23rd International Congress on Modelling and Simulation, Canberra, ACT, Australia, 1 to 6 December 2019
mssanz.org.au/modsim2019

Modelling time decay in uncertain implication rules of evidence theory with a focus on cyber attacks

E. El-Mahassni [a]

[a]*Cyber and Electronic Warfare Division, Defence Science and Technology Group, PO Box 1500, West Avenue, Edinburgh, SA, 5111*

Email: Edwin.El-Mahassni@dst.defence.gov.au

Abstract: Key to a successful organisation is the ability of its cyber security to respond in a timely manner. Hence, speed with regards to having up-to-date knowledge of the system state is paramount. However, this knowledge may not always be perfect or complete. A way of quantifying this uncertainty is through the use of a reasoning under uncertainty framework. One such framework is the theory of belief functions such as evidence theory or Dempster-Shafer theory. It has connections to other frameworks such as probability, possibility and imprecise probability theories. But, in cyber security, uncertainty modeling is not enough. A ubiquitous problem present in this domain is in modeling cyber attacks so that causal statements that follow a kill chain (a well-known paradigm to describe the phases of a cyber attack) can be accurately represented. One way this could be done is through implication rules. More precisely, a law of implication is a type of relationship between two statements or sentences. Hence, Dempster-Shafer Theory which incorporates implications rules with respect to time would be desirable in the cyber domain as a way to model cyber attacks. The aim of this paper is to be able to model the different stages of a cyber attack using Dempster-Shafer Theory that also incorporates a time parameter. Examples are also provided to illustrate the novelty of this work.

Keywords: *Cyber security, Dempster-Shafer Theory, time-decay models*

1 INTRODUCTION

Within the area of autonomous cyber operations, one of the emerging key factors is the need to do things quickly. Recent cyberattacks like Petya and Wannacrypt succeeded because of the speed and scope by which they were able to inflict damage (Simons, 2018). For instance, Petya caused 200M-300M USD damage in about an hour, spreading throughout the entire enterprise. Of course, this is not the first time a cyber attack has taken place. A well-known model for describing the steps required to carry about an effective cyber threat is Lockheed Martin's "kill chain" (Hutchins et al., 2011). It "describes phases of intrusions, mapping adversary kill chain indicators to defender courses of action, identifying patterns that link individual intrusions, mapping adversary kill chain indicators to defender courses of action" (Hutchins et al., 2011). From an attacker's perspective, there are 7 stages listed. These are: recoinnaissance, weaponization, delivery, exploitation, installation, command and control and actions on objectives. In this paper, any of the steps of the kill chain can be modelled but here we focus on the first, third and fourth elements of the kill chain. These can be described in the following manner:

- **Recoinnassance.** Investigating and analysing available information about the target in order to identify potential vulnerabilities.
- **Delivery.** Transmission of a weapon to the targeted environment.
- **Exploitation.** After the weapon is delivered to the victim host, exploitation triggers intruders' code.

In a fast moving environment like cyber, it would be unrealistic to expect that a defender will have perfect knowledge of the system. There might be ignorance and uncertainty as to the current state of the environment. Hence, a framework for reasoning under uncertainty like the theory of belief functions could be suitable for handling such ambiguities. The theory of belief functions, also referred to as evidence theory or Dempster-Shafer Theory, is a general framework for reasoning with uncertainty, with understood connections to other frameworks such as probability, possibility and imprecise probability theories. However, it would be incorrect to think it is a generalisation of Bayesian probability/fusion (Hoffman and Murphy, 1993; Dezert et al., 2012).A good summary of Dempster-Shafer Theory can be found in Rakowsky (2007). There we read "(Dempster-Shafer Theory) is well-known for its usefulness to express uncertain judgment of experts." Further, Dempster-Shafer Theory makes it possible to model: 1) Single pieces of evidence within single hypothesis relations or 2) Single pieces of evidence within multi hypotheses relations as uncertain assessments of a system in which exactly one hypothesis is objectively true. Dempster-Shafer Theory describes the subjective viewpoint as an assessment for an unknown objective fact (Rakowsky, 2007). Many have also noted that Dempster-Shafer theory is a generalisation of the Bayesian theory of probability.

