Decision Fusion using Dempster-Schaffer Theory

Prof. D. J. Parish
High Speed networks Group
Department of Electronic and Electrical Engineering
D.J.Parish@lboro.ac.uk

Loughborough University
Overview

- Introduction
- Why Fusion?
- Possible Approaches
  - Bayesian
  - Dempster-Schaffer Theory
    - Origin
    - Main Characteristics
- DS worked example
- DS Issues
  - Data Independence
  - Some issues with conflicting evidence
- Basic Belief Assignment
  - Possible Approaches
    - A “Light Weight” Approach
- Examples from Research
WHY DATA FUSION

- Multi-Metric or Cross-layer Anomaly Based IDSs outperform Single-metric detection results [3]

- Although there are cases in which IDSs that utilise information from a single metric might give good detection results, the presence of attacks is rarely accurately detectable by examining a single metric from one layer of the protocol stack.

- Multi-Metric IDSs combine information from two or more layers of the protocol stack

- The higher the number of metrics, the greater the chances to identify intrusions
Data fusion:
- Process of gathering information from multiple and heterogeneous sources and combining them towards obtaining a more accurate final result
- The most common data fusion techniques
  - Bayesian Theory
  - Dempster-Shafer (D-S) Theory of Evidence

**Bayesian Theory**
- Calculates the probability of occurrence of a certain event, based on the experience extracted from previous events
  - Previous event probabilities is very difficult or impossible to determine
  - Does not directly assign probability to *uncertainty*
Dempster-Shafer Theory

- From Wikipedia, the free encyclopedia

- The **Dempster–Shafer theory (DST)** is a mathematical theory of evidence.[1] It allows one to combine evidence from different sources and arrive at a degree of belief (represented by a belief function) that takes into account all the available evidence. The theory was first developed by Arthur P. Dempster[2] and Glenn Shafer.[1][3]

- In a narrow sense, the term **Dempster–Shafer theory** refers to the original conception of the theory by Dempster and Shafer. However, it is more common to use the term in the wider sense of the same general approach, as adapted to specific kinds of situations. In particular, many authors have proposed different rules for combining evidence, often with a view to handling conflicts in evidence better.
Dempster-Shafer (D-S) Theory of Evidence
- Mathematical discipline that combines evidences of information from multiple events to calculate the belief of occurrence of another event

- PROS:
  - High potential for managing Uncertainty
  - Assigns probability to Uncertainty
  - Does not require a priori knowledge
  - Suitable for detecting previously unknown attacks

- CONS:
  - Computation complexity increases exponentially with the number of possible event outcomes
  - Conflicting beliefs management assigning empty belief value
  - Evidences should be completely independent

A comparative study of different data fusion methods is presented in [3]
This work concludes that D-S theory is more promising than Bayesian in tasks of IDS
Frame of Discernment $\Theta = \{\Theta_1, \Theta_2, \ldots, \Theta_n\}$

- Finite set of all possible mutually exclusive outcomes about some problem domain
- All the observers must use the same frame of discernment

$2^\Theta$, refers to every possible mutually exclusive subset of the elements of $\Theta$
- If $\Theta = \{\text{Attack, Normal}\}$, then $2^\Theta = \{\text{Attack, Normal, Uncertainty, } \emptyset\}$

- Each subset is defined as an Hypothesis and receives a belief value within $[0, 1]$
- Assignment is known as the Basic Probability Assignment (BPA)

\[
m : 2^\Theta \rightarrow [0, 1] \quad \text{if} \quad \begin{cases} 
m(\emptyset) = 0 \\
m(A) \geq 0, \ \forall A \subseteq \Theta \\
\sum_{A \subseteq \Theta} m(A) = 1\end{cases}
\]
From the mass assignments, the upper and lower bounds of a probability interval can be defined. This interval contains the precise probability of a set of interest (in the classical sense), and is bounded by two non-additive continuous measures called belief (or support) and plausibility:

- The belief $\text{bel}(A)$ for a set $A$ is defined as the sum of all the masses of subsets of the set of interest:
- The plausibility $\text{pl}(A)$ is the sum of all the masses of the sets $B$ that intersect the set of interest $A$:
- The two measures are related to each other as follows:
- And conversely, for finite $A$, given the belief measure $\text{bel}(B)$ for all subsets $B$ of $A$, we can find the masses $m(A)$ with the following inverse function:
- where $|A - B|$ is the difference of the cardinalities of the two sets.[4]
### DEMPSTER-SHAFER - EXAMPLE

