On Aggregating Information in Actor Networks[*]

Stephan Olariu, Shahram Mohrehkesh, Xianping Wang and Michele C. Weigle
{olariu, smohrehk, wxianpin, mweigle}@cs.odu.edu
Department of Computer Science, Old Dominion University, Norfolk, VA

This paper provides a way to think formally about the aggregation processes that take place in networks where individual actors (whether sensors, robots, or people) possess data whose value may decay over time. The various actors use data to make decisions: the larger the value, the better (i.e. more informed) the decision. At every moment, individual actors have the choice of making a decision or else to defer the decision to a later time. However, the longer they wait, the lower the value of the data they hold. To counter-balance the effect of time discounting, we define an algebraic operation that we call aggregation, *whereby two or more actors integrate their data in the hope of increasing its value.*

Our main contribution is a formal look at the value of time-discounted information and at the algebra of its aggregation. We allow aggregation of time-discounted information to proceed in an arbitrary, not necessarily pairwise, manner. Our model relates aggregation decisions to the ensuing value of information and suggests natural thresholding strategies for the aggregation of the information collected by sets of network actors. Extensive simulations have confirmed the accuracy of our theoretical predictions.

I. Introduction

The designers of mission-oriented systems know that timely information aggregation is important, but they do not have an obvious way of measuring its value. There is a good reason for this: aggregation is a complex concept whose semantics are time-driven and context-dependent. Yet failure to aggregate information in a timely manner may have catastrophic consequences: terrorists penetrate our perimeters, natural disasters overwhelm local governments, and foreign hackers tamper with our national infrastructure.

The value of information has been a perennial topic of research in economics [11]. The economic argument is that information is a good that has value and hence can be traded [7]. Moreover, this value is often subject to deteriorating over time [2]. In the light of this, it is somewhat surprising that the dynamics of information deterioration over time has received very little attention in the literature [4].

It has long been recognized that assessing the value of time-discounted information is hard to evaluate being too dependent on subjective human valuation [3]. It is, thus, important to consider formal ways of defining the aggregation of time-discounted information so that aggregation might be better described and under-stood. This may lead to better designs for mission-oriented systems.

Against the backdrop of time pressure is the act of aggregation itself. While it is commonly accepted that the value of information often increases as a result of aggregation, almost nothing is known about the dynamics of aggregation and about the value of the resulting information, especially in the presence of time discounting.

The main contribution of this work is to provide a theoretical base upon which decisions about information aggregation can be made, measurements taken, and experiments performed. More specifically, our main contribution is a formal look at time-discounted information and at the algebra of its aggregation. In addition, this work suggests decision strategies for the aggregation of the information collected by sets of actors, whether sensors, robots, or people.

Our results find applications to fields in which hybrid networks of humans and machines need to be formed, such as mobile ad hoc and sensor network designs deployed in support of various missions. In both areas, there is a strong need to model not only aggregation, but also the dynamics of aggregation, as the links between different nodes shift in response to the actions of the network actors, who are in turn shaped by the mission at hand [8].

While our results have a wide applicability, we use sensor networks as an illustration of the concepts we discuss. This choice was suggested, in part, by the fact

that sensor networks are key ingredients in mission-oriented systems, and also by the fact that in these networks the value of the information often changes so rapidly as to defy deterministic modeling [6].

The remainder of the paper is organized as follows: Section II presents our motivating scenarios. We discuss our assumption about aggregation dynamics in Section III. Further, Section IV provides a formalism for time-discounted information. Section V investigates the aggregation operator independent of time discounting. Section VI introduces in a formal way the algebra of aggregation. Next, Section VII discusses aggregation in conjunction with the effect of time discounting. In Section VIII we put our theoretical results to work by considering a simple but realistic aggregation scenario. In Section IX, we present our simulation results. Finally, Section X offers concluding remarks and directions for future investigations.

II. Motivating Scenarios

Consider a sensor network deployed with the specific mission of protecting a power plant. The information about a possible attack on the power plant will be more valuable the less it is delayed. It will lose value continuously as there will not be adequate time to prepare. Thus, there are powerful incentives for reporting an attack as soon as possible. However, the cost of a false alarm is considered to be prohibitive in terms of the amount of human attention it requires. Thus, there are powerful incentives for aggregating individual actor information before reporting.

To begin, imagine that the power plant is threatened by an intruder who intends to sabotage the turbines. Again, the less delay in the information about the intrusion event, the better. The sensors that have detected the event need to decide whether to report an intrusion (and risk triggering a false alarm) or wait until several other sensors have corroborated the intrusion. With each moment of delay in notification of the intrusion, the ability to find the intruder decreases, as the intruder may be moving, and the area to search increased quadratically in the time since detection.

Next, imagine a foreign hacker who launches an attack on the network equipment controlling the power plant. The earlier the cyber-attack is detected, the higher the chance of thwarting the intruder. But as time goes on, the worse the attack gets. This type of cyber-attack may well double or triple the malicious network traffic with each time increment. Thus, the value of the information to the decision maker will deteriorate rapidly, as it becomes harder and harder to fight the attack as the network becomes overwhelmed.

The common characteristic of all the above scenarios is that getting information quickly has value. On the other hand, there are costs associated with obtaining information.

III. Aggregation dynamics

We assume a set of actors associated with particular pieces of information. This may be both sensor networks, where the actors are sensors, and social networks, where the actors are people. Computer networks can also be modeled in this way. Such situations are complex, with new and emerging external inputs, feedback loops, etc. Here we simplify by considering a situation in which the network actors have values, which are deteriorating with time, and must decide to aggregate or not with others. There are powerful incentives for aggregation. For example, individual actors may report an event based solely on the information they hold but they risk reporting a false positive. In many systems, the cost of a false alarm is considered to be prohibitive and must be avoided. On the other hand, aggregation involves costs, too. Chief among these costs is inter-actor communication which, in wireless environments, may be high. Also, aggregation takes time and during this time the value of the information continues to decay. As a result, occasionally, aggregation may not return as much value as it costs to perform the aggregation [10].

IV. Time-Discounted Information

Before we begin, we feel it is important to point out that in this paper we do not distinguish between *raw data* and *processed data*, using the term *information* for both. We trust that the context will disambiguate.

The phenomena we discuss occur in continuous time. Consider an arbitrary sensor and let X be the random variable that describes the "amplitude" of the sensed attribute. To specify that the sensor has collected the data at time r we shall write $X(r)$ and refer to it as the *value* of the data at time r. To avoid trivialities, in the sequel of this work we shall assume that $X(r) \neq 0$.

