization.

Probability represents the likelihood that a threat will manifest itself. Over time, the probability of a vulnerability being exploited increases, so they should be revisited over time.

It is important to realize that these numbers are subjective, so the same people, or a committee should review them periodically.

Identify data feeds

Identify the primary data feeds that might provide NSM detection and analysis value.

During the previous definition phase, you should have identified the device's characteristics. (Location in the network, OS, services running, etc. . .)

Based upon this information, you can examine network-based and host-based data feeds.

Narrow focus

Now, we have to get intimately granular with the sources we have selected.

It is the most technical step of the process, and involves reviewing every data source individually to Gauge it's value.

There are sources that require too much work for the value they provide, so you have to decide weather it's collection is worth the time and money investment.

Specific questions that might arise in this step:

- What can you filter out of PCAP traffic from a specific network traffic.
- 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

You should also define the amount and duration of each data type you would like to retain.

The documents generated from this process are never finalized.

Case Scenario: On-line Retailer

We have a on-line retail store, that resells products they like. Their only income source is their website.

Identify Organizational Threats

1. All customers ID cards being stolen (Confidentiality) (Pay fine, and no customer trust)
2. Inaccessible website (Availability, loss in revenue)
3. Order without money transfer (Integrity)

Theft of Customer PII (Personally Identifiable Information) (Confidentiality)

- Database compromised through the website by an external person.
- Attacker compromises computer of internal team member with database access (Developer)

Disruption of E-commerce Service (Availability)

- DDOS
- Compromise externally facing asset, that renders the service unavailable
- Compromises internal network, pivots to the E-commerce servers, and renders servers unavailable.

Unintended use of E-commerce service (Integrity)

- Bug on website
- Compromised internal user with access to the backend

Quantify Risk

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

Figure 1: PDI Threats

Identify Data Feeds

Identify the data sources that are useful for NSM detection and analysis.

Theft of Customer PII - Web Application Compromise

Based on the analysis, done in previous phases, we can collect a series of data flows to inspect.

- Web server transactions with external users
- Server actions (collect application-specific log data)
- Indirect user access to back-end database.

Now, we know what to monitor, we have to know where to monitor.

- 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

Narrow Focus

Take the primary data sources that have been identified and refine those so that only useful aspects of that data are collected.