#### Sensor 1

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Attack</td>
<td>0.32</td>
</tr>
<tr>
<td>Normal</td>
<td>0.25</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>0.43</td>
</tr>
</tbody>
</table>

\[
m(E) = \sum_{X \cap Y = E} m_{1}(X) \cdot m_{2}(Y) / 1 - \sum_{X \cap Y = \emptyset} m_{1}(X) \cdot m_{2}(Y) \quad \forall E \neq \emptyset
\]

#### Sensor 2

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Attack</td>
<td>0.35</td>
</tr>
<tr>
<td>Normal</td>
<td>0.1</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>0.55</td>
</tr>
</tbody>
</table>

\[
\begin{array}{c|ccc}
   m_{2} & m_{1} & \{A\} & \{N\} & \{A, N\} \\
   \{A\} & 0.35 & 0.11 & 0.09 & 0.15 \\
   \{N\} & 0.1 & 0.03 & 0.025 & 0.04 \\
   \{A, N\} & 0.55 & 0.18 & 0.14 & 0.24 \\
\end{array}
\]

\[
m(A) = 1.136 \times (0.11 + 0.15 + 0.18) = 0.5 \\
m(N) = 1.136 \times (0.025 + 0.04 + 0.14) = 0.233 \\
m(A|N) = 1.136 \times (0.24) = 0.272
\]
DEMPSTER-SHAFER - EXAMPLE

### Sensor 1

<table>
<thead>
<tr>
<th></th>
<th>Attack</th>
<th>Normal</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attack</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncertainty</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Sensor 2

<table>
<thead>
<tr>
<th></th>
<th>Attack</th>
<th>Normal</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attack</td>
<td>0.35</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncertainty</td>
<td>0.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
m(E) = \sum_{X \cap Y = E} m_1(X) \cdot m_2(Y) / \sum_{X \cap Y = \emptyset} m_1(X) \cdot m_2(Y) \forall E \neq \emptyset
\]

<table>
<thead>
<tr>
<th>( m_2 \setminus m_1 )</th>
<th>{A}: 0.32</th>
<th>{N}: 0.25</th>
<th>{A, N}: 0.43</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}</td>
<td>0.11</td>
<td>0.09</td>
<td>0.15</td>
</tr>
<tr>
<td>{N}</td>
<td>0.03</td>
<td>0.025</td>
<td>0.04</td>
</tr>
<tr>
<td>{A, N}</td>
<td>0.18</td>
<td>0.14</td>
<td>0.24</td>
</tr>
</tbody>
</table>

\[
m(A) = 1.136 \times (0.11 + 0.15 + 0.18) = 0.5
\]

\[
m(N) = 1.136 \times (0.025 + 0.04 + 0.14) = 0.233
\]

\[
m(A|N) = 1.136 \times (0.24) = 0.272
\]
DEMPSTER-SHAFER - EXAMPLE

### Sensor 1

<table>
<thead>
<tr>
<th></th>
<th>Attack</th>
<th>Normal</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.32</td>
<td>0.25</td>
<td>0.43</td>
</tr>
</tbody>
</table>

### Sensor 2

<table>
<thead>
<tr>
<th></th>
<th>Attack</th>
<th>Normal</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.35</td>
<td>0.1</td>
<td>0.55</td>
</tr>
</tbody>
</table>

\[
m(E) = \sum X \cap Y = E \uparrow \downarrow m_1 (X) \ast m_2 (Y) / 1 - \sum X \cap Y = \emptyset \uparrow \downarrow m_1 (X) \ast m_2 (Y) \quad \forall E \neq \emptyset
\]

\[
m(A) = 1.136 \ast (0.11 + 0.15 + 0.18) = 0.5
\]

\[
m(N) = 1.136 \ast (0.025 + 0.04 + 0.14) = 0.233
\]

\[
m(A|N) = 1.136 \ast (0.24) = 0.272
\]
DEMPSTER-SHAFER - EXAMPLE

Sensor 1

| Attack   | 0.32 |
| Normal   | 0.25 |
| Uncertainty | 0.43 |

Sensor 2

| Attack   | 0.35 |
| Normal   | 0.1  |
| Uncertainty | 0.55 |

\[
m(E) = \sum_{X \cap Y = E \uparrow} m_1(X) \cdot m_2(Y) / 1 - \sum_{X \cap Y = \emptyset \uparrow} m_1(X) \cdot m_2(Y) \quad \forall E \neq \emptyset
\]

