

Toward Aggregating Time-Discounted Information*

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

ABSTRACT

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 is discounted 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 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. A sensor network with the mission of intrusion detection is used throughout as an illustrative example. The accuracy of our theoretical predictions was confirmed by simulating a number of realistic scenarios.

Categories and Subject Descriptors

C.2.1 [Network Architecture and Design]: Wireless communication; C.2 [COMPUTER-COMMUNICATION NETWORKS]: General; H.1.1 [Systems and Information Theory]: Value of information

Keywords

Time-discounted information, information aggregation, mission-oriented systems, sensor networks

*Work supported by NSF grant CNS-1116238

1. 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 [10]. The economic argument is that information is a good that can be traded, and hence has value [7]. Moreover, this value is 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 [5].

It has long been recognized that assessing the value of time-discounted information is hard to evaluate being too dependent on subjective human valuation [4]. It is, thus, important to consider formal ways of defining the aggregation of time-discounted information so that aggregation might be better described and understood. 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 a theoretical one: by proposing a model where there was none before, we 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 theoretical 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 time is of the essence and, consequently, the value of the information often changes so rapidly as to defy deterministic modeling [6].

As it turns out, in mission-oriented sensor networks, the primary agents are computational: the individual sensors collect, share and aggregate information. In every mission-oriented sensor network there is, ultimately, a human component, but before reaching the human the sensor network already aggregates and values information. This process consumes power that needs to be minimized [11]. Even in situations where power is not a major issue, human attention often is. In other words, it may be that not everything can be monitored all the time, and therefore human participation in the network needs to be minimized. Thus, the aggregation problem for sensor networks is often a multi-attribute one: try to detect events with minimal delay and aggregate detection, while at the same time try to minimize the use of energy [7].

The remainder of the paper is organized as follows: Section 2 presents our motivating scenarios. We discuss our assumption about aggregation dynamics in Section 3. Further, Section 4 provides a formalism for time-discounted information. Section 5 investigates the algebra of aggregation. In Section 6 we put our theory to work by considering a simple but realistic aggregation scenario. In Section 7, we present our simulation results. Finally, Section 8 offers concluding remarks and directions for future investigations.

2. 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 intrusion event 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 also incentives for aggregating individual sensor readings 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. For example, it is customary to assume that the sensors have at their disposal a modest, non-renewable energy budget. The sensors are usually programmed to preserve energy while, at the same time, they try to detect intrusion data early and aggregate the data collected [7].

3. ASSUMPTIONS ABOUT 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 also can 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 nodes have values, which are deteriorating with time, and must decide to aggregate or not with others. There is a cost to aggregation. In the case of sensor networks, the cost has to do with the power needed to communicate and process information and also with the risk of false positives. Indeed, the cost of a false alarm is considered to be prohibitive. We assume that the sensors run on small, non-rechargeable batteries, and so their energy budget cannot be replenished. In other networks, the resource to be consumed may be human attention; the idea, though, is the same. It is worth noting that occasionally aggregation may not return as much value as it costs to perform the aggregation [9].

4. TIME-DISCOUNTED INFORMATION

Before we begin, we feel it is important to point out that in this paper we do not distinguish between *data* (a.k.a *raw information*) and *processed data*, using the term information for both. We trust, the context will disambiguate. In particular, the sensory information collected by individual sensors should be interpreted as data, while the result of aggregating data collected by several sensors should be construed as bona-fide processed data, i.e., information.

The phenomena we discuss occur in continuous time. Having witnessed an event and having collected relevant information, the sensors have the choice to report a decision or else to defer the decision to a later time. Consider a sensor that, having witnessed an event, has collected relevant information. 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 $X(r) \neq 0$.

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

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

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

In many 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 such that

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

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

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 $X(t)$.