Within the cyber kill chain outlined earlier, it is obvious that a step cannot be taken unless the previous one is executed. Dempster-Shafer Theory in its original form cannot represent causal relationships between events. One of the earliest papers to address this issue is Benavoli et al. (2008), through the use of implication rules. A law of implication is a type of relationship between two statements or sentences. If A and B represent statements then "if A then B" represents an uncertain implication rule with a certain degree of confidence with this statement holding true with a probability p such that $p \in [\alpha, 1]$. This is referred to as a one degree of freedom (*1dof*) uncertain implication rule. Hence, the implication rule is believed to be true with probability α and $1 - \alpha$ represents the uncertainty. *1dof* simply means that the uncertainty can be represented by means of a single parameter, i.e. α. Similarly, a 2 degree of freedom (*2dof* uncertain implication rule) can be defined as: "if A then B" which holds with probability p such that $p \in [\alpha, \beta]$, with $0 \leq \alpha \leq \beta \leq 1$. Thus, this implication rule is believed to be true with degree α. false with degree $1 - \beta$ and uncertain with degree $\beta - \alpha$. Hence, Dempster-Shafer Theory together with implications rules are an attractive option for modeling uncertainty within causal events. However, in cyber, as noted earlier, time is crucial in determining the ability to carry out attacks. Thus, in the cyber domain, in addition to having an uncertainty framework that models causal events, a question of relevance would be: how does the passage of time affect the credibility or reliability of information?

2 BASIC CONCEPTS OF DEMPSTER-SHAFER THEORY

The basic concepts of Dempster-Shafer Theory described here can also be found in Rakowsky (2007). The frame of discernment (FoD) is defined as the set of all possible states within the system. It is a model that describes the set of possible hypotheses and is typically denoted by Θ_D. For instance, its construction can be

given by $\Theta_D = \{\theta_1, \theta_2, \ldots, \theta_k\}$, where the $\theta_i, 1 \leq i \leq k$, are mutually exclusive and contained in the set labelled Θ_D.

The most common means of encoding evidence is via the *basic belief assignment* (bba). This can be defined in the following manner.

Definition 1. *Let a function m be defined on the power set of the frame of discernment Θ_D as follows:* $m : \wp(\Theta_D) \rightarrow [0,1], X \mapsto m(X)$, *where* $\sum_{X \subseteq \Theta_D} m(X) = 1$.

For a given set $X \subseteq \Theta_D$, the belief mass $m(X)$ represents the proportion of all relevant and available evidence that supports the claim that a particular element of Θ_D belongs to the set X, but to no particular subset of X.

Dempster's original rule of combination is used to combine evidence from two independent belief functions represented as basic belief assignments (bbas) m_1 and m_2. This combination rule is given by the following definition.

Definition 2. *Let m_1, m_2 be bbas over a frame of discernment Θ_D. Then the resulting belief function after combining both bbas is given by the following direct sum:* $(m_1 \oplus m_2)(X) = \frac{1}{(1-\kappa)} \sum_{Y \cap Z = X \in \wp(\Theta_D)} m_1(Y) m_2(Z)$, *with the conflict mass (which quantifies the amount of disagreement between both sets of bbas) given by* $\kappa = \sum_{Y \cap Z = \emptyset} m_1(Y) m_2(Z)$.