In general, the value of information may decay with time. In its most general form, for $t \geq r$, the *discounted* value, $X(t)$, of $X(r)$ at time t is given by

$$X(t) = X(r)g(r,t) \qquad (1)$$

where $g : \mathbb{R}^+ \cup \{0\} \times \mathbb{R}^+ \cup \{0\} \to [0,1]$ is referred to as a *discount function*.

In a variety of practical applications, the discount function in (1) is, actually, a function of the difference $t - r$ only, that is, a function of the difference between the time of data collection and the current time. With this in mind, in this paper we are interested in discount functions satisfying the condition

$$X(t) = X(r)\delta(t - r) \qquad (2)$$

with $\delta : \mathbb{R}^+ \cup \{0\} \longrightarrow [0, 1]$.

Equation (2) tells us that the *penalty* of waiting for $t - r$ time is that the value of the information collected by the sensor decreases from $X(r)$ to $X(t)$. Refer to Figure 1 for an illustration.
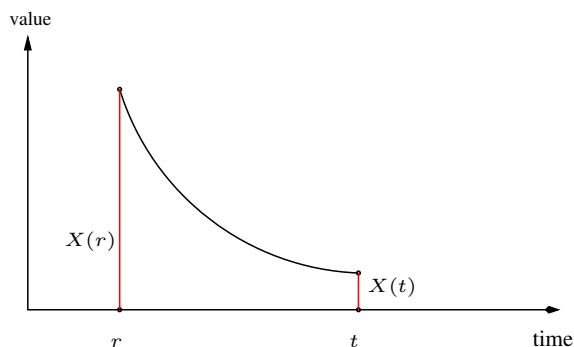

Figure 1: *Illustrating information decay.*

Obviously, $X(r) = X(r)\delta(0)$ implies

$$\delta(0) = 1. \qquad (3)$$

Further, we assume that after a very long time, the value of information vanishes. Formally, we assume that

$$\lim_{x \to \infty} \delta(x) = 0. \qquad (4)$$

We begin by proving the following useful result that will be instrumental in obtaining a closed form for δ.

Lemma IV.1. *If* $X(r) \neq 0$ *then for all* r, s, t *with* $0 \leq r \leq s \leq t$

$$\delta(t - r) = \delta(s - r)\delta(t - s). \qquad (5)$$

Proof. Applying (2) to the pairs (r, s), (s, t), (r, t) we obtain $X(s) = X(r)\delta(s - r)$, $X(t) = X(s)\delta(t - s)$ and $X(t) = X(r)\delta(t - r)$ which, combined, yield

$$X(r)\delta(t - r) = X(r)\delta(s - r)\delta(t - s).$$

Since $X(r) \neq 0$ the conclusion follows. □

Observe that by virtue of (4), δ cannot be identically 1 on $\mathbb{R}^+ \cup \{0\}$. Our next result shows that, in fact, δ takes on the value 1 if and only if $x = 0$.

Lemma IV.2. $\delta(x) = 1$ *if and only if* $x = 0$.

Proof. Recall that by (3), if $x_0 = 0$ then $\delta(x_0) = 1$. To prove the converse, let x_0 be the *largest* non-negative real for which $\delta(x_0) = 1$. It suffices to show that $x_0 = 0$. Suppose not and consider $\delta(2x_0)$. We can write

$$
\begin{aligned}
\delta(2x_0) &= \delta(2x_0 - 0) \\
&= \delta(2x_0 - x_0)\delta(x_0 - 0) \quad \text{[by (5)]} \\
&= \delta(x_0)\delta(x_0) \\
&= 1, \quad \text{[since } \delta(x_0) = 1]
\end{aligned}
$$

contradicting the maximality of x_0. Thus, $x_0 = 0$ and the proof of the lemma is complete. □

Corollary IV.3. *For all* $x > 0$, $0 < \delta(x) < 1$.

Proof. Follows immediately from (3) and Lemma IV.2, combined. □

For all r, s, t with $0 \leq r \leq s \leq t$ let x and y stand for $s - r$ and $t - s$, respectively. In this notation, $t - r = x + y$ and (5) can be written in the equivalent form

$$\delta(x + y) = \delta(x)\delta(y) \qquad (6)$$

with both x and y non-negative. As it turns out, the functional equation (6) has a simple solution that we discuss next.

Theorem IV.4. *If the function* $f : [0, \infty) \longrightarrow \mathbb{R}$ *satisfies the functional equation* $f(x + y) = f(x)f(y)$ *and is not identically zero then there exists a constant* a *such that*

$$f(x) = e^{ax} \qquad (7)$$

Our proof of Theorem IV.4 can be found in [9].

We are now in a position to show that the discount function δ is, in fact, an exponential. The details are spelled out by the following theorem.

Theorem IV.5. *For all* r *and* t *with* $0 \leq r \leq t$,

$$\delta(t - r) = e^{-\mu(t-r)}$$

where

$$\mu = -\ln \delta(1) > 0$$

Proof. Recall that by (6) the discount function δ satisfies the conditions of Theorem IV.4. Moreover, by Corollary IV.3 $0 < \delta(1) < 1$ and so $\ln \delta(1) < 0$. Thus, with

$$\mu = -\ln \delta(1) > 0$$

the expression of $\delta(t - r)$ becomes $\delta(t - r) = e^{-\mu(t-r)}$, as claimed. □

Theorem IV.5 shows that, under mild assumptions, the discount function is an exponential. We note that a similar result was derived by [6, 10] in the case of discrete time.

IV.A. A Related Discount Function

The goal of this subsection is to present a variant of the discount function δ defined in (2). Indeed, in some applications, the discounted value of the information at time $t > r$ may be naturally expressed as

$$X(t) = X(r) - X(r)\psi(t-r) \qquad (8)$$

where $\psi : \mathbb{R}^+ \cup \{0\} \to [0,1]$. As we are about to show, the discount functions δ and ψ are related in a simple fashion. To see this, consider three reals r, s, t with $0 \leq r < s < t$ and assume that $X(r) \neq 0$. It is easy to see that, by virtue of (8), we can write

$$
\begin{aligned}
X(t) &= X(r)[1 - \psi(t-r)] \\
&= X(r)[1 - \psi(t-s)][1 - \psi(s-r)]
\end{aligned}
$$

implying that

$$1 - \psi(t-r) = [1 - \psi(t-s)][1 - \psi(s-r)].$$

By writing $\xi(t-r) = 1 - \psi(t-r)$, $\xi(s-r) = 1 - \psi(s-r)$ and $\xi(t-s) = 1 - \psi(t-s)$ we obtain

$$\xi(t-r) = \xi(s-r)\xi(t-s)$$

which is, essentially, identical to (5). As an immediate consequence of Theorem IV.5

$$1 - \psi(t-r) = e^{-\lambda(t-r)}$$

where $\lambda = -\ln[1 - \psi(1)] > 0$.

