12TH ICCRTS
"Adapting C2 to the 21st Century"

Title of Paper: The Impact of Entropic Drag on Command and Control (DRAFT)

Paper ID: I-020

Track(s): Network Centric Experimentation and Applications, C2 Concepts Theory and Policy, Organizational Issues

Authors: David Scheidt, Michael Pekala, Joseph Peri

Point of Contact: David Scheidt
Name of Organization: Johns Hopkins University Applied Physics Laboratory
Complete Address:
Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road,
Laurel, MD 20723

Telephone: (240) 228-7301
E-mail: david.scheidt@jhuapl.edu

Abstract: A commonly believed axiom in signal detection theory is that "more information is good" [1]. That is, when attempting to determine the state of a partially observable system the addition of correct information monotonically improves the correctness of the state assessment. When diagnosing static systems the assertion that "the effect of information is to increase the likelihood of getting correct diagnosis, while reducing the likelihood of incorrect diagnosis" holds. However, when diagnosing a dynamic system, the improvement in diagnosis is offset by increases in uncertainty that is generated by the dynamic forces within the system being observed. A similar effect occurs when *controlling* dynamic, stochastic systems. The act of exercising control requires a finite amount of time, during which uncertainty enters into the system reducing the efficacy of the control policy. The impact of these information processing delays increases in relevance as the complexity, and pace of both the system and control apparatus increase. This paper defines a mathematical framework describing the effect of information and information processing on the diagnosis of dynamical systems.

Note to Reviewers: We apologize for the incomplete state of this paper. This paper is a formal documentation of research conducted at during 2006. Prior to the final paper deadline in April this paper will undergo several substantial modifications, notably: (1) incorporation of a series of brief, easily understood examples that explain the principles described; (2) incorporation of a section on information entropy and control; (3) incorporation of the section on entropy and channel communications that is included as an appendix; incorporation of graphs with more accurate curves [Fig 1-7]; references and associated bibliography. If requested, a more intermediate version of the paper can be provided for comprehensive review on March 1st.

[1] Heeger, D., Signal Detection Theory, http://www.cns.nyu.edu/~david/sdt/sdt.html, New York University.

Information and Entropy

An understanding of the information required to represent a system begins with a formal definition of the system. We refer to the system being diagnosed and/or controlled as the *world (W)*. A world is composed of a finite set of *elements* $W=\{e_1,e_2,..e_n\}$ from the set of possible elements E $e \in E$. Elements, in turn, are described by a tuple of *n* attributes $x_e=(x_1,x_2,..x_n)$ which are in turn each members of their own defining set $\forall x_i \in X_i=\{a_{i,1},\ a_{i,2},...,a_{i,n}\}$ Attributes provide descriptions of elements such as location and orientation. Attributes are finite, constrained values which collectively define the scope of the world. The *state* of an element (x_e) is a unique set of values for all attributes. Transitions between states are made by events (E). Laws that govern state transitions are mapped through a transition function

$$f : X \times E \rightarrow X \qquad (1)$$

The state space (P_e) of an element is the Cartesian product of the attribute sets which is the number of unique, states within X_e.

$$P_e = X_{e_1} \times X_{e_2} \times ... \times X_{e_n} \qquad (2)$$

The state space of the world is the power set of the elemental state space for all elements:

$$P_w = P_{e_1} \times P_{e_2} \times ... \times P_{e_n} \}; \forall P_e \qquad (3)$$

A world's state space may be divided into states which are feasible (P_f) and infeasible (P_i). P_f and P_i are mutually exclusive and comprehensive s.t.

$$P_f \cap P_i = \{\} \qquad (4)$$

$$P_w = P_f \cup P_i \qquad (5)$$

The information content (I) of a message is the amount of information required to uniquely identify a state. At this point we only consider messages whose information content is ideally coded for the set of possible values that could be stored in the message. That is, the representation scheme uses the minimum number of bits to completely represent the set of possible values. Information content is defined by Brillion as:

$$I = K \ln |P| \qquad (6)$$

In which *K* is a constant associated with the size of the language used to represent a world state. In modern computers and communications this language is binary allowing us to define the information content is defined as:

$$I = \log_2 |P| \qquad (7)$$

Again through Brillion we can see that the amount of feasible states (P_f) that exist for a world after a message (m) describing that world has been received is:

$$|P_f| = |P_w| - 2^{I_m} \qquad (8)$$

The feasible space represents disorder, lack of knowledge, or *information entropy* (H). Using (6) we define the post-message entropy of a world as:

$$H = \log_2 |P_f| = \log_2 |P_w| - I_m \qquad (9)$$

Information entropy is a useful construct that we will extend to collections of information and information exchanges.

Information entropy is similar to uncertainty (U) in that they both measure the disorder in a system. They differ in the units of measurement. Information entropy is measured in bits while uncertainty is measured in natural units for the system being described (e.g. meters).

Equation (9) shows how new information can be used to reduce entropy within the representation of a world. In static worlds the information content of a message that contains positive information (I_m>0) decreases information entropy. For ideally coded messages, providing information about a previously unknown world (worlds in which H = $\log_2 P_w$) the decrease in information entropy is equal to the size of the message.

$$\Delta H = \begin{cases} I_m ; |P_w| \geq 2^{I_m} \\ 0 ; |P_w| < 2^{I_m} \end{cases} \qquad (10)$$

If a priori knowledge of the world exists prior to receipt of the message some portion of the message may be redundant, reducing the message's information content. The information content of a message about a world for which a priori information is expressed as:

$$\Delta H = \log_2 \lfloor P_f \cap p(I_m) \rfloor - I_m \quad (11)$$

Where $p(I_m)$ is a decoding function that produces a set of <u>*feasible*</u> states from a message I_m.