In Dempster's rule of combination, the empty set mass is zero, i.e. $m(\emptyset) = 0$ (Rakowsky, 2007). This means that Θ_D has to be complete and contains all possible hypotheses of the scenario listed (Rakowsky, 2007). Dempster's rule is commutative, i.e. $m_1 \oplus m_2 = m_2 \oplus m_1$ and associative, i.e. $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$, when the basic belief masses are compatible, or not in complete conflict. The commutative property means that the final result is not affected by the order of the bbas involved and the associative property means that the final result is not affected by the order of the operations that need to be performed. Two other useful encodings in one-to-one correspondence with the basic belief assignment m are the *belief* and *plausibility* functions which are defined as $\mathrm{bel}(A) = \sum_{0 \neq B \subseteq A} m(B)$ and $\mathrm{pl}(A) = \sum_{B \cap A \neq 0} m(B)$, respectively. For a given subset $A, \mathrm{bel(A)}$, quantifies the extent to which the evidence supports A, while $\mathrm{pl}(A)$ quantified the extent to which the evidence does not contradict A. In addition, there are two multivariate functions which will be used in this paper. A vacuous extension extends a given domain D to a larger domain D' and marginalisation projects bbas from a domain D to D' where $D \supseteq D'$ Benavoli et al. (2008).

3 Implication Rules in Evidence Theory

The Dempster-Shafer Theory framework on its own is unable to provide a conclusion or effect based on a preceding statement. This can be achieved through the use of implication rules which takes a statement in the form of premise and causally links it to an effect; representing this connection in a quantifiable manner which can then be used within the existing Dempster-Shafer Theory framework (Ristic and Smets, 2005; Benavoli et al., 2007; Almond, 1995; Benavoli et al., 2008).

In determining whether the methodology for constructing implications rules is valid, there are three axioms which must be considered — reflexivity ($A \Rightarrow A$ is always true (tautology)), transitivity (if $A \Rightarrow B$ and $B \Rightarrow C$, then $A \Rightarrow C$) and contrapositivity ($A \Rightarrow B = \bar{B} \Rightarrow \bar{A}$).

There has been some previous research in the development and calculation of implication rules in evidence theory. Transformations for converting *1dof* and *2dof* uncertain implication rules into the evidence theory framework have been presented in Ristic and Smets (2005); Benavoli et al. (2007); Almond (1995). However, according to Benavoli et al. (2008), the contrapositivity rule was not obeyed in (Almond, 1995). The *1dof* and the *2dof* uncertain implication rules can now be given (Benavoli et al., 2008).

Definition 3. *A 1dof implication rule, $A \Rightarrow B$ with confidence $[\alpha, 1]$, can be expressed as a bba function consisting of 2 focal sets on the joint domain $D_1 \cup D_2$.*

$$m^{D_1 \cup D_2}(C) = \begin{cases} \alpha, & \text{if} \quad C = (A \times B) \cup (\bar{A} \times \Theta_{D_2}) \\ 1 - \alpha, & \text{if} \quad C = \Theta_{D_1 \cup D_2} \end{cases}$$

where $\bar{A}$ is the complement of A in Θ_{D_1}.

Definition 4. *A 2dof implication rule, $A \Rightarrow B$, with confidence $[\alpha, \beta]$ can be expressed as a bba function*

consisting of 2 focal sets on the joint domain $D_1 \cup D_2$. This representation is defined in the following way:

$$m^{D_1 \cup D_2}(C) = \begin{cases} \alpha, & \text{if} \quad C = (A \times B) \cup (\bar{A} \times \Theta_{D_2}) \\ 1-\beta, & \text{if} \quad C = (A \times \bar{B}) \cup (\bar{A} \times \Theta_{D_2}) \\ \beta - \alpha, & \text{if} \quad C = \Theta_{D_1 \cup D_2} \end{cases}$$

where $\bar{A}$ is the complement of A in Θ_{D_1}, and accordingly $\bar{B}$ is the complement of B in Θ_{D_2}.

If there are two implication rules then note the following idea from Benavoli et al. (2008):

Lemma 1. *If $A_1 \Rightarrow A_2$ with with a probability $p_1 \in [\alpha_1, \beta_1]$ and $A2 \Rightarrow A_3$ with a probability $p_2 \in [\alpha_2, \beta_2]$ then $A_1 \Rightarrow A_3$ with probability $p \in [\alpha_1\alpha_2, 1-(1-\beta_1)(1-\beta_2)]$.*

The proof follows from the fact that the lower bound probability of both $A_1 \Rightarrow A_2$ and $A_2 \Rightarrow A_3$ being true is $\alpha_1\alpha_2$ and the lower bound probability of both being false is $(1-\beta_1)(1-\beta_2)$. Similarly, for N implication rules where $A_1 \Rightarrow A_2, \ldots, A_N \Rightarrow A_{N+1}$ we have that $A_1 \Rightarrow A_{N+1}$ with probability $[\alpha_1 \ldots \alpha_N, 1-(1-\beta_1)\ldots(1-\beta_N)]$.