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

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

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

$$\lim_{x \rightarrow \infty} \delta(x) = 0. \quad (4)$$

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

Lemma 4.1. *If $X(r) \neq 0$ then for all r, s, t with $0 \leq r \leq s \leq t$*

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

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. \square

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 4.2. *$\delta(x) = 1$ 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. \square

Corollary 4.3. *For all $x > 0$, $0 < \delta(x) < 1$.*

Proof. Follows immediately from (3) and Lemma 4.2, combined. \square

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

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

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

Theorem 4.4. *If the function $f : [0, \infty) \rightarrow \mathbb{R}$ satisfies the functional equation $f(x+y) = f(x)f(y)$ and is not identically zero then there exists a constant a such that*

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

Our proof of Theorem 4.4 relies on several lemmas.

Lemma 4.5. *If $f(x) \neq 0$, then $f(1) \neq 0$.*

Proof. We prove the contrapositive. Assume that $f(1) = 0$. First, we claim that $f(y) = 0$ for all $y \geq 1$. This follows since for $x = y-1 > 0$, $f(y) = f(1+x) = f(1)f(x) = 0$. Next, observe that $f(1) = f(\frac{1}{2} + \frac{1}{2}) = f(\frac{1}{2})^2 = 0$ and, thus, $f(\frac{1}{2}) = 0$. This, in turn, implies that $f(x) = 0$ for all $\frac{1}{2} \leq x \leq 1$. A similar argument shows that $f(\frac{1}{4}) = 0$ and that $f(x) = 0$ for all $\frac{1}{4} \leq x \leq \frac{1}{2}$. Now a simple inductive argument shows that for all positive integers $n > 0$, $f(\frac{1}{2^n}) = 0$ and that $f(x) = 0$ for all x in the range $\frac{1}{2^n} \leq x \leq \frac{1}{2^{n-1}}$. The conclusion follows. \square

By Lemma 4.5, we assume $f(1) \neq 0$; in fact, $f(1) = f(\frac{1}{2})^2$ guarantees that $f(1) > 0$. For all $x \geq 0$ define the function $\phi(x) = \frac{f(x)}{f(1)^x}$. Our goal is to show that $\phi(x) \equiv 1$ on $[0, \infty)$.

Lemma 4.6. *For all $x, y \geq 0$, $\phi(x+y) = \phi(x)\phi(y)$.*

Proof. By definition, we have $\phi(x+y) = \frac{f(x+y)}{f(1)^{x+y}} = \frac{f(x)f(y)}{f(1)^{x+y}} = \frac{f(x)}{f(1)^x} \frac{f(y)}{f(1)^y} = \phi(x)\phi(y)$ as claimed. \square

Lemma 4.7. *For all positive integers n , $\phi(n) = 1$.*

Proof. By induction: first, $\phi(1) = \frac{f(1)}{f(1)} = 1$. Assuming $\phi(n-1) = 1$, it follows that $\phi(n) = \phi(n-1+1) = \phi(n-1)\phi(1) = 1$. \square

Corollary 4.8. *$\phi(0) = 1$.*

Proof. $1 = \phi(1) = \phi(1+0) = \phi(1)\phi(0) = 1 \cdot \phi(0) = \phi(0)$. \square

Lemma 4.9. *For every $z > 1$ there exists $z_0 \in [0, 1]$ such that $\phi(z_0) = \phi(z)$.*

Proof. For every real z , there exists a unique positive integer n such that $n \leq z < n+1$. Clearly, $z_0 = z-n \in [0, 1]$. Now, $\phi(z_0) = \phi(z_0)\phi(n) = \phi(z_0+n) = \phi(z)$. \square

Corollary 4.10. *If $\phi(x) \equiv 1$ on $[0, 1]$ then $\phi(x) \equiv 1$ on $[0, \infty)$*

Lemma 4.11. *$\phi \equiv 1$ on $[0, 1]$.*

Proof. Otherwise, the set $A = \{y \in [0, 1] \mid \phi(y) \neq 1\}$ is non-empty. Let $x = \inf A$. By Corollary 4.8, $x > 0$. However, now $\phi(x) = \phi(\frac{x}{2} + \frac{x}{2}) = \phi(\frac{x}{2})^2 \neq 1$ and so $\frac{x}{2} \in A$, contradicting the choice of x . \square

Proof of Theorem 4.4 Suppose $f(x)$ is not identically zero. By the definition of ϕ , Lemma 4.11 and Corollary 4.10 it follows that for all $x \geq 0$, $f(x) = f(1)^x$. Moreover, since $f(1) > 0$ we can write

$$f(x) = e^{ax}$$

where $a = \ln f(1)$. This completes the proof of Theorem 4.4.