<table>
<thead>
<tr>
<th>(m_2\backslash m_1)</th>
<th>{A}: 0.32</th>
<th>{N}: 0.25</th>
<th>{A, N}: 0.43</th>
</tr>
</thead>
<tbody>
<tr>
<td>{A}: 0.35</td>
<td>0.11</td>
<td>0.09</td>
<td>0.15</td>
</tr>
<tr>
<td>{N}: 0.1</td>
<td>0.03</td>
<td>0.025</td>
<td>0.04</td>
</tr>
<tr>
<td>{A, N}: 0.55</td>
<td>0.18</td>
<td>0.14</td>
<td>0.24</td>
</tr>
</tbody>
</table>

\[
m(A) = 1.136 \cdot (0.11 + 0.15 + 0.18) = 0.5
\]

\[
m(N) = 1.136 \cdot (0.025 + 0.04 + 0.14) = 0.233
\]

\[
m(A \mid N) = 1.136 \cdot (0.24) = 0.272
\]
### An example with more sensors

#### Hypothesis

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Hypothesis</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>0.3</td>
<td>0.217</td>
<td>0.667</td>
<td>0.667</td>
<td>0.217</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>Attack</td>
<td>0.4</td>
<td>0.567</td>
<td>0.167</td>
<td>0.167</td>
<td>0.567</td>
<td>0.567</td>
</tr>
<tr>
<td></td>
<td>Uncertainty</td>
<td>0.3</td>
<td>0.216</td>
<td>0.166</td>
<td>0.166</td>
<td>0.216</td>
<td>0.216</td>
</tr>
</tbody>
</table>

#### Iteration

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Hypothesis</th>
<th>#1 - #2</th>
<th>R - #3</th>
<th>R - #4</th>
<th>R - #4</th>
<th>Final Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>0.262</td>
<td>0.857</td>
<td>0.187</td>
<td>0.746</td>
<td>0.475</td>
</tr>
<tr>
<td></td>
<td>Attack</td>
<td>0.65</td>
<td>0.107</td>
<td>0.751</td>
<td>0.247</td>
<td>0.524</td>
</tr>
<tr>
<td></td>
<td>Uncertainty</td>
<td>0.088</td>
<td>0.036</td>
<td>0.062</td>
<td>0.007</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Current Techniques
- Empirical approach
- Expert opinion
- Manually assignment
- Fixed Thresholds
- Fixed Scales
- Fixed Linear functions
  - Unable to automatically adapt without IDS administrator
- Data Mining techniques
  - Require Gathering data, Processing, Training, Perform analysis, etc.
  - Unable to automatically adapt in Real-Time
We proposed a novel BPA methodology [4]

- Three independent Statistical approaches
- Automatically adapt detection capabilities
- No intervention from IDS administrator
- Light weight profiling process
- Tested with diverse number of Wireless Network Attacks
Sliding window of ~30 frames

- If current frame is Legal → Slides
- If current frame is Malicious → Drops the frame
Degree of dispersion of Dataset

- Similar to Boxplot method
- Quartiles define the scales boundaries
- Length of scales varies
  - Automatically adjust to the network behaviour changes
BASIC PROBABILITY ASSIGNMENT - BELIEF IN ATTACK - ANGLE

- Frequency and Euclidean Distance
- Mean or Mode - Reference point
- Angle $\alpha$ - Reference of maximum belief
- Angle $\beta$ - Reference of belief to current analysed frame
- Linear function between $\alpha$ and $\beta$ generates belief in Attack
Belief in *Uncertainty* is used as adjustment value

- Provisional *Uncertainty* value:
  \[
  Belief_{(Unc.)} = \frac{0.5 \cdot \text{Belief Min}}{\text{Belief Max}}
  \]

- Condition of D-S Theory:
  \[
  \sum_{A \subseteq \Theta} m(A) = 1
  \]

- Adjustment value:
  \[
  \mu = \frac{(X - 1)}{3}
  \]