4 TIME DECAY AND ENTROPY

Previously in Dempster-Shafer Theory, one way of modeling time decay is through exponential decay (Kurdej and Cheraoui, 2012). This decay occurs widely in the natural sciences and dynamic system theory. This is expressed in the following form: $N(t) = N_0 e^{-\lambda t}$, where λ is the decay rate, $N(t)$ is the quantity at time t, and $N_0 = N(0)$ is the initial quantity. An intuitive characteristic of exponential decay for many people is the time required for the decaying quantity to fall to one half of its initial value. This is called the half-life and is often denoted by $t_{1/2}$. This is given by $t_{1/2} = \frac{\ln(2)}{\lambda}$.

Entropy was originally introduced by Shannon (1948) to calculate the amount of information in a transmitted message, it is also used as a measure of uncertainty. Information entropy for a discrete set of probabilities $p_i, i = 1, ..., n$ is given by $H = -\sum_i p_i \ln\{p_i\}$. The minimum value of entropy is zero and is obtained when we have complete knowledge of a random variable, namely its probability density function is a delta function. The maximum value is $-\ln\{n\}$ where n is the number of hypotheses or elements in the frame of discernment.

The standard entropy definition operates over probabilities, so is not appropriate for the Dempster-Shafer Theory without some modification.

An entropy measure based on the plausibility transformation Cobb and Shenoy (2006) has been proposed in Jirousek and Shenoy (2018). For a frame of discernment $\Theta = \{x_1, x_2, \ldots, x_n\}$, the plausibility transformation is defined as $P^{\text{pl}}(x_i) = K^{-1}\text{pl}(x_i)$, for $i = 1, \ldots, n$ and where $K = \sum_{i,=1,\ldots,n}^{x_i \in \Theta,} \text{pl}(x_i)$. This new measure satisfies many properties that would be desired, such as: consistency with DS theory semantics, non-negativity, monotonicity, probability consistence and additivity (on a weak basis). The formula for this entropy measure is given by $H' = \sum_{X \in \Theta_D} P^{\text{pl}}(X) \ln\left\{\frac{1}{P^{\text{pl}}(X)}\right\} + \sum_{Y \in 2^{\Theta_D}} m(Y) \ln\{|Y|\}$.

4.1 Time Decay within Dempster-Shafer Theory

Background. In this section, we aim to propose a framework for describing confidence values within the Dempster-Shafer framework through the use of time decay. This is not the first time that time decay has been implemented in Dempster-Shafer Theory. In (Kurdej and Cheraoui, 2012), three discounting methods were proposed with exponential time decay being used in the examples. In this paper, the concept of time decay, in particular exponential time decay, is extended further to include implication rules. The motivation is as follows: in cyber, speed is critical and the confidence of a belief should be correlated to when the information was obtained. Further, it is also desirable to model causal relationships so that cyber attacks can be accurately represented. That is, at time $t = 0$, the instantaneous moment when a sensor reading was obtained, the bba for a certain non-trivial proposition should be at its maximum value. However, over time, the confidence placed in this value should decrease and instead be allocated or moved towards the ignorance bba or, at the very least, a bba where the cardinality of the focal element is greater. As time passes, it would also be useful to quantify the entropy of such a system. Intuitively we would expect the entropy to increase as time, t, passes. This is because an increase in time also leads to an increase in the ignorance mass while every other single mass decreases in value.

Time Decay in Uncertain Implication Rules. The idea of time decay explained above can now be extended to uncertain implication rules in Dempster-Shafer Theory. The principle is the same: as time increases, the mass of the universal set increases at the expense of those masses with non-trivial elements from the frame of discernment. Hence, the following definitions are given.