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 4.12. 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 4.4. Moreover, by Corollary 4.3 $0 < \delta(1) < 1$ 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. \square

Theorem 4.12 shows that, under mild assumptions, the discount function is an exponential. We note that a similar result was derived by [6, 9] in the case of discrete time.

4.1 An Alternate 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

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

where $\psi : \mathbb{R}^+ \cup \{0\} \rightarrow [0, 1]$. 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 4.12

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

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.

5. AGGREGATION

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. We refer to this operation as *aggregation* and denote the result by $X(t) \diamond Y(t)$ where \diamond is an application-dependent operator. Natural instances of \diamond include $+$, \max , \min , XOR , OR , among many similar ones.

5.1 Basics of Aggregation

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

$$X(t) \diamond Y(t) = [X(r)\delta(t-r)] \diamond [Y(s)\delta(t-s)]. \quad (9)$$

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

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

Associativity: $[X(t) \diamond Y(t)] \diamond Z(t) = X(t) \diamond [Y(t) \diamond Z(t)]$ for all $X(t), Y(t), Z(t)$. 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 $X(t) \diamond Y(t) \diamond Z(t)$ instead of the cumbersome parenthesized expressions. More generally, we shall write $\diamond_{i=1}^n X_i(t)$ instead of $X_1(t) \diamond X_2(t) \diamond \dots \diamond X_n(t)$.

Idempotency: If $Y(t) = 0$ then $X(t) \diamond Y(t) = X(t)$. Aggregation with information of value 0 has no effect. This is the view we take in this paper. We note however, that the Idempotency law may be extended to read $X(t) \diamond Y(t) = X(t)$ whenever $Y(t) \leq X(t)$, mirroring our intuitive idea that one does not stand to gain by aggregating with information of lesser value.

It is clear that the aggregation operator induces an abelian semi-group structure on the set of non-negative reals.

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 $0 \leq t \leq \tau$, 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 following result is a direct consequence of Lemma 4.1 and of the distributivity property.

Lemma 5.1. For all $0 \leq r \leq s \leq t \leq \tau$, we have

$$[X(t) \diamond Y(t)] \delta(\tau-t) = [X(r)\delta(\tau-r)] \diamond [Y(s)\delta(\tau-s)]. \quad (10)$$

Proof. By using distributivity we write

$$\begin{aligned} & [X(t) \diamond Y(t)] \delta(\tau-t) \\ &= [X(r)\delta(t-r) \diamond Y(s)\delta(t-s)] \delta(\tau-t) \quad [\text{by (2)}] \\ &= [X(r)\delta(t-r)\delta(\tau-t)] \diamond [Y(s)\delta(t-s)\delta(\tau-t)] \\ &= X(r)\delta(\tau-r) \diamond Y(s)\delta(\tau-s) \quad [\text{by Lemma 4.1}] \end{aligned}$$

\square

The left-hand side of (10) is the discounted value of $X(t) \diamond Y(t)$ at time τ , while the right-hand is the aggregated value of the discounted values of $X(r)$ and $Y(s)$ at time τ .

5.2 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, \dots, X_n , collected, respectively, at times t_1, t_2, \dots, 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 $t \geq \max\{t_1, t_2, \dots, t_n\}$.

Theorem 5.2. *The discounted value $V(t)$ of the aggregated information at time t is*

$$V(t) = \diamond_{i=1}^n X_i(t_i) \delta(t - t_i), \quad (11)$$

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 5.1. 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, \dots, G_p each of them the result of a previous aggregation at times, respectively, u_1, u_2, \dots, u_p . Observe that we can always relabel the groups in such a way that their aggregation times are ordered as $u_1 < u_2 < \dots < u_p$.