Uncertainty Due to Fidelity

In practical terms worlds are represented in terms of abstractions. In modern communications semantics for encoding attributes are defined (elements and attributes in this paper) and encoded in bits. Some real-world information is naturally discrete. Other phenomenon is naturally continuous and can only be approximated with binary representation. The amount of uncertainty associated with discrete representation of continuous attributes (ε_r) is limited by the number of bits used to encode the data and the dimensionality of the data. For example, when communicating attributes describing positions in Cartesian space our message must be transmitted digitally at some point limiting the fidelity by the least precise representation[1] used in the communications path which is the unit distance ($\hat{r}$) of the representation. The unit distance defines the minimum uncertainty of our knowledge such that the minimum amount of error is:

$$\varepsilon_r = \frac{\sqrt{n}}{2}\hat{r}; r \in Z^n \quad (12)$$

where r is the unit size of representation and n is the dimensionality of the space.

[1] In this discussion we focus on integer representation. Other representations, such as floating point, can provide higher fidelity for portions of their range but do not provide higher fidelity *on average*.

Entropic Drag

So far we have limited our discussion to static worlds. When observing static worlds information entropy is decreased by each successive message that describes the world. In dynamic worlds the decrease in entropy achieved from a message is offset by an increase in entropy due to the world's dynamic forces. We call this increase in information *entropic drag* (Γ) as it is a drag on an observer's ability to understand the world. Entropic drag is derived from the rate at which the dynamic forces in the world create unpredictable change. To define these dynamic forces recall that a world's state space P_w is divided into feasible and infeasible states and that the transition between these states is Markovian, defined by the transition function. For any set of infeasible states there exists a boundary layer between feasible and infeasible states. The rate at which infeasible states transition to feasible states is the driving force behind entropic drag. We formally define entropic drag as the log of the rate at which previously infeasible world states become feasible:

$$\Gamma_w(t) = H_w(i)\frac{di}{dt} = \log_2 P_f(x)\frac{dx}{dt} \quad (13)$$

The impact of entropic drag on information is shown as a decrease in the information content on a world after an observation s.t.

$$I = (I_0 + I_m)\cdot\left(1 - \int_0^{t_m} \Gamma(t)dt\right) \quad (14)$$

where t_m is the time required to process the message, I_0 is the information content on the world prior to processing the message and I_m is the information content of the message. An important feature of this equation is that, if the time required to process a message is sufficiently long, the loss of information associated with entropic drag will exceed the information content of the message and the net result of processing the message will be an information loss!

Obtaining and using information takes some finite amount of time. The process of obtaining and using information was described by Boyd as the Observe, Orient, Decide and Act (OODA) loop. The first part of this loop consists of a sensor observing the environment, transmitting a message to an information sink where it is fused with additional information to create a consistent model of the world. The delay between an observation (t_o) and the incorporation (t') of the observation's data is the sum of the sensor processing (δ_s), communications (δ_c) and data fusion (δ_f) delays.

$$t' = t_0 + \delta = t_0 + \delta_s + \delta_c + \delta_f \quad (15)$$

The sensor processing and communications channels are assumed to be able to process information at a fixed rate measured in bits per second. The data processing profile for data fusion varies by technique, with some techniques such as nearest-time replacement operating at a fixed rate and others, such as search-based techniques, incurring exponential increases in latency as the number of observations increase. For the purpose of this paper we will assume that each element in the observe-orient chain operates at a fixed rate in bits per second.

$$\beta = \frac{I_m}{\delta} \quad (16)$$

By applying this definition to (14) we express information content in terms of message size and bandwidth.

$$I = (I_0 + I_m) \cdot \left(1 - \int_0^{\frac{I_m}{\beta}} \Gamma(t)dt \right) \quad (17)$$

and

$$\Delta I = I_m - (I_0 + I_m) \int_0^{\frac{I_m}{\beta}} \Gamma(t)dt \quad (18)$$

From (18) we can see that information is only gained from a single message if the rate of information entropy is less than the bandwidth.

$$\Delta I > 0 \text{ iff } \beta > \Gamma \quad (19)$$

Information entropy progresses along a frontier between the feasible state space and the infeasible state space. The frontier (**S**) is set of piecewise smooth surfaces ($\mathbf{s}_i \in \mathbf{S}$) within P_w that are defined as the boundaries between infeasible states $x_i \in P_i$ that are reachable from a feasible state $x_f \in P_f$ through an event in accordance to the transition function. Each piecewise smooth surface is either a member of a set of surface that form a closed n-dimensional surface that encapsulates positive or negative information or a member of a semi-closed surface whose edges abut limits of P_w space. **S** changes at a rate that is defined by the entropic drag with the boundary perpetually growing the amount of feasible space and shrinking the infeasible space.

The entropic drag for a surface $\mathbf{s}_i$ is

$$\Gamma(s_i) = H(s_i)\frac{dh}{dt} = \log_2 s_i \cdot \vec{v}\frac{dx}{dt} \quad (20)$$

where $\vec{v}$ is the n-dimensional vector of unpredictable change normalized to s_i. The entropic drag for the entire world is the sum of the entropic drag for each surface:

$$\Gamma_w(s_i) = \sum_{\forall s_i} H(s_i)\frac{dh}{dt} \quad (21)$$

Positive and Negative Information

Positive information (P^+) are assertions that one or more states are true.

$$P^+ = \{x_1, x_2, .., x_n\} \quad (22)$$

For example, a message stating that an ant was observed at time t_0 is positive information. Negative information (P^-) are assertions that a set of states are false.

$$P^- = \{\neg x_1, \neg x_2, .., \neg x_n\} \quad (23)$$

For example, a message stating that an observation at time t_0 at a specific location found an element did not exist is negative information. World states can be expressed either through positive information or negative information and in fact observations frequently produce a mix of positive and negative information. Positive and negative information both create infeasible regions in the state space and the information gain from a message is equal to the reduction in feasible states. If the message encodes negative information the information gain is equal to:

$$I_m^- = \log_2 \left| P^- \right| \qquad (24)$$

whereas the information gained from positive information is equal to the size of the space outside of the positive information set:

$$I_m^+ = \log_2 \left| P_w \right| - \log_2 \left| P^+ \right| \qquad (25)$$

As positive information entropies information is increasingly lost at a rate that increases in proportion to the dimensionality of the original message [Fig. 1].