Definition 5. *A 1dof implication rule, $A \Rightarrow B$ with confidence $[\alpha, 1]$ can be expressed as a bba function consisting of 2 focal sets on the joint domain $D_1 \cup D_2$.*

$$m^{D_1 \cup D_2}(C) = \begin{cases} \alpha \mathrm{e}^{-\lambda t} & , \quad \text{if} \quad C = (A \times B) \cup (\bar{A} \times \Theta_{D_2}) \\ 1 - \mathrm{e}^{-\lambda t} + (1-\alpha)\mathrm{e}^{\lambda t}, & \quad \text{if} \quad C = \Theta_{D_1 \cup D_2} \end{cases}$$

where $\bar{A}$ is the complement of A in Θ_{D_1} and λ is the decay rate.

Definition 6. *A 2dof implication rule, $A \Rightarrow B$ with confidence $[\alpha, \beta]$ can be expressed as a bba function consisting of 2 focal sets on the joint domain $D_1 \cup D_2$. This representation is defined in the following way:*

$$m^{D_1 \cup D_2}(C) = \begin{cases} \alpha \mathrm{e}^{-\lambda t}, & \text{if} \quad C = (A \times B) \cup (\bar{A} \times \Theta_{D_2}) \\ (1-\beta)\mathrm{e}^{-\lambda t}, & \text{if} \quad C = (\bar{A} \times \bar{B}) \cup (A \times \Theta_{D_2}) \\ 1 - \mathrm{e}^{-\lambda t} + (\beta - \alpha)\mathrm{e}^{-\lambda t}, & \text{if} \quad C = \Theta_{D_1 \cup D_2} \end{cases}$$

where $\bar{A}$ is the complement of A in Θ_{D_1}, $\bar{B}$ is the complement of B in Θ_{D_2} and λ is the decay rate.

With this definition, it can be seen that over time (as t increases) then $\alpha \to \alpha \mathrm{e}^{-\lambda t}$ and $\beta \to 1 - (1-\beta)\mathrm{e}^{-\lambda t}$.

In Lemma 1, we described how two implications rules are combined. In Lemma 2, we extend this idea to include a time parameter.

Lemma 2. *If $A_1 \Rightarrow A_2$ with with a probability $p_1 \in [\alpha_1, \beta_1]$ with decay rate λ_1 and $A_2 \Rightarrow A_3$ with a probability $p_2 \in [\alpha_2, \beta_2]$ and decay rate λ_2 then $A_1 \Rightarrow A_3$ with probability*
$p \in [\alpha_1 \alpha_2 \mathrm{e}^{-(\lambda_1+\lambda_2)t}, 1 - (1-\beta_1)(1-\beta_2)\mathrm{e}^{-(\lambda_1+\lambda_2)t}]$.

The idea here is that as t increases then the implication rule approaches a trivial result. That is, the lower bound approaches 0 and the upper bound approaches 1. Now, the result of the lower bound is rather easy to see. For the upper bound, we note that as before $(1-\beta_1)(1-\beta_2)$ is the probability of both implications rules not being true. When time decay rates are introduced we see that as t increases then $(1-\beta_1)(1-\beta_2)\mathrm{e}^{-(\lambda_1+\lambda_2)t}$ decreases and so the upper bound which is given by $1 - (1-\beta_1)(1-\beta_2)\mathrm{e}^{-(\lambda_1+\lambda_2)t}$ also increases.

5 EXAMPLES

In this section, a couple of examples focused on cyber attacks will be demonstrated. The first uses a *1dof* implication rule. It involves the concept of an attacker carrying out a *reconnaissance* and then this implies the *delivery* of weapon to a targeted environment. The second example is more complex and involves the stages of reconnaissance, delivery and exploitation. For each example, we also demonstrate that as time increases, the overall entropy also increases which demonstrates the intuitive notion that uncertainty also increases over time.