Finally, simple algebra reveals that

$$X(t) = X(r)e^{-\lambda(t-r)}$$

which is of the same form as (2) and so the discount functions δ and ψ are, essentially, the same.

V. Instantaneous Aggregation

Information aggregation can be thought of as being orthogonal to the effect of time on its value. In this section we discuss the basics of aggregation and look at the aggregation of information independent of time discounting. The full effect of time discounting and its effect on aggregation will be discussed in detail in Section VII.

Assume a number of sensors have witnessed a short-lived event (e.g. an explosion) and have collected relevant information. In some applications it may well be feasible to proceed with a relatively simple aggregation. Specifically, let $X_1, X_2, \cdots, X_n, \cdots$ be the readings of a relevant attribute (e.g., seismic tremors caused by the explosion mentioned above) collected by the sensors. Due to spatial diversity of sensor locations, we assume that the X_is are independent random variables with a common underlying distribution function F. We further assume that the X_is have finite first and second moments.

In order to proceed to an instantaneous aggregation the data collected by the sensors, we define an application-dependent threshold Δ and seek to characterize those readings X_n that exceed Δ. What is the expected value and variance of such an X_n? To answer this question, let G be the distribution function of X_n.

For $x \geq \Delta$,

$$
\begin{aligned}
G(x) &= \Pr[X_n \leq x] \\
&= \Pr[X \leq x \mid X > \Delta] \\
&= \frac{\Pr[\{X \leq x\} \cap \{X > \Delta\}]}{\Pr[X > \Delta]} \\
&= \frac{\Pr[\Delta < X \leq x]}{1 - F(\Delta)} \\
&= \frac{F(x) - F(\Delta)}{1 - F(\Delta)}. \qquad (9)
\end{aligned}
$$

By (9), we write

$$
G(x) = \begin{cases}
0 & \text{for } x < \Delta \\
\frac{F(x)-F(\Delta)}{1-F(\Delta)} & \text{for } x \geq \Delta.
\end{cases} \qquad (10)
$$

Now that we have the distribution of X_n we turn our attention to computing both $E[X_n]$ and $Var[X_n]$. First, by definition

$$
\begin{aligned}
E[X_n] &= \int_0^\infty u \, dG(u) \\
&= \int_0^\Delta u \, dG(u) + \int_\Delta^\infty u \, dG(u) \ \text{[by (10)]} \\
&= \int_\Delta^\infty u \, dG(u) \\
&= \int_\Delta^\infty u \frac{dF(u)}{1 - F(\Delta)} \ \text{[by (10)]} \\
&= \frac{1}{1 - F(\Delta)} \int_\Delta^\infty u \, dF(u). \qquad (11)
\end{aligned}
$$

To evaluate $Var[X_n]$ we proceed as follows:

$$
\begin{aligned}
Var[X_n] &= E[X_n^2] - E[X_n]^2 \\
&= \int_\Delta^\infty u^2 \, \mathrm{d}G(u) - E[X_n]^2 \\
&= \frac{1}{1 - F(\Delta)} \int_\Delta^\infty u^2 \, \mathrm{d}F(u) \\
&\quad - \left(\frac{1}{1 - F(\Delta)} \int_\Delta^\infty u \, \mathrm{d}F(u) \right)^2 \quad (12)
\end{aligned}
$$

We note that if the threshold Δ as well as the distribution function F are known to the sensors, each of them can determine the conditional expectation and variance of those readings that exceed Δ. This trick allows the sensors to judge if the values they hold are "relevant". The sensors with relevant data will transmit them without delay to a local sink which will make the final determination.

VI.   The Algebra of Aggregation

Consider two sensors that have collected data about some attribute of an event they have witnessed and let X and Y be, respectively, the values collected. It is usually important for X and Y to be integrated in order to obtain a more reliable and, perhaps, more relevant information about the event at hand. We refer to the operation of integrating sensor data as *aggregation* and denote the result by $X \Diamond Y$. Here, $\Diamond$ is an application-dependent operator. Natural instances of $\Diamond$ include $+$, $\max$, $\min$, XOR, OR, among many similar ones. It is obvious that the aggregation operator can be extended to an arbitrary number of values that need to be aggregated.

To conform to our empirical experience, the aggregation operator $\Diamond$ is assumed to have the following fundamental properties:

Commutativity: $X \Diamond Y = Y \Diamond X$ for all X and Y. The result of the aggregation does not depend on the order in which the values are aggregated.

Associativity: $[X \Diamond Y] \Diamond Z = X \Diamond [Y \Diamond Z]$ for all X, Y, Z. If several values are aggregated in groups, the value of the aggregated information does not depend on the order in which groups are formed. It is customary to write $X \Diamond Y \Diamond Z$ instead of the cumbersome parenthesized expressions. More generally, we shall write $\Diamond_{i=1}^n X_i$ instead of $X_1 \Diamond X_2 \Diamond \cdots \Diamond X_n$.

Idempotency: If $Y = 0$ then $X \Diamond Y = X$. In other words, aggregation with information of value 0

has no effect. This is the view we take in this paper. We note however, that the idempotency property may be extended to read $X \Diamond Y = X$ whenever $Y \leq X$, mirroring our intuitive idea that one does not stand to gain by aggregating with information of lesser value.

As an example, it is easy to verify that the operator $\Diamond$ defined as

$$
X \Diamond Y = X + Y - XY \quad (13)
$$

satisfies the associativity, commutativity and idempotency properties defined above. A straightforward inductive argument shows that for an arbitrary collection $X_1, X_2, \cdots, X_n$, we have

$$
\begin{aligned}
\Diamond_{i=1}^n X_i &= X_1 \Diamond (X_2 \Diamond X_3 \Diamond \cdots \Diamond X_n) \\
&= (X_2 \Diamond X_3 \Diamond \cdots \Diamond X_n) \Diamond X_n
\end{aligned}
$$

and, moreover,

$$
\begin{aligned}
\Diamond_{i=1}^n X_i &= \sum_{i=1}^n X_i - \sum_{1 \leq i < j \leq n} X_i X_j \\
&\quad + \sum_{1 \leq i < j < k \leq n} X_i X_j X_k \\
&\quad + \cdots + (-1)^{n+1} X_1 X_2 \cdots X_n. \quad (14)
\end{aligned}
$$

For later reference, we now state and prove a technical result involving the operator $\Diamond$ defined in (13).