This can be envisioned as the feasibility frontier growing outward from a clustered set of feasible states.

As negative information entropies information is decreasingly lost in proportion to the dimensionality of the world [Fig. 2]. This can be envisioned as the feasibility frontier growing inward, removing a diminishing set of clustered infeasible states.

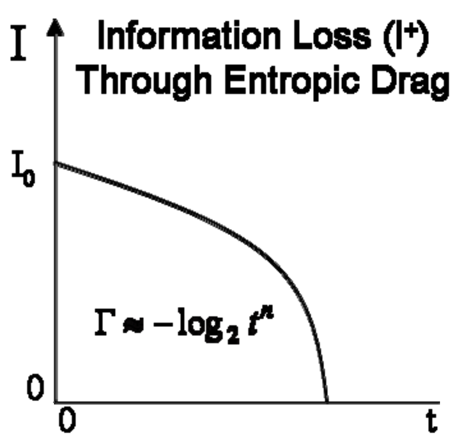

Figure 1

Each closed surface $\mathbf{s_i}$ either grows (for I^+) or shrinks (I^-) due to entropic drag until terminates in a singularity, terminates against other surfaces, or terminates against the boundaries of the state space. The information loss for the world is a piecewise regular continuous function that is the sum of the information loss for each surface.

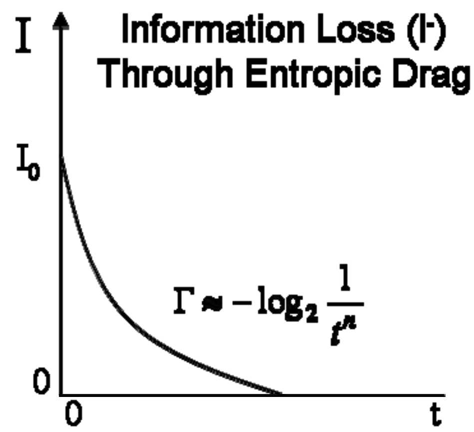

Figure 2

Message Streams

Information acquisition is typically bandwidth limited, where bandwidth is defined as the rate at which information can be obtained as shown in (15). In bandwidth limited environments a decision-maker has the ability to accept information at a continuous rate of β. In ideal conditions, information in static world increases linearly until the information capacity of the world is reached [Fig. 3].

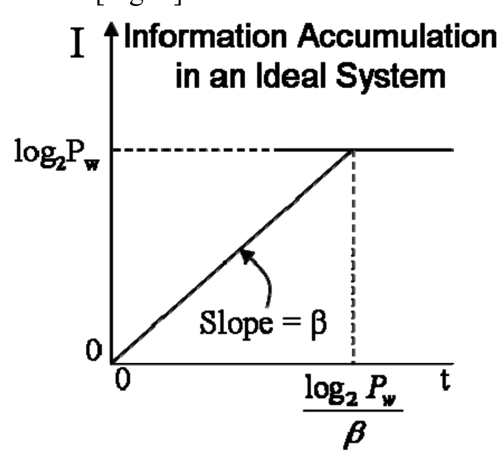

Figure 3

Concurrently with the increase in information content through the arrival of new messages information is lost through information entropy. The amount of information available on the world provided a constant stream of information is:

$$I = I_0 + \int_0^{\frac{I_m}{\beta}} \left(I_m - \Gamma(I_0 + I_m)\right)dt \quad (26)$$

As mentioned briefly above, edge effects can play an important role in the rate of entropic drag. We highlight two basic cases here. In the first case the information gathering bandwidth is sufficient to exhaustively describe the world before the information content of the first message has vanished in a singularity or against an edge. In this case the information bandwidth exceeds entropic drag. In this case the information content will become maximized when the information space has been transmitted.

$$t_{\max} \geq \frac{\log_2(P_w)}{\beta} \quad (27)$$

Information will monotonically increase until time t_{max} after which further messages can only maintain the current amount of information as the information gained by a successive message cannot exceed the combination of information lost due to entropic drag and the information loss through redundancy with prior information [Fig. 4].

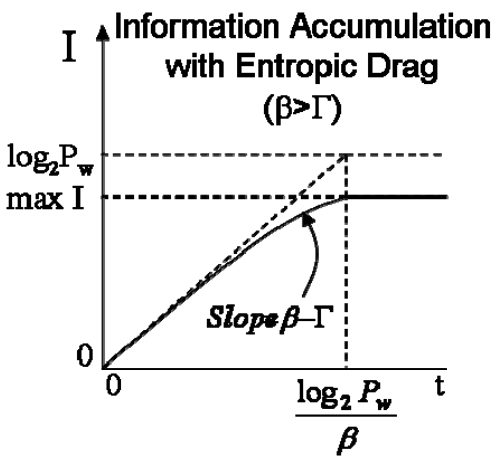

Figure 4

In the second case the bandwidth is less than the entropic drag. In this case the information content will reach a stable point of maximum information when the entropic drag equals the bandwidth [Fig. 5].

Note that in both cases the maximum information reached is less than the information space of the world. This gap is the minimum information loss associated with describing a dynamic world.