5.1 Example 1

At any given moment, there is a warning system that alerts an administrator that a computer might be surveyed for a cyber attack. As time progresses, and in the absence of any updates, the administrator is less certain that such a survey is being conducted. From an attacker's perspective, the intent of such a survey is to find or detect a vulnerability which may then be exploited. However, again, as time progresses, the lack of a forthcoming attack suggests that it becomes less certain that a vulnerability could be found.

This problem is now encoded using Dempster-Shafer Theory in the following manner. Let $\Theta_R = \{r, \bar{r}\}$ denote the frame of discernment of whether an attacker is conducting a *reconnaissance* or not respectively and let $\Theta_D = \{d, \bar{d}\}$ describe the frame of discernment for delivering a weapon *delivery*. The confidences expressed as masses by an attacker doing a survey are $m_1(r) = 0.8, m_1(r, \bar{r}) = 0.2$. An administrator in charge of protecting the network knows that there is a chance that the attacker might find a vulnerability. For instance, consider the following example: if there is reconnaissance being done then a weapon which will be delivered targeting a vulnerability has a confidence of at least 0.4. However, in the absence of any further updates, the confidence decreases with respect to time according to some rate decay λ.

Using the formula given in Definition 5, this rule can be converted into the following bba functions: $m_2((r,d),(\bar{r},d),(\bar{r},\bar{d})) = 0.4\mathrm{e}^{-\lambda_d t}, m_2((r,d),(\bar{r},d),(r,\bar{d}),(\bar{r},\bar{d})) = (1-\mathrm{e}^{-\lambda_d t})+0.6\mathrm{e}^{-\lambda_d t} = 1-0.4\mathrm{e}^{-\lambda_d t}$. For the case when $t = 0$, we have $m_2((r,d),(\bar{r},d),(\bar{r},\bar{d})) = 0.4, m_2((r,d),(\bar{r},d),(r,\bar{d}),(\bar{r},\bar{d})) = 0.6$. Letting $m_1 \oplus m_2 = m_3$ and using the vacuous extension in Definition 4, we have $m_3((r,d)) = 0.32, m_3((r,d),(r,\bar{d})) = 0.48, m_3((r,d),(\bar{r},d),(\bar{r},\bar{d})) = 0.08, m_3((r,d),(\bar{r},d),(r,\bar{d}),(\bar{r},\bar{d})) = 0.12$. Marginalising this with respect to *recoinnassance* we are left with $m(d) = 0.32, m(d,\bar{d}) = 0.68$. At any time t, and performing similar calculations, these masses are simply: $m(d) = 0.32\mathrm{e}^{-\lambda_d t}, m(d,\bar{d}) = 1-0.32\mathrm{e}^{-\lambda_d t}$.

For this example,

$$H' = \frac{\ln\{(2-0c\mathrm{e}^{-\lambda t})\}}{(2-c\mathrm{e}^{-\lambda t})} + \frac{(1-0c\mathrm{e}^{-\lambda t})}{(2-c\mathrm{e}^{-\lambda t})} \ln\left\{\frac{(2-c\mathrm{e}^{-\lambda t})}{(1-c\mathrm{e}^{-\lambda t})}\right\} + (1-c\mathrm{e}^{-\lambda t})\ln\{2\}, \tag{1}$$

where $c = 0.32$ As t increases, we get $H' = \frac{1}{2}\ln\{2\} + \frac{1}{2}\ln\{2\} + \ln\{2\} = 2\ln\{2\} \approx 1.386$.

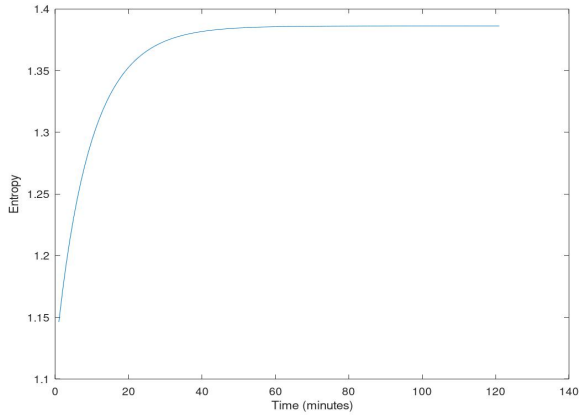

Figure 1. Entropy of Example 1 as a Function of Time for $\lambda = 0.1$

This diagram shows that entropy increases to a limit over time which is a desired property of the entropy function. Using the definition of belief, we note that at $t = 0, \mathrm{bel(v)} = 0.32$. This means that the lower probability of a vulnerability is 0.32. As time passes, then $m(v,\bar{v})$ tends to 1 which represents complete ignorance.