Lemma VI.1. *Let $X_1, X_2, \cdots, X_n$ be reals with $0 \leq X_i \leq 1$ for all $i = 1, 2, \cdots n$. Then*

$$
1 - \Diamond_{i=1}^n X_i = \Pi_{i=1}^n (1 - X_i).
$$

The proof of Lemma VI.1 can be found in the Appendix.

In order to illustrate the discussion above, consider an emergency event, say fire, witnessed by a number of sensors. Each of the sensors obtains a temperature reading and we wish to aggregate this information as soon as possible. For this purpose, consider a generic sensor, in close proximity to the fire, that has obtained a temperature reading t_i. Consider the event T_i that the sensor temperature reading is t_i. We define the "value" X_i of this temperature reading as follows:

$$
X_i = \Pr[T_i \mid F]. \quad (15)
$$

In other words, the value of the information collected by the sensor is the conditional probability of a sensor witnessing fire to obtain a temperature reading of t_i given that the event F (i.e., fire) has occurred.

Importantly, individual sensors, once their temperature reading is available, can evaluate the corresponding X_i by a simple table lookup.

The aggregated value of the various X_is is

$$
\begin{aligned}
\Diamond_{i=1}^n X_i &= \sum_{i=1}^n \Pr[T_i \mid F] - \sum_{1 \le i < j \le n} \Pr[T_i \cap T_j \mid F] \\
&+ \sum_{1 \le i < j < k \le n} \Pr[T_i \cap T_j \cap T_k \mid F] \\
&+ \cdots + (-1)^{n+1} \Pr[T_1 \cap T_2 \cap \cdots \cap T_n \mid F] \\
&= \Pr[T_1 \cup T_2 \cup \cdots \cup T_n \mid F]
\end{aligned}
$$

Assuming that the T_is are independent, Lemma VI.1 allows us to write

$$
\begin{aligned}
\Diamond_{i=1}^n X_i &= \sum_{i=1}^n \Pr[T_i|F] \\
&- \sum_{1 \le i < j \le n} \Pr[T_i|F]\Pr[T_j|F] \\
&+ \sum_{1 \le i < j < k \le n} \Pr[T_i|F]\Pr[T_j|F]\Pr[T_k|F] \\
&+ \cdots \\
&+ (-1)^{n+1} \Pr[T_1|F]\Pr[T_2|F]\cdots\Pr[T_n|F] \\
&= 1 - \Pi_{i=1}^n (1 - \Pr[T_i|F]) . \quad (16)
\end{aligned}
$$

As a final step, a suitable thresholding procedure determines whether or not the event is worth reporting. The thresholding strategy is suggested by Lemma VI.1. The intention is to prevent potentially numerous relatively small temperature readings from aggregating into a value that exceeds the threshold. To implement this idea, it suffices to set the threshold

$$
\Delta = 1 - 10^{-k} \quad (17)
$$

where the choice of k will be explained shortly.

For example, we may well decide to take the conditional probability of a reading in the *critical range*, say 100°C to 500°C, given that fire has occurred as 0.9 and 0 for readings outside of this range. By Lemma VI.1 for a given value of k, the threshold Δ in (17) is reached if and only if at least k sensors have obtained readings in the critical range.

VII. Aggregating time-discounted information

To counter-balance the effect of time discounting, we define an algebraic operation on sensor data that we call *aggregation*. Consider two sensors that have collected data about an event at times r and s. Let $X(r)$

and $Y(s)$ be, respectively, the values of the information collected by the two sensors. At some later time t, the two sensors decide to integrate their information.

Observe that since $t \ge r$ and $t \ge s$, what is being aggregated at time t are the discounted values $X(r)\delta(t-r)$ and $Y(s)\delta(t-s)$ as illustrated in Figure 2. Thus, we write

$$
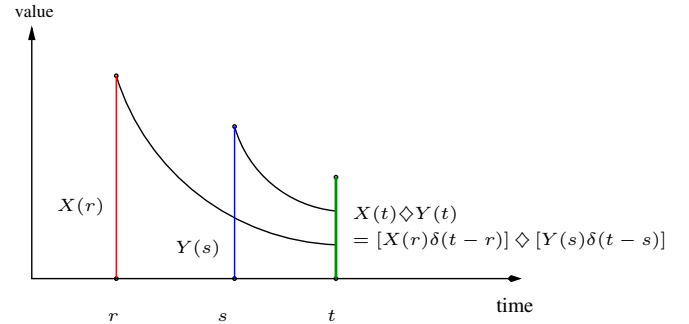X(t)\Diamond Y(t) = [X(r)\delta(t-r)] \Diamond [Y(s)\delta(t-s)] . \quad (18)
$$

Figure 2: *Illustrating the aggregation at time t of $X(r)$ and $Y(s)$.*

In order to be able to understand how time discounting affects aggregated values we shall find it convenient to assume that the discount function δ distributes over $\Diamond$.

Distributivity: For all t, τ with $0 \le t \le \tau$, we can write $[X(t)\Diamond Y(t)]\delta(\tau - t) = [X(t)\delta(\tau - t)] \Diamond [Y(t)\delta(\tau - t)]$. The discounted value at time $\tau - t$ of the information $X(t)\Diamond Y(t)$ aggregated at time t matches the aggregated value at time τ of $X(t)\delta(\tau - t)$ and $Y(t)\delta(\tau - t)$. In other words, it does not matter whether we first aggregate and then discount the aggregated information or vice versa.

The distributivity property is fundamental in understanding the interplay between time discounting and aggregation. We mention in passing that, in general the distributivity property need not be verified. However, in this paper we look specifically at aggregation operators where distributivity holds.

Lemma VII.1. *Assuming the distributivity property, for all $0 \le r \le s \le t \le \tau$, we have*

$$
[X(t)\Diamond Y(t)]\delta(\tau-t) = [X(r)\delta(\tau - r)] \Diamond [Y(s)\delta(\tau - s)] . \quad (19)
$$

Proof. By using distributivity we write

$$[X(t) \lozenge Y(t)] \, \delta(\tau - t)$$
$$= [X(t)\delta(t - r) \lozenge Y(t)\delta(t - s)] \, \delta(\tau - t) \text{ [by (2)]}$$
$$= [X(t)\delta(t - r)\delta(\tau - t)] \lozenge [Y(t)\delta(t - s)\delta(\tau - t)]$$
$$= X(r)\delta(\tau - r) \lozenge Y(t)\delta(\tau - s) \text{ [by Lemma IV.1]}$$

$\square$

The left-hand side of (19) is the discounted value of $X(t)\lozenge Y(t)$ at time τ, while the right-hand is the aggregated value of the discounted values of $X(r)$ and $Y(s)$ at time τ.