- \( X = \text{Summation of the three beliefs} \)
Potential Problems

- All data and sensors used by DS should be independent.
- This is difficult to achieve in practice and there is considerable literature to indicate that total independence is not always required in practice.
- Misleading results can be generated if there is contradictory evidence, or certain values are 0.

\[
\% = \{A,B,C\} \\
m_1 = \{A\} (0.99), \{B\} (0.01); \quad \{C\} (0) \\
m_2 = \{C\} (0.99), \{B\} (0.01); \quad \{A\} (0)
\]

\[m_1 + m_2 = \{B\} (1.0)\]

In some cases this may be sensible, in others it will not be.
Aims to use multiple metrics from different layers to improve abuse detection in Wireless networks.

- Data Fusion based on Dempster-Schaffer theory of evidence.
  - Data Mining approaches could be evaluated
Date from other Wireless Access Networks & locations

Cross Layer IDMS (Data Fusion/Event Correlation)

IDS Notifications
Mitigation Response(s)

Attack

OTHER (APPS)

TRANSPORT (TCP/UDP)

NETWORK (IP)

MAC

PHY
Methodology

Capture Packets

Get metrics:

Construct statistics (mode-avg)

Distance of metric from (mode/avg of metric)

Assign belief in attack for each metric

Fuse beliefs for each metric with Dempster-Shafer

RSSI       Most Volatile
RATE
TTL per flow
NAV
SEQ #       Least Volatile
Data Fusion

- Dempster-Shafer because:
  - Deals with uncertainty
  - No a priori knowledge
Man in the Middle (MitM)

- Takes advantage of lag time
- Injects its own content
## Man-In-The Middle Attack Results

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Type</th>
<th>%</th>
<th>Result %</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAV + SEQ</td>
<td>FN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>7/63</td>
<td>11.1</td>
</tr>
<tr>
<td>RSSI + NAV + SEQ</td>
<td>FN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>8/63</td>
<td>12.7</td>
</tr>
<tr>
<td>RSSI + TTL + RATE</td>
<td>FN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>All metrics</td>
<td>FN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>FP</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Rogue Access Point

1. Disassociates client

2. Responds to Probes Requests

Attacker BackTrack 4 Atheros Card

Monitor BackTrack 4 Atheros Card

Access Point AP

Client Atheros Card

INTERNET

INTERNET
## Rogue Access Point Attacks

<table>
<thead>
<tr>
<th>Method</th>
<th>Rate</th>
<th>ESSID Spoof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airbase</td>
<td>Fixed at 1 Mbps</td>
<td>No</td>
</tr>
<tr>
<td>Airbase -a</td>
<td>Fixed at 1 Mbps</td>
<td>Yes</td>
</tr>
<tr>
<td>Host AP</td>
<td>Normal Rate</td>
<td>No</td>
</tr>
</tbody>
</table>
## Rogue Access Point Results

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Type</th>
<th>Airbase</th>
<th>Airbase ESSID Spoof</th>
<th>HostAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAV + SEQ</td>
<td>Detected?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FP</td>
<td></td>
<td>0/405</td>
<td>0/246</td>
<td>0/57</td>
</tr>
<tr>
<td>RSSI + NAV + SEQ</td>
<td>Detected?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FP</td>
<td></td>
<td>35/405</td>
<td>2/246</td>
<td>3/57</td>
</tr>
<tr>
<td>RSSI + TTL + RATE</td>
<td>Detected?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>FP</td>
<td></td>
<td>100%</td>
<td>0/246</td>
<td>100%</td>
</tr>
<tr>
<td>All metrics</td>
<td>Detected?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>FP</td>
<td></td>
<td>0/405</td>
<td>0/246</td>
<td>0/57</td>
</tr>
</tbody>
</table>
## Benefits of Extra Metrics

<table>
<thead>
<tr>
<th>No. of Metrics</th>
<th>Beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attack</td>
</tr>
<tr>
<td>NAV-SEQ</td>
<td>0.569</td>
</tr>
<tr>
<td>RSSI - NAV - SEQ</td>
<td>0.664</td>
</tr>
<tr>
<td>RSSI - TTL - Rate</td>
<td>0.575</td>
</tr>
<tr>
<td>5 metrics</td>
<td>0.710</td>
</tr>
</tbody>
</table>
Summary and Conclusions