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

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

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

$$\begin{aligned} V_k(t) &= \left[\diamond_{j=1}^{n_k} X_{k_j} \delta(u_k - t_{k_j}) \right] \delta(t - u_k) \\ &= \diamond_{j=1}^{n_k} X_{k_j} \delta(u_k - t_{k_j}) \delta(t - u_k) \quad [\text{by distributivity}] \\ &= \diamond_{j=1}^{n_k} X_{k_j} \delta(t - t_{k_j}) \quad [\text{by Lemma 4.1}]. \end{aligned}$$

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 5.2, 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 5.2 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.

6. LET'S PUT 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, \dots be, respectively, be the sensed temperature values collected by the various sensors at times t_1, t_2, \dots . Since the sensors have witnessed the same event, it is natural to assume that the

random variables X_1, X_2, \dots come from the same underlying distribution X with finite expectation $E[X] < \infty$.

We do not assume that X_i s are independent. And with good reason: because of their proximity to the fire and of the fact that the collected data pertain to the same event, they are likely highly correlated. However, we assume that the X_i s are independent of the times t_1, t_2, \dots .

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 (4.12) the discounted value of this information at a later time t is

$$X_i(t) = X_i(t_i) e^{-\mu(t-t_i)}. \quad (12)$$

To make information aggregation as efficient as possible, and to minimize collisions of irrelevant packets, we mandate the sensors whose collected temperature reading is between, say, 70°C and 150°C to attempt to aggregate their readings with those of other sensors. The other sensors will not participate in aggregation. This strategy is justified by the fact that temperatures less than 70°C may not be indicative of fire.

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 100°C , this requirement is tantamount to insisting on accumulating a total of $\Delta = k \times 100$ "temperature point" as a result of aggregation. In turn, this suggests $\diamond = "+"$ as a suitable aggregation operation.

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 $t \geq \max\{t_1, t_2, \dots\}$. To answer this natural question we make the simplifying assumption that t_1, t_2, \dots 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. For simplicity, we assume that, in fact, all the sensors have detected data in the desired reporting range. The general case, where a random subset of the sensors have data in the reporting range, will be dealt with in the journal version of this work.

Theorem 6.1. *The expected time-discounted value, $E[V(t)]$, of the information collected by sensors at times t_1, t_2, \dots is*

$$E[V(t)] = \frac{\lambda}{\mu} E[X] [1 - e^{-\mu t}] \quad (13)$$

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, \dots .

Proof. Assume that the sensors collect their data at the times of a Poisson process with parameter λ . By the Law of Total Expectation,

$$E[V(t)] = \sum_{n \geq 1} E[V(t) | \{N = n\}] P[\{N = n\}] \quad (14)$$

where N is the random variable that counts the number of sensors that have data ready for aggregation by time t . By (12), Theorem 4.4 the conditional expectation, $E[V(t) | N = n]$, can be written as

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

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

$$\begin{aligned}
& E[V(t)|\{N = n\}] \\
&= \sum_{i=1}^n E[X_i(t_i)e^{-\mu(t-t_i)}] \\
&= \sum_{i=1}^n E[X_i(t_i)e^{-\mu(t-U_i)}] \\
&\quad \text{[where the } U_i\text{s are uniform in } (0, t]\text{]} \\
&= \sum_{i=1}^n E[X_i(t_i)]E[e^{-\mu(t-U_i)}] \\
&\quad \text{[because the } X_i\text{s and } U_i\text{s are independent]} \\
&= \sum_{i=1}^n E[X]E[e^{-\mu(t-U_i)}] \\
&\quad \text{[recall, } X \text{ is the common distribution of the } X_i\text{s]} \\
&= E[X]e^{-\mu t} \sum_{i=1}^n E[e^{\mu U_i}] \\
&= e^{-\mu t} E[X] \sum_{i=1}^n \int_0^t e^{\mu u} \frac{du}{t} = e^{-\mu t} E[X] \sum_{i=1}^n \frac{e^{\mu t} - 1}{\mu t} \\
&= \frac{E[X]}{\mu t} \sum_{i=1}^n [1 - e^{-\mu t}] \tag{15} \\
&= \frac{nE[X]}{\mu t} [1 - e^{-\mu t}]. \tag{16}
\end{aligned}$$