VII.A. Discounted Value of Aggregated Information

Consider an event witnessed by n, $(n \geq 2)$, sensors and let the sensed values be $X_1, X_2, \cdots, X_n$, collected, respectively, at times $t_1, t_2, \cdots, t_n$. Assume, further, that various groups of sensors have aggregated their information before time t and that, finally, at time t the aggregation has been completed. We are interested in evaluating the time-discounted value of the information collected by the sensors at time t, where $t \geq \max\{t_1, t_2, \cdots, t_n\}$. The answer to this natural question is provided by the following fundamental result.

Theorem VII.2. *Assuming distributivity, the discounted value $V(t)$ of the aggregated information at time t is*

$$V(t) = \lozenge_{i=1}^{n} X_i(t_i)\delta(t - t_i), \qquad (20)$$

regardless of the order in which the values were aggregated.

Proof. The proof is by induction on n. For $n = 2$, the conclusion follows at once from Lemma VII.1. Now, let $n \geq 2$, be arbitrary and assume the statement true for all m, $(m < n)$. We assume, without loss of generality, that the last aggregation takes place at time t. This aggregation must have involved a number of disjoint groups $G_1, G_2, \ldots, G_p$ each of them the result of a previous aggregation at times, respectively, $u_1, u_2, \ldots, u_p$. Observe that we can always relabel the groups in such a way that their aggregation times are ordered as $u_1 < u_2 < \cdots < u_p$.

Let us look at group G_k. By the induction hypothesis, the value of information in group G_k aggregated at time u_k was

$$V(u_k) = \lozenge_{j=1}^{n_k} X_{k_j}\delta(u_k - t_{k_j})$$

where, of course, we assume that group G_k involves n_k sensors whose values were aggregated.

Assuming $t \geq u_k$, the discounted value of $V(u_k)$ at time t is

$$V_k(t) = \left[\lozenge_{j=1}^{n_k} X_{k_j}\delta(u_k - t_{k_j})\right]\delta(t - u_k)$$
$$= \lozenge_{j=1}^{n_k} X_{k_j}\delta(u_k - t_{k_j})\delta(t - u_k)$$
$$= \lozenge_{j=1}^{n_k} X_{k_j}\delta(t - t_{k_j}) \text{ [by Lemma IV.1]}.$$

which is exactly the discounted value of information collected by sensors in group G_k, had it been aggregated at time t. Since G_k was arbitrary, the conclusion follows. $\square$

Theorem VII.2, in effect, says that the order in which the values are aggregated does not matter as long as each is aggregated only once. In practical terms, Theorem VII.2 gives the algorithm designer the freedom to schedule aggregation in a random manner, much in line with the stochastic nature of wireless communication and sensor data aggregation.

VII.B. Thresholding

Recall that, as already mentioned, we assume that reporting a false positive involves a huge overhead and is considered prohibitively expensive. Mindful of this state of affairs, having aggregated, at time t, the information collected by the various sensors, it is important to decide whether this information warrants reporting.

One of the natural strategies employed is thresholding. Specifically, a policy is followed of first setting up an application-dependent threshold Δ and then reporting an event only if the aggregated information exceeds Δ.

As Figure 3 illustrates, the time at which the aggregation is performed is critical. Indeed, in the figure the aggregated value $X_1(t)\lozenge X_2(t)$ barely exceeds the threshold Δ. Aggregation at a slightly later time would not exceed the threshold and a relevant event would go unreported.

VIII. Putting all this to work

The main goal of this section is to show how the theoretical concepts developed in the previous sections apply to a practically relevant scenario.

Consider a fire event witnessed by a number of sensors deployed in a given area. For simplicity, assume that each sensor has collected a temperature value. Let $X_1, X_2, \cdots$ be, respectively, the sensed temperature values collected by the various sensors at times $t_1, t_2, \cdots$. Since the sensors have witnessed the same
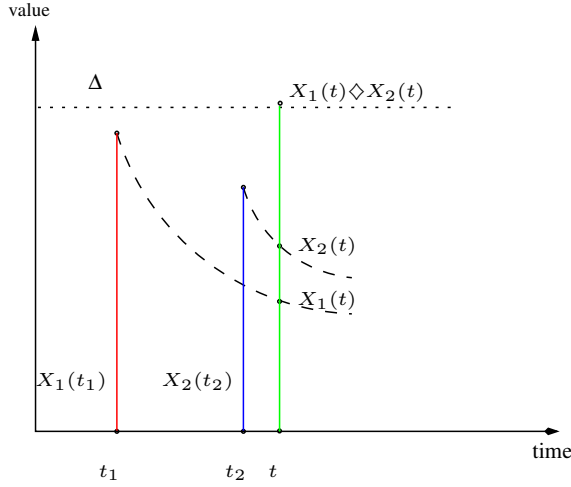
Figure 3: *Illustrating thresholding.*

event, it is natural to assume that the random variables $X_1, X_2, \cdots$ come from the same underlying distribution X with finite expectation $E[X] < \infty$.

We assume that the X_is are independent and, moreover, they are independent of the times $t_1, t_2, \cdots$ at which the data was collected. These assumptions can be justified by the spatial diversity of the sensors.

For a generic sensor that has collected data at time t_i, we let $X_i(t_i)$ denote the value of this information when it was collected. By Theorem (IV.5) the discounted value of this information at a later time t is

$$X_i(t) = X_i(t_i)e^{-\mu(t-t_i)}. \qquad (21)$$

Further, as a QoS parameter intended to avoid reporting a false positive, we need a minimum of k individual temperatures to be aggregated. Given an expected temperature reported of 300°C, this requirement is tantamount to insisting on accumulating a total of $\Delta = k \times 300$ "temperature points" as a result of aggregation. In turn, this suggests $\diamond = "+"$ as a suitable aggregation operator. It is easy to confirm that for the chosen $\diamond$ the distributivity property holds and so the results of Section VII apply.

In this context, we are interested in evaluating the *expected* time-discounted value, $V(t)$, at time t, of the information collected by the sensors where $t \geq \max\{t_1, t_2, \cdots\}$. To answer this natural question we make the simplifying assumption that $t_1, t_2, \cdots$ are the times of a Poisson process with parameter $\lambda > 0$. In other words, λ is the rate at which the sensors that witnessed an event are ready to report their sensory data.