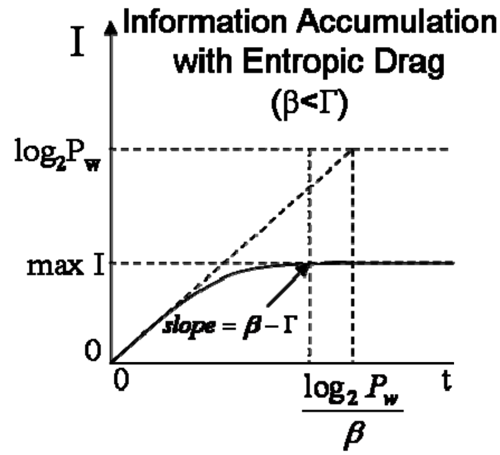

Figure 5

Uncertainty Redux

One of the design decisions involved in constructing a C2 system is selecting the fidelity of information being communicated. The fidelity is adjustable as the designer can arbitrarily decide the unit measure for each dimension in the world. For example, should the lowest bit of information about an entity be equivalent to a millimeter, meter or kilometer? Earlier we showed how a minimum amount of uncertainty (ε) is associated with the choice of representation. The minimum uncertainty of a system at some time t is the aggregation of the representational uncertainty of a system and the feasible state space translated from bits to real world units.

$$U = U_r + \hat{r} \cdot 2^H \quad (28)$$

The representational uncertainty is the product of the information frontier and the informational error:

$$U_r = S_t \cdot \varepsilon_r \quad (29)$$

By substitution uncertainty becomes:

$$U = S_t \cdot \varepsilon_r + \hat{r} \cdot \int 2^\Gamma dt \quad (30)$$

Both portions of the uncertainty equation (30) associated vary as a function of size unit vector. However, the rate at which they vary differs, with the uncertainty due to representational error becoming predominant as the unit representation grows and the uncertainty due to entropy becoming predominant as the unit representation shrinks. This duality allows us to identify the optimal unit representation for maximizing uncertainty.

Observation Inefficiencies

So far we have assumed that messages have been ideal, with each bit of an observation translating to a message bit that provides a single bit of information. In practice this is only the case when regular homogeneous worlds are being observed and no a priori about the world. When a priori knowledge exists, the efficiency of an observation is reduced by the number of bits that are redundant with the a priori information. When observations are controllable, as in the case of vehicle borne sensors, the order of observation may be controlled to observe

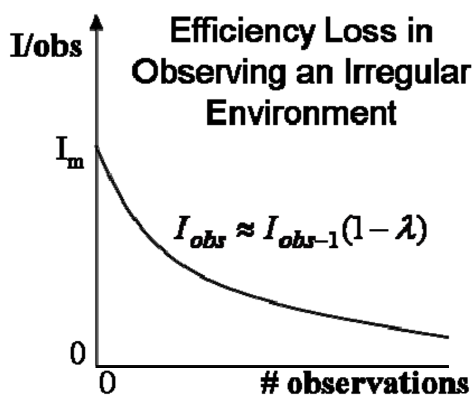

Figure 6

the portions of the world with the biggest potential payoff (least amount of a priori redundancy) first. This loss in efficiency is highly dependent upon the situation. In order to explore entropic relationship inefficient systems we notionally we describe this efficiency loss of as a decay function [Fig 6]. We replace the constant information value in equations (17) with a function $g(i,t) \rightarrow i$ to show the relationships between information variant messages and information entropy.

$$I = \int_0^t \left(g(I_m, t) - \Gamma(g(I_m, t)) \right) dt \quad (32)$$

When deploying a sensor motion strategy that exhaustively searches the world the entropic drag can become so large that it overcomes information gain. This effect is shown in [Fig 7] which shows the effect of a linear entropic drag on the observation environment shown in the previous figure. As shown in the figure shows that the maximum amount of information is found after an ideal number of observations have taken occurred. If the goal of a command and control system is to enable a decision maker to make decisions based upon the most complete information possible, immediately after this observation would be the time to make decisions.

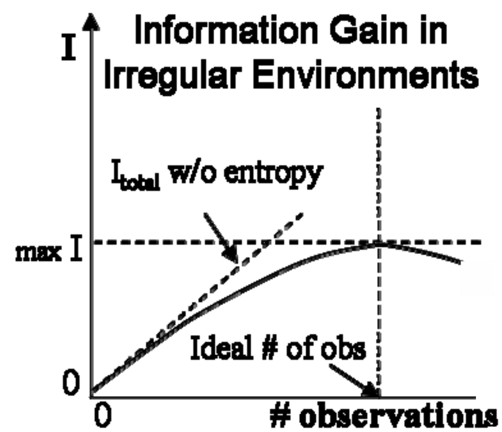

Figure 7

While we are use a linear function in our example, we recognize that the information function may be highly non-linear. However, non-linearity does not negate the principle shown here.

Entropic Drag and Network Topologies

Simulation experiments were conducted to examine the effect of entropic drag on a simple command and control network infrastructure. These experiments examined the flow and relevance of *information* throughout the network. The bulk of the simulation's requirements lie in the terms "C2 network infrastructure" and "information".

A command and control network infrastructure consists of a set of communicating entities (or nodes) whose communications topology forms an acyclic tree. The lines of communication between nodes (edges) also correspond to the chain of command. Thus, a node's authority in the network is inversely proportional to its depth in the communications tree (i.e. the root node of the tree has the highest authority and leaf nodes have the least authority). An additional property of the communications/command tree is that all nodes at the same depth in the tree have the same number of directly reporting subordinates (child nodes). This mapping of rank to number of immediate subordinates is the property that distinguishes alternate C2 infrastructures in the simulator.

Within the simulator, information consists of data gathered from the environment by leaf nodes in the C2 infrastructure. Each piece of information is an abstract quantity that is independent of other pieces of information (e.g. information does not overlap or correspond to multiple measurements of a known target in the environment as might be the case in a filtering problem). Information can be generated by leaf nodes periodically or stochastically, depending on the simulator configuration. The value of each piece of information as a number in the interval [0,1], where 1 corresponds to maximum value and 0 corresponds to no value.

Another property of information is that it only becomes useful to an entity after the information has been fused into the local world model. Thus, information is subject to two primary sources of latency before it can increase the knowledge of a network entity: latency due to network communications and latency due to the local information fusion algorithm.

Simulation Structure

The C2 network information simulation is a modular, discrete timestep simulation whose primary components are: network entities, communications links and information processing algorithms. Figure 1 depicts the actions that occur each timestep. In addition, the simulation tracks a number of metrics for each piece of information. Two key metrics include the information area (the number of nodes that finished fusing the information in the current timestep) and the information volume (the number of nodes that have fused the information at or before the current timestep). These information area and volume are metrics are taken from the literature on the performance of real-world networks. The overall value of a piece of information is derived by multiplying the information volume by the associated entropic drag. Since these area and volume metrics are affected by the number of nodes in the network, when comparing different C2 topologies we typically consider topologies with the same number of leaf nodes and then restrict the results to leaf node areas and volumes.

