Chapter 6
Alert Management and Correlation

Alert management includes functions to cluster, merge and correlate alerts. The clustering and merging functions recognize alerts that correspond to the same occurrence of an attack and create a new alert that merges data contained in these various alerts. The correlation function can relate different alerts to build a big picture of the attack. The correlated alerts can also be used for cooperative intrusion detection and tracing an attack to its source.

6.1 Data Fusion

Data Fusion is the process of collecting information from multiple and possibly heterogeneous sources and combining them in order to get a more descriptive, intuitive and meaningful result[40]. According to Bass [2], the output of fusion-based IDSs are estimates of current security situation including the identity of a threat source, the malicious activity, attack rate and an assessment of the potential severity of the projected target.

Siaterlis et al. [40] propose the use of Dempster-Shafer’s Theory of Evidence (D-S theory for short) as the underlying data fusion model for creating a DDoS detection engine. The engine takes into account knowledge gathered from totally heterogeneous information sources. The prototype reported in this paper consists of two sensors: a Snort preprocessor-plug in and a SNMP data collector. The D-S inference engine fuses the knowledge collected from the reports of various sensors, in order to infer the state of the monitored network. Sensors try to leverage on what network operators empirically know as signs of flooding attacks. The implementation is evaluated against a UDP flooding attack in an academic research network.

D-S theory can be considered as extension of Bayesian inference [40], but it differs from the Bayesian inference in that it allows the explicit representation of ignorance and combination of evidence. Therefore, D-S theory is more suitable for complex systems whose states are hard to be modeled. In addition to the DDOS detection engine proposed by Siaterlis et al.[40], the “intrusion early warning model”
proposed by Zhai et al. [22], and the distributed intrusion detection system proposed by Wang et al.[53] are other two examples in which the D-S theory has been used for data fusion in intrusion detection. The mathematical framework of D-S theory is as follows:

- **Frame of discernment (FOD)** is a set $\Theta$ that consists of all possible, mutually exclusive system states $\theta_1, \theta_2, \ldots, \theta_n \in \Theta$. A hypothesis $H_i$ is a subset of $\Theta$, denoted by $2^\Theta$.

- **Basic Probability Assignment (bpa)** is a mass function $m$ which computes belief in hypothesis,

  \[ m : 2^\Theta \rightarrow [0, 1] \]

  \[ m(\emptyset) = 0, m(H) \geq 0, \sum_{H \subseteq \Theta} m(H) = 1 \]

- **Belief Function** is defined as function $Bel : 2^\Theta \rightarrow [0, 1]$, which describes the belief in a hypothesis $H$:

  \[ Bel(H) = \sum_{B \subseteq H} m(B) \]

- **Plausibility** represents the degree of reliability of a hypothesis, accordingly, the plausibility function is defined as:

  \[ Pl : 2^\Theta \rightarrow [0, 1] \]

  \[ Pl(H) = \sum_{B \cap H \neq \emptyset} m(B) \]

  The relationship between $Bel$ and $Pl$ can be expressed as follows:

  \[ Pl(H) = 1 - Bel(\overline{H}) \]

  Where $\overline{H}$ is the complement of $H$. $Bel(\overline{H})$ indicates the degree of doubt for hypothesis $H$.

- **Rule of Combination** in D-S theory defines a rule to combine independent evidences in order to obtain a new belief in a hypothesis.

  \[ m_{12}(H) = \frac{\sum_{B \cap C = H} m_1(B)m_2(C)}{\sum_{B \cap C \neq \emptyset} m_1(B)m_2(C)} \]

Defining FOD is the first important step when using D-S theory for intrusion detection, it is actually classifying the attacks [53, 22]. For example, \{Probe, DDOS, Worm, R2L, unknown\} and \{Normal, SYN-flood, UDP-flood, ICMP-flood\} [40] are two different attack classifications for FOD. The assignments of basic probabilities depend on the evidences that are available. The evidences are results of our observations for information provide by INFOSEC devices such as IDSs and SNMP collectors[22, 40]. Consequently, the final states of the system are then estimated based on the evidences to support the decision making.