Theorem VIII.1. *The expected time-discounted value, $E[V(t)]$, of the information collected by sensors*

at times $t_1, t_2, \cdots$ is

$$E[V(t)] = \frac{\lambda}{\mu}E[X]\left[1 - e^{-\mu t}\right] \qquad (22)$$

where $\lambda > 0$ is the rate at which the sensors collect their data and $E[X]$ is the common expectation of $X_1, X_2, \ldots$.

Proof. Recall that we assumed that the sensors collected their data at the times of a Poisson process with parameter λ. By the Law of Total Expectation,

$$E[V(t)] = \sum_{n \geq 1} E[V(t)|\{N = n\}]P[\{N = n\}] \qquad (23)$$

where N is the random variable that counts the number of sensors that have data ready for aggregation by time t. By (21), Theorem IV.4 the conditional expectation, $E[V(t)|N = n]$, can be written as

$$
\begin{aligned}
E[V(t)|\{N = n\}] &= E[\sum_{i=1}^{N} X_i(t_i)e^{-\mu(t-t_i)}|\{N = n\}] \\
&= \sum_{i=1}^{n} E[X_i(t_i)e^{-\mu(t-t_i)}]
\end{aligned}
$$

It is well known that, given that n Poisson events were recorded in $(0, t]$, their conditional distribution is uniform. Thus,

$$
\begin{aligned}
&E[V(t)|\{N = n\}] \\
&= \sum_{i=1}^{n} E[X_i(t_i)e^{-\mu(t-t_i)}] \\
&= \sum_{i=1}^{n} E[X_i(t_i)e^{-\mu(t-U_i)}] \\
&\quad \text{[where the } U_i\text{s are uniform in } (0, t]] \\
&= \sum_{i=1}^{n} E[X_i(t_i)]E[e^{-\mu(t-U_i)}] \\
&\quad \text{[because the } X_i\text{s and } U_i\text{s are independent]} \\
&= \sum_{i=1}^{n} E[X]E[e^{-\mu(t-U_i)}] \\
&\quad \text{[recall, } X \text{ is the common distribution of the } X_i\text{s]} \\
&= E[X]e^{-\mu t}\sum_{i=1}^{n} E[e^{\mu(U_i)}] \\
&= e^{-\mu t}E[X]\sum_{i=1}^{n} \int_{0}^{t} e^{\mu u}\frac{du}{t} \\
&= e^{-\mu t}E[X]\sum_{i=1}^{n} \frac{e^{\mu t} - 1}{\mu t} = \frac{E[X]}{\mu t}\sum_{i=1}^{n}[1 - e^{-\mu t}] \\
&= \frac{nE[X]}{\mu t}[1 - e^{-\mu t}]. \qquad (24)
\end{aligned}
$$

On plugging (24) back into (23), we obtain

$$
\begin{aligned}
E[V(t)] &= \sum_{n \geq 1} \frac{nE[X]}{\mu t}[1 - e^{-\mu t}]P[\{N = n\}] \\
&= \sum_{n \geq 1} \frac{nE[X]}{\mu t}[1 - e^{-\mu t}]\frac{(\lambda t)^n}{n!}e^{-\lambda t} \\
&= \frac{e^{-\lambda t}E[X][1 - e^{-\mu t}]}{\mu t}\sum_{n \geq 1}\frac{(\lambda t)^{(n-1)}}{(n-1)!} \\
&= \frac{e^{-\lambda t}E[X][1 - e^{-\mu t}]\lambda t}{\mu t}e^{\lambda t} \\
&= \frac{\lambda}{\mu}E[X][1 - e^{-\mu t}].
\end{aligned}
$$

□

There are a number of interesting things to note here:

- The actual distribution of the X_is does not appear explicitly in Theorem VIII.1. This is telling us that two quite different distributions with the same expectation are equivalent as far as Theorem VIII.1 is concerned;

- $E[V(t)] = \frac{\lambda}{\mu}E[X][1 - e^{-\mu t}]$ is an increasing function of time and

$$
\lim_{t \to \infty} E[V(t)] = \frac{\lambda}{\mu}E[X].
$$

Thus, for every application-dependent threshold Δ, there exists an earliest time when Δ is exceeded.

Note that, as mentioned above, Theorem VIII.1 allows us to evaluate the earliest time t at which the expected discounted value of the information collected by the sensors exceed an application-dependent threshold Δ. Thus, at time t, $E[V(t)] \geq \Delta$, or equivalently,

$$
\frac{\lambda}{\mu}E[X][1 - e^{-\mu t}] \geq \Delta.
$$

Solving for t, we obtain

$$
t \geq \frac{1}{\mu}\ln\frac{\lambda E[X]}{\lambda E[X] - \Delta\mu} \qquad (25)
$$

In fact, (25) states that a value of t exists only if $\lambda E[X] > \Delta\mu$ or, equivalently,

$$
\Delta < \frac{\lambda}{\mu}E[X]. \qquad (26)
$$

Observe that in practice it is the case that $E[X] < \Delta$, for otherwise there are no incentives for aggregation.

IX. Simulation Results

In this section, we evaluate our theoretical model through simulation.

IX.A. Discounted Value of Information

As a simple example, assume that the sensors are monitoring temperature to detect fire. The sensors report their sensed temperature over time, modeled as a uniform random variable in the range $[30, 450]°C$. For simplicity, temperature values are used as the value of information, and as discussed in Section VIII, we use $\diamondsuit = $ "$+$" as the aggregation function. In our scenario, an alarm is generated if the value of aggregated information for temperatures above $100°C$ exceeds 400. We assume that the value of information degrades exponentially with parameter $\mu = 5 \times 10^{-3}$.