On plugging (15) back into (14), 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_i s does not appear explicitly in Theorem 6.1. This is telling us that two quite different distributions with the same expectation are equivalent as far as Theorem 6.1 is concerned;
- $E[V(t)] = \frac{\lambda}{\mu} E[X] [1 - e^{-\mu t}]$ is an increasing function of time and

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

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

Note that, as mentioned above, Theorem 6.1 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

$$t \geq \frac{1}{\mu} \ln \frac{\lambda E[X]}{\lambda E[X] - \Delta \mu} \tag{17}$$

In fact, (17) states that a value of t exists only if $\lambda E[X] > \Delta \mu$ or, equivalently,

$$\Delta < \frac{\lambda}{\mu} E[X]. \tag{18}$$

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

7. SIMULATION RESULTS

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

7.1 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 from $\mathbb{R}[30, 150]^\circ\text{C}$. For simplicity, temperature values are used as the value of information, and as discussed in Section 6, we use $\diamond = "+"$ as the aggregation function. In our scenario, an alarm is generated if the value of aggregated information for temperatures above 70°C exceeds 400. The value of information degraded exponentially with parameter $\mu = 5 \times 10^{-3}$.

Figure 1 shows the value of the discounted information over time (as described in Section 4). 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.

7.2 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-negligible 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

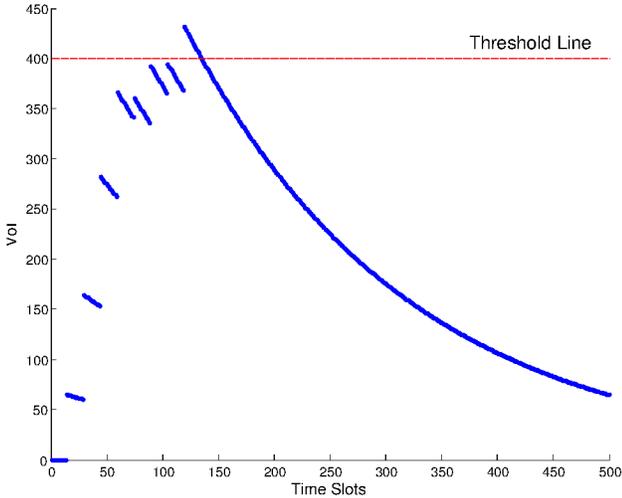


Figure 1: Value of aggregated information over time

for aggregation. For example, the aggregation method may use one of the following strategies:

- I) wait until at least k nodes have reported;
- 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 sensor nodes can decide about reporting in a way that only the effective ones report. For example, in the former strategy, 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 policy, transmitting fewer messages would reduce the traffic and result in faster reception of required information by the aggregator. 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 time to aggregate, which consequently avoids losing the value of information. This simulation scenario is designed in a way that is independent of aggregation method and function.

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 the aggregator. The number of sensors varies 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, if sensors are programmed to report their temperature only if they sense temperature between 100°C to 150°C , then in a fire detection scenario, only sensors that

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)

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 2 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 also increases to 4.7s. Figure 2 also 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 longer average time to aggregate means lower value of aggregated information.