The simulation has a number of input parameters that can be varied. Primary simulation parameters include: Simulation duration (timesteps), *C2* topology, entropic drag (value lost/timestep), Latency along a network communications link (timesteps/observation) and Latency due to fusion. Unlike other simulation parameters which are constant, fusion latency can optionally be a function of the number of prior fusion operations (allowing for nonlinear fusion complexity).

Simulation Outputs

Figure 8 shows a sample information volume plot for a single simulation run. These plots generally contain three pieces of information: the ideal information volume (red line), the information volume without accounting for entropic drag (blue line) and the information volume with entropic drag (green line). The ideal information volume depicts the spread of information throughout the network assuming no latency due to communications or fusion and hence no entropic drag. Thus, it provides an (admittedly unrealistic) upper bound on the amount of information in the network. Nonetheless, it provides a strong indication that an upper bound exists. The undecayed information volume depicts the network information volume that would result when considering latency but not entropic drag. The decayed information volume takes into account both the effects of latency and entropic drag, and represents the main result of interest. One expects both the ideal and undecayed information volumes to monotonically increase, with the entropic drag volume trailing behind the ideal volume. The entropic drag volume shows that there exists a maximum information volume (time=175) for the system. Further, it shows that this peak occurs prior to the complete distribution of information across the network across the network (time=240).

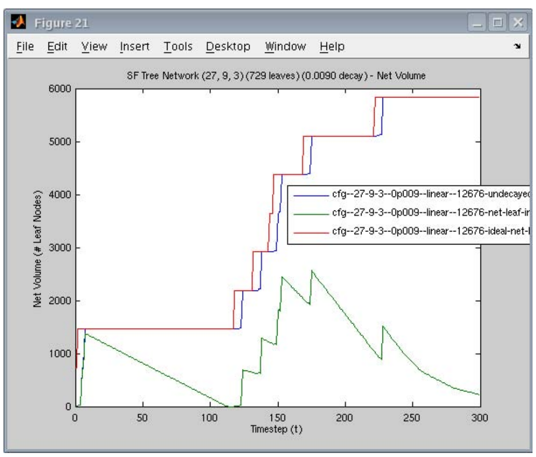

Figure 8

A key feature of the simulation output is the ability to compare one or more simulation runs. Figure 9 is a plot comparing multiple runs of the simulation where the information decay rate (δ) was varied while the other simulation properties were held constant. In the plot the upper lines in the plot have progressively lower decays. Figure 10 shows the same data plotted in three dimensions with the information decay rate plotted along the y axis. The appropriate plot type can be selected based on the parameters being varied.

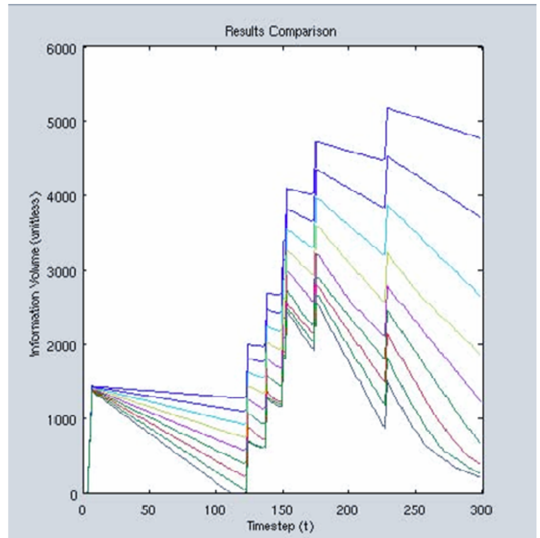

Figure 9

In figure 9 one can again see the existence of maximum information volume prior to full dissemination of information across the network. We can also see that the time at which the information maximum occurs varies with respect to the entropic drag as he information maximum occurs at time step 230 for the runs with lower information drag (purple lines at the top of the graph), at time step 175 for the runs with higher entropic drag (lines at the bottom) and equally across time step 230 and 175 for the cyan line.

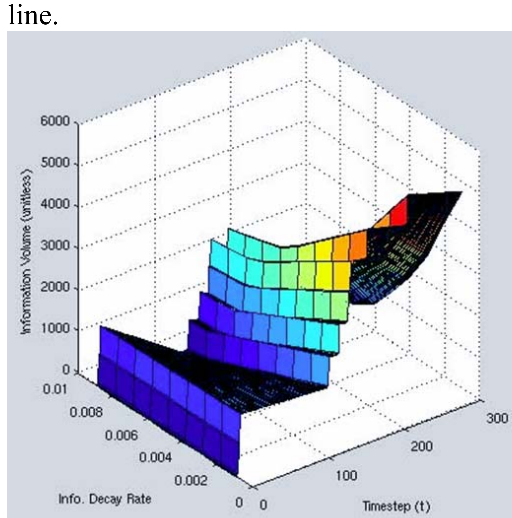

Figure 10

Future Efforts

This paper is a first step in the application of information theoretic entropy to command and control. Large bodies of work in information management, particularly data fusion, networking and the pantheon of group control strategies needs to be looked at though the lens of entropic drag. Our forward looking hypothesis is that an understanding of the entropic effects of information will allow C2 designers and in-the-field decision makers to employ command and control strategies that are optimized for any given situation. Further work will also be required to improve the metrics that are used to measure the dynamicism within an environment. Environmental dynamics are driven by complexity and the pace of environmental change; however, is not well understood how complexity and pace should be measured in real-world environments. These further advancements should enable the pursuit of our long-range objective, the construction of an adaptive command and control system that autonomously observes the environment and changes the network topology and information and decision-making strategies to optimize C2 performance. Finally, while we have theoretical and simulated evidence that indicates the importance of entropic drag to C2, to date we have only investigated theoretical environments and have not attempted to apply these principles to real world problems.