Figure 4 shows the value of the discounted information over time (as described in Section IV). Each jump indicates an aggregation operation, increasing the value of the information. Then, the value decreases as time goes by until the next aggregation operation. This behavior repeats until the total value of the aggregated information exceeds 400. The increase in the value of information could also cease if there are no sensors left to report new information. After this point in time, the value of aggregated information decreases over time.
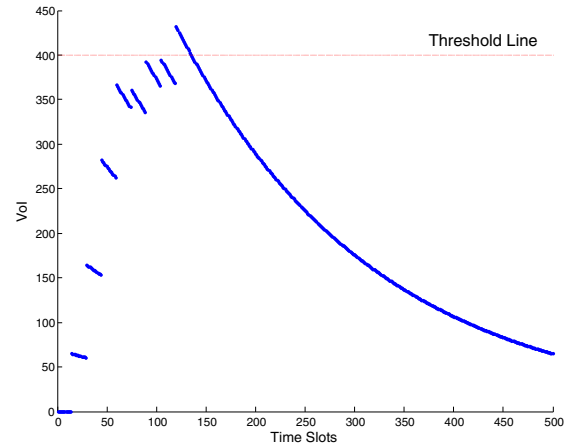

Figure 4: *Value of aggregated information over time*

IX.B. Effect of Aggregation Method

In this section, we simulate a scenario in which the sensors collect information and report it to a central node. We want to evaluate the effect of wireless communication and aggregation methods on the *average time to aggregate*. Collisions in the wireless channel

typically garble messages beyond recognition. Thus messages need to be retransmitted which will increase the time to receive all pieces of information to be aggregated. Clearly, the denser the traffic the more collisions and the more retransmissions causing delays in aggregation. A non-negligeable side-effect of all this is that due to collision-caused time delays, the value of individual pieces of information decays and what is being aggregated has lesser value.

Furthermore, the aggregation method could affect the time to achieve the defined aggregation. This latency is due, to a large extent, to the logic behind the transfer of information for aggregation. For example, the aggregation method i may use one of the following strategies:

(I) : wait for reports from at least k nodes;

(II) : wait until the aggregated value exceeds a threshold;

(III) : wait until there is enough spatial diversity in the reported information.

In all these examples, the sensors can decide about reporting in a way that only the effective ones report. For example, in the former method, it would be enough if only one node out of several neighbor nodes transmits the message. This can be performed through many methods such as cluster head or a probabilistic transmission policy. Independent of the policy, transmitting fewer messages would reduce the traffic and result in faster reception of the required information. In this section, we develop a general simulation scenario that will show the effect of these two factors, i.e. wireless channel condition and aggregation method, on the increase in the time to aggregate, which consequently avoids reducing the value of information. This simulation scenario is designed in a way that is independent of the aggregation method.

In our simulation, the sensors are deployed uniformly at random over an area of $750m \times 750m$. The sensors can communicate with the aggregator node in a single hop. At least two nodes are required to report to aggregator. The number of sensors vary from 4 to 100. A sensor will periodically send its sensed information with rate λ. We simulate this process in NS 2.29 [1]. The simulation parameters are shown in Table 1.

To show the effect of the aggregation method, we evaluate scenarios where only some portion (p) of the nodes send their sensed information. The aggregator waits until it receives the information from p portion of nodes. This is a general model for many aggregation strategies in sensor networks. For example,

Table 1: Simulation Parameters

Nodes	4-100
Simulation Time	2500 s
Data Rate	250 kbps
MAC	CSMA/CA
Communication Range	700 m
Packet Size	1000 B
Reporting rate (λ)	0.2 (1/s)

if sensors are programmed to report their temperature only if they sense temperature between $100°C$ to $500°C$, then in a fire detection scenario, only the sensors that are close to the fire and can detect it will send temperature readings. This strategy will avoid collisions among all nodes as well as save energy in comparison to the scenario where all nodes try to transmit information. Therefore, we present the results for several values of p. Note that $p = 1$ is the worst case scenario, where all nodes report their sensed information.

Figure 5 shows the average time to aggregate when various portions of nodes are involved in transmitting information for aggregation. With fewer nodes, of course, aggregation happens faster. For example, with $p = 0.25$ and 10 nodes, the average time to aggregate is 1.7s, but when the number of nodes is increased to 20, the average time to aggregate increases to 4.7s. Figure 5 shows the effect of the aggregation method (here, p) on the average time to aggregate. For example, with 4 nodes, the average time to aggregate changes from 1.7 seconds to 4.7 seconds when p is increased from 0.5 to 1. Recall that because of value decay, the longer the average time to aggregate, the lower the value of the aggregated information.
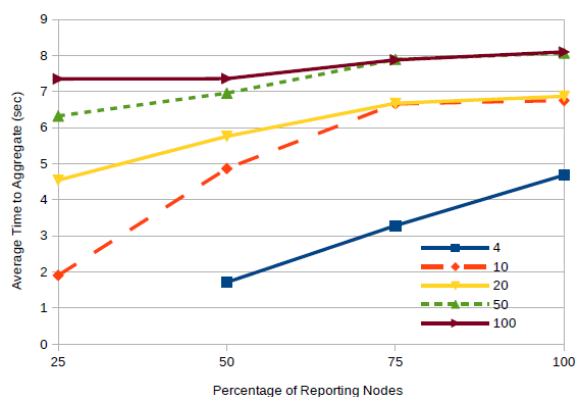

Figure 5: *Average time to aggregate for various number of nodes. At least two nodes are required for aggregation.*

The delay due to waiting for arriving pieces of information as well as collision in the reception of those

pieces of information will result in reducing the value of information. Figure 6 illustrates the effect of those delays on the loss of information value for several exponential decay functions.
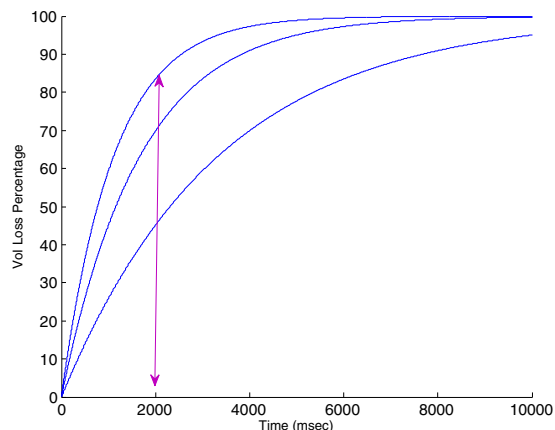

Figure 6: *Percentage of value of information loss over time for several exponential decay functions.*

As can be seen even with 2 seconds of delay, the loss of value can be up to 80%. This delay may be due to waiting for pieces of information to be generated by different sources and arriving at the aggregator, as well as delay due to the wireless channel conditions. In the next simulation, we investigate how the rate of information generation in relation to decay should be designed to achieve the value of information as soon as possible, so that a decision maker can make a proper determination.