5.2 Example 2

This example is inspired from Case Study 2 in Centre (2015) which described a large UK company's internal network being infected by remote access malware. Also, recall the three stages listed earlier: reconnaissance, delivery and exploitation.

For this example, we will calculate the confidences of a vulnerability being found using some arbitrary numbers. Let the frame of discernments be $\Theta_R = \{r,\bar{r}\}$ for reconnaissance and $\Theta_D = \{d,\bar{d}\}$ for delivery. Just as in Example 1, there is confidence of 0.8 that reconnaissance is being conducted and that: i) if there is reconnaissance being carried out then a weaponized delivery will be sent with a confidence between 0.4 and 0.7 with decay rate $\lambda_1 = 0.7$ and ii) if there is a delivery then there will be an exploitation with a confidence between 0.2 and 0.3 with decay rate $\lambda_2 = 1$. Using Definition 6, we have: $m_2((r,d),(\bar{r},d),(\bar{r},\bar{d})) = 0.4\mathrm{e}^{-0.7t}$, $m_2((r,d),(r,\bar{d}),(\bar{r},\bar{d})) = 0.3\mathrm{e}^{-0.7t}$, $m_2((r,d),(\bar{r},d),(r,\bar{d}),(\bar{r},\bar{d})) = 1-0.7\mathrm{e}^{-0.7t}$, $m_3((d,e),(\bar{d},e),(\bar{d},\bar{e})) = 0.2\mathrm{e}^{-t}$, $m_3((d,e),(d,\bar{e}),(\bar{d},\bar{e})) = 0.7\mathrm{e}^{-t}$ and $m_3((d,e),(\bar{d},e),(d,\bar{e}),(\bar{d},\bar{e})) = 1-0.9\mathrm{e}^{-t}$. Letting the common frame of discernment be $\Theta_C = \{(r,d,e),(\bar{r},d,e),(r,\bar{d},e),(r,\bar{d},\bar{e}),(\bar{r},\bar{d},\bar{e}),(\bar{r},\bar{d},e),$
$(r,d,\bar{e}),(\bar{r},d,\bar{e})\}$ then we can combine the two bbas listed above, i.e. $(m_2 \oplus m_3)$. And, after combining those

fused masses with m_1 and then marginalising with respect to *reconnaissance* and *delivery*, we are left with $m(e) = 0.064\mathrm{e}^{-1.7t}, m(e, \bar{e}) = 1 - 0.064\mathrm{e}^{-1.7t}$.

Another way to derive this result is by using Lemma 2. Using the transitive property, we have that the statement: If there is reconnaissance being carried out then there will be an exploit with probability between $0.08\mathrm{e}^{-1.7t}$ and $1 - (1 - 0.3)(1 - 0.7)\mathrm{e}^{-(0.7+1)t} = 1 - 0.21\mathrm{e}^{-1.7t}$. Using Definition 6, we then have $m_4((r, e), (\bar{r}, e), (\bar{r}, \bar{e})) = 0.08\mathrm{e}^{-1.7t}$, $m_4((r, e), (r, \bar{e}), (\bar{r}, \bar{e})) = 0.21\mathrm{e}^{-1.7t}, m_4((r, e), (r, \bar{e}), (\bar{r}, e), (\bar{r}, \bar{e})) = 1-0.29\mathrm{e}^{-1.7t}$. This is then combined with m_1. Then, marginalising with respect to *recinnaissance* and *delivery* we get the same results as above; that is: $m(e) = 0.064\mathrm{e}^{-1.7t}$, $m(e, \bar{e}) = 1 - 0.064\mathrm{e}^{-1.7t}$. The entropy equation is the same as in (1), (with $c = 0.064$) with the entropy increasing to a limit over time which is a desired property of the entropy function. Using the definition of belief, we note that at $t = 0, \mathrm{bel}(\mathrm{v}) = 0.064$. This means that the lowerbound probability of a vulnerability is 0.064. As time passes, then $m(v, \bar{v})$ tends to 1 which represents complete ignorance.