Conclusion

We have shown that the loss of information due to dynamic forces within an environment can have a substantial impact upon the information content of one or more messages about the environment. This effect, called entropic drag, fundamentally impacts the effectiveness of command and control systems. The principles outlined in this paper can be used provide a better understanding of the utility of existing command and control systems and to improve the design of future command and control systems.

Entropic drag impacts C2 systems in several important ways. First, entropic drag enforces a fundamental limit to the amount of information that can be known about a dynamic system. This limit can be used by C2 designers to identify the maximum useful fidelity of C2 semantics. Second, by expressing the relationship between information and time entropic drag allows a decision maker to identify when the optimal amount of information has been acquired. Third, entropic drag provides a framework for understanding the flow of information across a networked community, providing insight into the utility (or lack thereof) of sharing information with each member of the community as well as providing insight into the utility of alternative network topologies. Fourth, entropic drag provides tools to better understand the tradeoff between information and latency in control, allowing decision-makers to select the optimal amount of information to use for performing a specific task or set of tasks. Fifth, entropic drag provides tools to understand the utility of collaboration in shared decision making, allowing decision makers to correctly scope the degree of collaboration for optimal performance of a task. Finally, entropic drag provides a framework for the development of a next generation adaptive command and control infrastructures, infrastructures that autonomously adapt information and decision sharing strategies and the networking topology at run time in response to an environment's changing dynamic forces.

Acknowledgements

This research effort was supported by Johns Hopkins University Applied Physics Laboratory Internal research grant X9ANXC2M.

Appendix - Channel Communications

0. Abstract

We use information theory to discuss the technical problem of accuracy of transference of signals from sender to receiver. The development follows closely that of Kolmogorov. The basic concept is the quantity of information in one random object relative to another random object. The entropy of a random object is specialization of this quantity. From these concepts, one establishes conditions under which messages can be reproduced with arbitrarily small probability of error.

1. Introduction

The Observation, Orientation, Decision and Action (OODA) loop is a description of the ability to militarily act and react more rapidly than an opponent. Modern warfare is characterized by the sharing of information among teams. Wide-band, high-speed communication networks can rapidly disseminate large quantities of information. However, team performance can be affected by information overload, processing delay, complexity of information and complexity of decision. To address these issues, we go back to the basics of information modeling.

To model the flow of information from one node to another within the OODA loop, one cannot overlook the seminal contributions of C. E. Shannon (Reference j). The development of information theory was prompted by practical problems in the fields of electrical and radio communications. In this paper, we use information theory, or the Shannon theory of optimal coding of information, to calculate information rate. Specifically, we follow the approach developed by Kolmogorov (References g, i) and his students Dobrushin (References b, c, d), Gel'fand and Yaglom (Reference e) and Pinsker (Reference h). Kolmogorov develops a general definition of information for a relatively broad class of random objects, from which one derives the entropy of a random object. This approach maximizes the practical applicability of the concepts.

We begin with two practical examples for motivational purposes. Then, we describe the physical portion of the problem that leads us to probability values. The transmission problem is, then, described probabilistically, whereby the joint distribution of various random variables is developed. The joint distribution is then used to construct the quantity of information of one random object with respect to another. Finally, we discuss the fidelity criterion used in determining the optimal encoding and decoding schemes. We confine ourselves to the simple case of discrete random variables ranging over finite sets, to discrete memoryless sources and to discrete memoryless channels.

2. Motivation

A communications system consists of a message *source*, a *decoder* that converts the input message into a signal suitable for transmission, a *channel* through which the input signal propagates and becomes corrupted with noise, a *decoder* that takes the channel's output signal and converts it into an output message suitable for the *user*. Figure 1 contains a diagram of the elements of a communications system.

2.1 Telegraphy Example

A message consists of a sequence of letters and spaces from the English alphabet. The transmitter (including the telegraph key) encodes the sequence into mechanical dots and dashes, converts the mechanical dots and dashes into time-varying currents and injects these into the channel. The channel introduces noise that corrupts time-varying currents that propagate along the channel. The time-varying currents arrive at the receiver. The receiver (including the telegraph key) converts the time-varying currents into mechanical dots and dashes. The mechanical dots and dashes are decoded into a sequence of letters and spaces from the English alphabet. This sequence is an approximation of the message which the source originally sent.

2.2 Image Transmission Example

An optical device projects an image of the real world onto a focal plane array. The focal plane array is a matrix of pixels that encodes the optical image into a mosaic. Because each pixel ejects a number of electrons, depending upon the magnitude of the incident light, the system converts the mosaic into time-varying currents. Fluctuations make the potential deviate from the normal value dictated by the magnitude of the light. The encoded image is already a noise-bearing variant of the original optical image. The system scans the mosaic. The transmitter injects the encoded signal into the channel. Noise corrupts the input signal as propagation takes place. The channel outputs the corrupted signal that now enters the receiver. The receiver decodes the channel output signal into a mosaic that is an approximation of the original optical image.

3. <u>The Physical Portion of the Communications Problem</u>

A radar system or an optical imaging system converts a real-world scene into a mosaic. This conversion occurs because of the inherent limitations of the measuring device. If the system scans the mosaic in search of an object of interest within the scene, one may ask: What is the probability that the object resides within an element of the mosaic?

If the total area of the image is the number A_{tot} and that of an individual element of the mosaic is A_{pix} (let's say the mosaic is comprised of pixels), then one can, heuristically, define the probability that the object is within pixel E_i by

$$p(\{E_i\}) = \frac{A_{pix}}{A_{tot}} \tag{1}$$

Suppose a typical image consists of a 256 × 256 pixel array. We may define the set of pixels as

$$\Omega = \{E_1, E_2, \cdots, E_{256\times 256}\}. \tag{2}$$

We define the probability space as the triple (Ω, S_Ω, p), where $p : S_\Omega \to [0,1]$ and S_Ω is a σ-algebra of subsets of Ω, i.e., a non-empty set of subsets of Ω closed under the formation of complements and countable unions. For our needs, we let S_Ω be the power set of Ω, i.e., the set of all subsets of Ω.