IX.C.   Time to Exceed a Threshold

Here, we evaluate the applicability of the defined threshold time in (25). Recall that (25) shows the minimum time at which the expected value of information will exceed the threshold. This can be useful, for example, to plan for capturing an intruder. If we assume that the intruder's staying time in the monitored area is an exponential random variable S with expected value $E[S] < \infty$ and the reaction time of the security personnel to arrive at the location after receiving the alarm is a random variable R with expected value $E[R]$, then the expected detection time D can be written as $D \leq E[S] - E[R]$. In other words, the design of λ, i.e., sensor reporting periods, should not let t in (25) exceed D. In the following, we simulate this scenario to show how λ and μ can affect the effectiveness of the monitoring application. Assume that $E[S] = 300s$ and $E[R] = 100s$. This means that the detection time should be less than 200 seconds. In this

simulation, we assume that V is selected uniformly in the range $[V_{min}, V_{max}]$, which are set to 0 and 100, respectively. This can be a valid setting since it could mean a direct functional mapping from the percentage of confidence that a sensor has in detecting the intruder. Then, we could set the threshold Δ to 196, which means that we need to have a confidence equal to two reports with the 98% confidence in detection of intrusion before reporting. Letting $VoI(t)$ stand for the random variable that keeps track of the value of information at time t, Table 2 shows the moments in time when the expected value, $E[VoI(t)]$, of $VoI(t)$ will be above Δ for theory and simulation. We show the average time over 10,000 trials for various values of λ and μ.

Table 2: Time when $E[VoI(t)]$ exceeds a threshold

Config	λ	μ- decay par.	Theory	Sim.
1	0.01	2.50×10^{-3}	1475 s	1700 s
2	0.01	1.25×10^{-3}	534 s	557 s
3	0.02	2.50×10^{-3}	267 s	280 s
4	0.02	1.25×10^{-3}	223 s	247 s
5	0.10	2.50×10^{-3}	41 s	47 s
6	0.10	1.25×10^{-3}	40 s	46 s

As it turns out, the simulation results match the theoretical predictions well. In our defined scenario, we required that $E[VoI(t)]$ exceed Δ within 200 seconds. Table 2 shows that only configurations 5 and 6 meet that requirement, due to the arrival rate of 0.1, which has sensors reporting every 10 seconds.

X.   Concluding Remarks and Directions for Future Work

This paper provided a formal way of looking at aggregation of information in sensor networks. Of particular interest is the aggregation process in networks where individual actors possess information whose value decays over time. We offered a formal model for the valuation of time-discounted information and of the algebra of its aggregation. We allow aggregation of time-discounted information to proceed in an arbitrary, not necessarily pairwise, manner. We have shown that the resulting value of the aggregate does not depend on the order in which aggregation of individual values take place. Our results suggest natural thresholding strategies for the aggregation of the information collected by sets of network actors. Our theoretical predictions were confirmed by extensive simulation.

In spite of these results, a number of problems are open and are getting attention.

It would be of great theoretical interest to consider other time discounting regimens that do not conform (and are not amenable) to (2). Examples of such discount regimens are known to exist; step functions being a prime example. The question that we are addressing at the moment is that of approximating a step function and, indeed, other similar decay functions by polynomials. It is a classic result of Karl Weierstrass that every real function can be approximated by a suitable sequence of polynomials [5]. Since we have a good understanding a exponential decay functions and since every exponential can be approximated by a polynomial consisting of the first few terms in its Taylor expansion, this approach seems to be natural.

While a lot of attention had been devoted to detecting and mitigating the effects of collisions on the wireless channel, to the best of our knowledge, the effects of collisions on the value of information to be aggregated has not been addressed in the literature. As our simulations showed, due to the significant retransmission delays, the value of the information may be dramatically impacted.

Finally, it is of great practical interest to be able to retask the sensors as the mission dynamics evolve. Retasking may involve moving from one set of sensed attributes to another and also inferring an attribute for which the sensor does not have a direct sensing capability. This promises to be an exciting area of work.

References

[1] ns2: The network simulator ns-2, [online] available: http://www.isi.edu/nsnam/ns/.

[2] N. Ahituv. A systematic approach toward assessing the value of an information system. *MISQ*, 4(4):61–75, 1980.

[3] K. J. Arrow. The value of and the demand for information. In C. B. McGuire and R. Radner, editors, *Decision and organization. A volume in honor of Jacob Marschak*, pages 131–140. North-Holland Pub. Co., Amsterdam, 1972.

[4] S. Frederick, G. Loewenstein, and T. O'Donoghue. Time discounting and time preference: A critical review. *Journal of Economic Literature*, XL:351–401, 2002.

[5] J. E. Marsden. *Elementary Classical Analysis*. W. H. Freeman and Company, 1974.

[6] J. V. Nickerson and S. Olariu. A measure for integration and its application to sensor networks. In *Workshop on Information Technology and Systems (WITS)*, 2005.

[7] J. V. Nickerson and S. Olariu. Protecting with sensor networks: Attention and response. In *Proceedings of the 40th Annual Hawai'i International Conference on System Sciences*, 2007.

[8] S. Olariu, M. Eltoweissy, and M. Younis. ANSWER: AutoNomouS netWorked sEnsoR system. *Journal of Parallel and Distributed Computing*, 67(1):111–124, 2007.

[9] S. Olariu, S. Mokhrekesh, and M. Weigle. Toward aggregating time discounted information. In *ACM MiSeNet'2013*, September 2013.

[10] S. Olariu and J. V. Nickerson. A probabilistic model of integration. *Decision Support Systems*, 45(4):746 – 763, 2008.

[11] D. R. Raban and S. Rafaeli. The effect of source nature and status on the subjective value of information. *JASIST*, 57(3):321–329, 2006.

Appendix

A. Proof of Lemma VI.1

The proof is by induction. To settle the basis, consider $n = 2$ and assume that both X_1 and X_2 are reals in the interval $[0, 1]$. Observe that $1 - (X_1 \Diamond X_2) = 1 - (X_1 + X_2 - X_1 X_2) = (1 - X_1)(1 - X_2)$.

For the inductive step, let n be arbitrary and assume that the statement of the lemma true for the chosen value of n. With this in hand, we need to show that $1 - \Diamond_{i=1}^{n+1} X_i = \Pi_{i=1}^{n+1}(1 - X_i)$. Write $Y = 1 - \Pi_{i=1}^{n}(1 - X_i)$. In this notation

$$
\begin{aligned}
1 - \Diamond_{i=1}^{n+1} X_i &= 1 - (\Diamond_{i=1}^{n} X_i) \Diamond X_{n+1} \\
&= 1 - Y \Diamond X_{n+1} \\
&= 1 - Y - X_{n+1} + Y X_{n+1} \\
&= \Pi_{i=1}^{n}(1 - X_i) - X_{n+1} \Pi_{i=1}^{n}(1 - X_i) \\
&= \Pi_{i=1}^{n+1}(1 - X_i)
\end{aligned}
$$

and the proof of the lemma is complete.