6 Concluding Remarks

This is primarly a theoretical paper. It would be interesting to see how these theories would perform in more realistic scenarios which would include determining the value of λ. Also, although here we exclusively used the exponential decay, other types which might reflect a cyber system suddenly and rapidly decaying, as opposed to a smooth continuous function, could also be utilised. Dempster-Shafer Theory is an attractive framework for modeling uncertainty and ignorance. In addition, due to previous work, Dempster-Shafer Theory would allow us to model the complexity within a cyber environment through the use of implication rules. But, equally important within a cyber environment, timeliness of information is paramount. In this paper, we implemented a time-decay model within a Dempster-Shafer Theory framework that can be used with implications rules thereby allowing us to model in real-time the effect of time on data which slowly becomes outdated an its affect on our knowledge of the network. Future work could concentrate on the similarities or differences with other ways of modelling partial observability.

References

Almond, R. (1995). *Graphical Belief Modeling*. Chapman and Hall.

Benavoli, A., L. Chisi, A. Farina, and B. Ristic (2008). Modelling uncertain implication rules in evidence theory. In *Proc. 11th Intl. Conf. Information Fusion*, Cologne, Germany.

Benavoli, A., B. Ristic, A. Farina, M. Oxenham, and L. Chisi (2007). An approach to threat assessment based on evidential networks. In *Proc. 10th Intl. Conf. Information Fusion*, Quebec, Canda.

Centre, N. C. S. (2015). Common cyber attacks: Reducing the impact, cesg. *https://www.ncsc.gov.uk/guidance/white-papers/common-cyber-attacks-reducing-impact*.

Cobb, B. and P. Shenoy (2006). On the plausibility transformation method for translating belief function models to probability models. *Intl. J. Approx. Reasoning 41*(3), 313–330.

Dezert, J., A. Tchamova, D. Han, and J.-M. Tacnet (2012). Why dempster's fusion rule is not a generalization of bayes fusion rule. In *Proc. 16th Intl. Conf. Information Fusion*, Istanbul, Turkey.

Hoffman, J. and R. Murphy (1993). Comparison of bayesian and dempster-shafer theory of sensing: A practicioner's approach. In *SPIE Proc. on Neural and Stochastic Methods in Image and Signal Processing*, Volume II, pp. 266–279. SPIE.

Hutchins, E., M. Cloppert, and R. Amin (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *http://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf*.

Jirousek, R. and P. Shenoy (2018). A new definition of entropy of belief functions in the dempster-shafer theory. *Intl. J. Approx. Reasoning 92*(3), 49–65.

Kurdej, M. and V. Cheraoui (2012). Conservative, proportional and optimistic contextual discounting in the belief functions theory. In *Proc. 16th Intl. Conf. Information Fusion*, Istanbul, Turkey.

Rakowsky, U. (2007). Fundamentals of the dempster-shafer theory and its applications to system safety and reliability modelling. *Intl. J. of Reliability, Quality and Safety Engineering 14*(6), 579–601.

Ristic, B. and P. Smets (2005, Feb). Target identification using belief functions and implication rules. *IEEE Trans. on Aerospace and Electronic Systems 41*(3), 1097–1103.

Shannon, C. (1948). A mathematical theory of combination. *Bell System Technical Journal 27*, 379–423.

Simons, M. (2018, January). Overview of rapid cyberattacks. *http://www.microsoft.com/security/blog/2018/01/23/overview-of-rapid-cyberattacks/*.