In the information-theoretic context, Ω represents the information source, the elements of which are mapped into the set X of input messages by the random variable ξ. The probability space, (Ω, S_Ω, p) and the mapping $\xi : \Omega \to X$ induce the probability space (X, S_X, p_ξ), such that,

$$\forall A \in S_X, \xi^{-1}(A) = \{\omega \in \Omega : \xi(\omega) \in A\} \in S_\Omega, \tag{3}$$

that is, for any $A \in S_X$,(S_X is a σ-algebra of subsets of X) the inverse image of A is a collection of elements of Ω the image of each of which, under ξ, is an element of A, and this collection is an element of S_Ω. From the point of view of Measure Theory (Reference h), the random variable ξ is a *measurable* function. The nomenclature varies among mathematicians. Typically, if $X = \mathbb{R}$, the set of real numbers, then ξ is called a random variable. Some authors use measurable function and random variable interchangeably. Figure 2 shows a diagram of this situation.

The induced probability function, p_ξ, is defined as

$$\forall A \in S_X, p_\xi(A) = p \circ \xi^{-1}(A) = p(\{\omega \in \Omega : \xi(\omega) \in A\}). \tag{4}$$

Since the set $\{\omega \in \Omega : \xi(\omega) \in A\}$ belongs to S_Ω, it is mapped to the unit interval; its probability is an actual number, and, therefore, so is $p_\xi(A)$.

4. Fundamental Shannon Problem

Kolmogorov (Reference g) addresses the Shannon problem as follows. Given sets X, Y, Y′, X′ of possible values of input message, ξ, input signal, η, output signal, η′, and output message, ξ′, respectively; given the characteristics of the channel described by the conditional probability distribution, $p_{\eta'|\eta}$, and a class *V* of input signal distributions, $p_{\eta\eta'}$; given the input message distribution

$$p_\xi(A) = p(\xi \in A) = p(\{\omega \in \Omega : \xi(\omega) \in A\}) \tag{5}$$

and the fidelity criterion, $p_{\xi\xi'} \in W$, where *W* is a certain class of joint distributions

$$p_{\xi\xi'}(C_X \times C_{X'}) = p((\xi,\xi') \in C_X \times C_{X'}) = p(\{\omega \in \Omega : \xi(\omega) \in C_X \wedge \xi'(\omega) \in \times C_{X'}\}) \tag{6}$$

where $C_X \subset X$ and $C_{X'} \subset X'$. Is it possible, and if so, how, to find encoding and decoding rules (i.e. conditional distributions $p_{\eta|\xi}$ and $p_{\xi'|\eta'}$), such that by calculating $p_{\xi\xi'}$ in terms of p_ξ, $p_{\eta|\xi}$, $p_{\eta'|\eta}$, under the assumption that the sequence of random variables ξ, η, η', ξ' is Markovian, one will obtain $p_{\xi\xi'} \in W$?

Shannon tells us that if transmission is possible then $H_W(\xi,\xi') \le C$; that is, if the entropy of one random variable with respect to the other is greater than the channel capacity then transmission, with arbitrarily small error, is not possible. As Berger (Reference a) explains, if the system designer is required to fulfill the fidelity criterion and has been provided with a channel of capacity *C*, he need merely compare $H_W(\xi,\xi')$ with *C* to determine whether or not he has any chance of succeeding. Only if $C \ge H_W(\xi,\xi')$ can he possibly succeed. Hence, knowledge of $H_W(\xi,\xi')$ can prevent him from expending time and resources in a futile attempt to accomplish an impossible task.

The physical portion of the problem provides a probability space (Ω, S_Ω, p) and measurable mappings (i.e. random variables) ξ, η, η', ξ' from Ω to X, Y, Y′, X′, respectively. Let us construct $p_{\xi\xi'}$ in terms of p_ξ, $p_{\eta|\xi}$, $p_{\eta'|\eta}$.

$$\begin{aligned} p_{\xi\eta\eta'\xi'}(\xi = x_i, \eta = y_j, \eta' = y'_k, \xi' = x'_l) = {} & p_{\xi'|\eta'\eta\xi}(\xi' = x'_l \mid \eta' = y'_k, \eta = y_j, \xi = x_i) \cdot \\ & p_{\eta'|\eta\xi}(\eta' = y'_k \mid \eta = y_j, \xi = x_i) \cdot p_{\eta|\xi}(\eta = y_j \mid \xi = x_i) \cdot p_\xi(\xi = x_i) \end{aligned} \tag{7}$$

where $x_i \in X$, $x_j \in Y$, $y'_k \in Y'$, $x'_l \in X'$. Using the fact that the sequence of random variables ξ, η, η', ξ' is Markovian, the joint distribution for the pair (ξ, ξ') is

$$p_{\xi\xi'}\left(\xi = x_i, \xi' = x'_l\right) = \sum_{j,k} p_{\xi'|\eta'}\left(\xi' = x'_l \mid \eta' = y'_k\right) \cdot p_{\eta'|\eta}\left(\eta' = y'_k \mid \eta = y_j\right) \cdot p_{\eta|\xi}\left(\eta = y_j \mid \xi = x_i\right) \cdot p_{\xi}\left(\xi = x_i\right) \cdot \quad (8)$$

Similarly, the joint distribution for the pair (η, η') is

$$p_{\eta\eta'}\left(\eta = y_j, \eta' = y'_k\right) = \sum_{i,l} p_{\xi'|\eta'}\left(\xi' = x'_l \mid \eta' = y'_k\right) \cdot p_{\eta'|\eta}\left(\eta' = y'_k \mid \eta = y_j\right) \cdot p_{\eta|\xi}\left(\eta = y_j \mid \xi = x_i\right) \cdot p_{\xi}\left(\xi = x_i\right) \cdot \quad (9)$$

Dobrushin (Reference d) explains that the conditional probability distribution $p_{\eta|\xi}\left(\eta = y_j \mid \xi = x_i\right)$ is the distribution of the signal transmitted, into which the message x_i is transformed as a result of the operation of encoding. Similarly, $p_{\xi'|\eta'}\left(\xi' = x'_l \mid \eta' = y'_k\right)$ is interpreted as the probability distribution of the message received, which arises from the signal y'_k as a result of the operation of decoding.