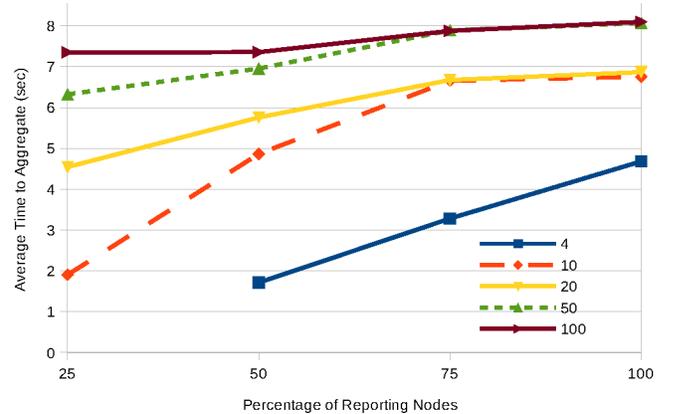


Figure 2: Average time to aggregate for various number of nodes. At least two nodes are required for aggregation.

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 losing the value of information. Figure 3 illustrates the effect of those delays on the loss of information value for several exponential decay functions.

As can be viewed even with 2 seconds of delay, the value loss can be up to 80% of the value of information. This delay can be because of waiting for pieces of information to be generated from different sources and arrive at the aggregator as well as delay due to wireless channel condition. In the next simulation, we investigate how the rate of information generation in relation with decay should be designed to

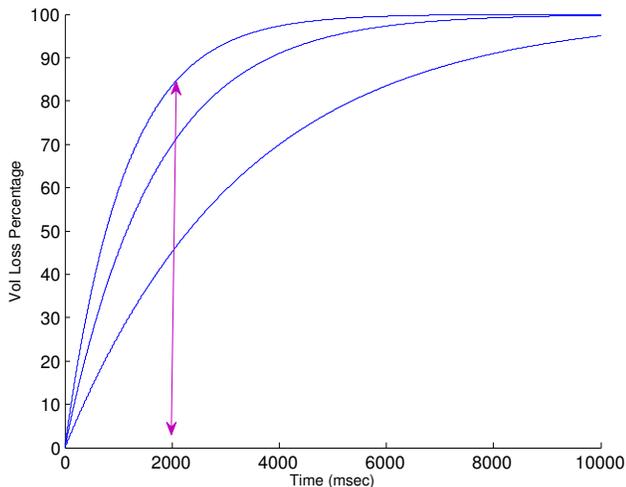


Figure 3: Percentage of value of information loss over time for several exponential decay functions

achieve the value of information as soon as possible, so that a decision maker can make a proper determination.

7.3 Time to Exceed a Threshold

Here, we evaluate the applicability of the defined threshold time in (17). Recall that (17) 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 an intruder 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 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 (17) 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 from $\mathbb{R}[V_{min}, V_{max}]$, 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 98% confidence in detection of intrusion before reporting. Letting $VoI(t)$ stand for the random variable that measures 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 .

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.

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

Config	λ	μ	Theory	Simulation
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

8. CONCLUDING REMARKS AND DIRECTIONS FOR FUTURE WORK

This paper provided a formal way of looking at the aggregation processes 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 takes 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. First, it would be natural to model the aggregation as a renewal-reward process where each aggregation brings a certain “reward”, specifically, an increase in the value of the corporate information stored by the sensor network. In order for this to yield meaningful results, we need a reasonable model for the distribution of the rewards. While this is highly application-dependent, we feel that a result in the spirit of Theorem 6.1 is possible.

Also, 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 Weierstrass’s classic approximation theorem [3] that every real function can be approximated by a suitable sequence of polynomials. Since we have a good understanding of 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 exiting area of work.

9. 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] T. Apostol. *Mathematical Analysis*. World student series. Addison-Wesley, 1974.
- [4] 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.
- [5] S. Frederick, G. Loewenstein, and T. O’Donoghue. Time discounting and time preference: A critical review. *Journal of Economic Literature*, XL:351–401, 2002.
- [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 and J. V. Nickerson. A probabilistic model of integration. *Decision Support Systems*, 45(4):746 – 763, 2008.
- [10] 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.
- [11] R. Rajagopalan and P. K. Varshney. Data aggregation techniques in sensor networks: A survey. *Comm. Surveys & Tutorials, IEEE*, 8:48–63, 2006.