The Markovian requirement means that for a fixed input signal, the probability distribution of the channel output signal does not depend on what message was encoded into the input signal. For a fixed output signal of the channel, the probability distribution for the message into which this signal is decoded does not depend upon what the transmitted signal and the coded message actually were.

We are now in a position to calculate the quantity of information of one random object with respect to another. We have

$$I(\xi, \xi') = \sum_{i,l} p_{\xi\xi'}\left(\xi = x_i, \xi' = x'_l\right) log_2 \frac{p_{\xi\xi'}\left(\xi = x_i, \xi' = x'_l\right)}{p_{\xi}\left(\xi = x_i\right) p_{\xi'}\left(\xi' = x'_l\right)} \quad (10)$$

and

$$I(\eta, \eta') = \sum_{j,k} p_{\eta\eta'}\left(\eta = y_j, \eta' = y'_k\right) log_2 \frac{p_{\eta\eta'}\left(\eta = y_j, \eta' = y'_k\right)}{p_{\eta}\left(\eta = y_j\right) p_{\eta'}\left(\eta' = y'_k\right)} \quad (11)$$

where $p_{\xi}, p_{\xi'}, p_{\eta}, p_{\eta'}$ are the marginal distributions based on $p_{\xi\xi'}$ and $p_{\eta\eta'}$.

Following Shannon, we define the channel capacity as

$$C = \sup_{p_{\eta\eta'} \in V} I(\eta, \eta'). \quad (12)$$

The maximum is taken with respect to all possible choices of distributions $p_{\eta\eta'}$. The quantity that Shannon calls the "rate of creating information relative to a fidelity criterion" per unit time is defined as

$$H_W(\xi, \xi') = \inf_{p_{\xi\xi'} \in W} I(\xi, \xi'). \quad (13)$$

The minimum is taken with respect to all possible choices of distributions $p_{\xi\xi'}$. The necessary condition of the possibility of transmission is

$$H_W(\xi, \xi') \leq C. \quad (14)$$

If $\xi = \xi'$ then $H_W(\xi) = H_W(\xi, \xi')$ is the entropy of the random object ξ for the accuracy of reproduction *W*. In this case,

$$I(\xi, \xi) = -\sum_i p_\xi(\xi = x_i) log_2 \, p_\xi(\xi = x_i) \qquad (15)$$

as to be expected.

The set *W* consists of distributions $p_{\xi\xi'}$ of the pair of random variables (ξ, ξ') such that the M-dimensional vector of mathematical expectations

$$(\mathrm{M}\rho_1(\xi, \xi'), \mathrm{M}\rho_2(\xi, \xi'), \cdots, \mathrm{M}\rho_\mathrm{M}(\xi, \xi')) \in \overline{W} \, . \qquad (16)$$

$\overline{W}$ is an M-dimensional vector of numbers and the ρ_i's (i = 1, ⋯,M) are real-valued measurable functions with respect to $S_X \times S_{X'}$. The ρ_i's behave like distance functions. The distributions $p_{\xi\xi'} \in W$ satisfy the condition of accuracy of reproduction.

In the set $Y \times Y'$ we are given N real-valued measurable functions, $\pi_i(y, y'), \; (i = 1, \cdots, \mathrm{N})$. We also have an N-dimensional vector, $\overline{V}$, of real numbers. The set *V* consists of all those distributions $p_{\eta\eta'}$ of the pair of random variables (η, η') such that the N-dimensional vector of mathematical expectations

$$(\mathrm{M}\pi_1(\eta, \eta'), \mathrm{M}\pi_2(\eta, \eta'), \cdots, \mathrm{M}\pi_\mathrm{M}(\eta, \eta')) \in \overline{V} \, . \qquad (17)$$

The distributions $p_{\eta\eta'} \in V$ satisfy the restriction on the distribution of the signal. For example, the transmitting power has a specified upper bound.

It would be wonderful to find that random variable that would map the original message from the source directly to the user without any distortion. Unfortunately, one must take the indirect route, that is, by considering the encoding, channel and decoding portions of the problem. The mathematical formulation must include all these steps. The correct random variable, ξ', is the one that permits the fulfillment of the condition of accuracy of reproduction and of the restriction on the distribution of the signal.

Summary

We have used Shannon's theory of optimal coding of information to address the issue of relaying information from a source to a user, in the context of the OODA loop. The approach taken was that of Kolmogorov, which maximizes the practical applicability of the concepts. The salient feature is to find optimal encoding and decoding schemes that result in the fulfillment of the fidelity criterion and of the restriction of the distribution of the signal. We have restricted ourselves to the case of discrete random variables ranging over finite sets, to discrete memoryless sources and to discrete memoryless channels. Future reports will address the issues of rate distortion functions, continuous information, continuous channels and continuous sources.

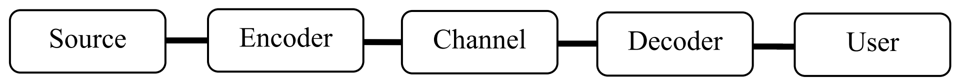

Figure 1. Diagram of a Communications System.

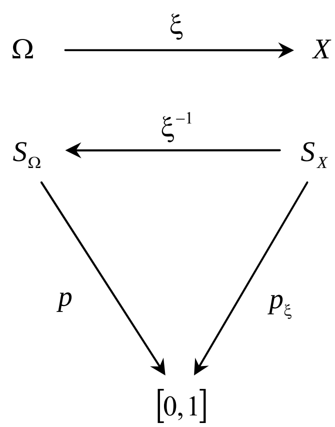

Figure 2. Measurable Function Diagram