REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
13-08-2013

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
15-05-2010 --- 14-05-2013

4. TITLE AND SUBTITLE
Information Fusion from the Point of View of Communication Theory

5a. CONTRACT NUMBER
FA9550-10-1-0231

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Gubner, John A.; Scharf, Louis L.; Chong, Edwin K. P., Fonseca; Benedito J. B. Jr.; Bunnag, Jittpat; Bhaskar, Badri N.; Hao, Kei; Yang, Luyu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Board of Regents of the University of Wisconsin System
21 North Park Street
Suite 6401
Madison, WI 53715 USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF
AFRL AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
USAF AFRL AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution A - Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Several aspects of communication theory have been investigated, with the main emphasis on multisensor systems. Among the fundamental results achieved, it has been shown when resolving multiple hypotheses at a desired probability of error, the number of hypotheses to be resolved can be traded off against available signal-to-noise ratio (SNR). It has been shown that the number of resolvable hypotheses can grow no faster than the SNR raised to a certain power. The research also showed that the error probability decreases exponentially fast in the number of sensors, and the error exponent was characterized in general and computed for several examples. In another multisensor setting with randomly deployed sensors, it was desired to detect the possible presence of a randomly located signal emitter. This results in a composite hypothesis-testing problem in which the measurements usually are conditionally dependent. However, it has been shown that if the region of interest is circular or a regular convex polygon, then there are simple least-favorable distributions for the emitter location that reduce the composite hypothesis to a simple one and result in the measurements becoming conditionally independent, resulting in a mathematically tractable problem.

15. SUBJECT TERMS
distributed systems, hypothesis resolution, information fusion, large deviations, least-favorable distribution, mis-assessment, ranc

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	11	John A. Gubner
					19b. TELEPHONE NUMBER *(include area code)* (608) 263-1471

Final Report

FA9550-10-1-0231

15-May-2010 through 14-May-2013

Information Fusion from the Point of View of Communication Theory

Principal Investigator: John A. Gubner
Department of Electrical and Computer Engineering
University of Wisconsin–Madison

Contents

1. Objectives

Objectives are little changed from the originally proposed objectives: derive algorithms for fusion and compression of measurements in distributed sensor arrays, with a view to trading off resolution, performance and signal-to-noise ratio.

2. Accomplishments

2.1. Information Fusion in Randomly Deployed Sensor Detection Systems

Problem Statement

This research focus considers the problem of fusing measurements from randomly placed sensors to make a binary decision as to whether or not a low-power signal emitter is present at some unknown location in a region of interest. We assume that the sensor measurements are to be combined at a fusion center. The fusion center's lack of knowledge of the emitter location (if present) makes the sensor measurements statistically dependent. Furthermore, the fusion center's lack of knowledge of the distribution of the emitter location makes the hypothesis-testing problem composite.

A traditional approach for dealing with composite hypotheses is to use a least-favorable distribution. However, most work in this area assumes statistically independent measurements, which we do not have.

Main Results

We have shown that for certain regions of interest, there are simple distributions for the emitter location that are "least favorable" for the fusion center to use. Furthermore, the use of such distributions renders the sensor measurements statistically independent, thus facilitating the design and analysis of decision rules and system performance.

Publication [1]

Our initial results [1] considered the situation in which the emitter is known only to lie in a disk S_e of known radius. We showed that any distribution for the emitter location that puts the emitter on the circular boundary of this disk is least favorable, and renders the sensor measurements statistically independent. These results assumed that the sensors were randomly deployed in S_e. We also considered the possibility that the sensors could be deployed in a larger disk S_s (with the same center as S_e). We showed that this can improve the worst-case performance.

Publication [2]

In order to quantify how conservative the use of least-favorable distributions is, in [2] we used the theory of asymptotic relative efficiency (ARE) to compare the performance of least-favorable distributions to most-favorable distributions. We showed that a distribution that places the emitter at the center of the disk S_e makes the sensor measurements statistically independent. Although the ARE depends on the radius of the disk, the ARE is bounded above by 9. We also investigated the ARE of the optimal detector under a least-favorable distribution with respect to a simple averaging detector under the same least-favorable distribution. We found that the ARE was one. Hence, the use of an averaging detector did not decrease the asymptotic performance (as the number of sensors tends to infinity).

Publication [7]

In [7], we began extending our initial results to noncircular regions. Specifically, we focused on regular convex polygons. We first showed that the least-favorable distribution for the emitter location placed the emitter only on the vertices. As in the circular case, such a distribution makes the sensor measurements statistically independent. We then used the ARE to compare regions with different numbers of sides. When the number of sensors is large, we showed that a system designer can decide between candidate regions of interest using a simple formula. When the size of the region of interest is large compared to the distance in which the signal approaches 0, we showed that regular convex polygons with a higher number of vertices would require a smaller number of sensors to reach a given asymptotic performance. For instance, if a system designer is to partition a large region into multiple regions of interest, each one monitored by a sensor detection system, then it is better to use hexagonal regions instead of square or triangular ones.

Publications [4] and [9]

Our most unified, complete, and general results can be found in [4] and [9]. As an example of a new result in [9], we used the ARE to show that for a large number of sensors and a large deployment region, there is very little performance difference between assuming the most-favorable distribution and deploying sensors in S_e and deploying in a slightly larger region S_s and using the least-favorable distribution.

Future Work and Applications

All of our results assume that the region of interest is either a disk or a regular convex polygon. Although the results can still be useful in some situations and as guidelines for a system design, it is important to consider further alternatives to deal with the problems associated with the statistical dependence of measurements and the composite hypothesis when the region of interest is not a disk or a regular convex polygon. Most of our results assume that the communication subsystem can offer a dedicated and error-free channel to each of the sensors. Since many sensor detection systems being envisioned by others consider non-dedicated or error-prone communication channels, it would be useful if our results could be extended to show that the least favorable emitter location distribution identified for the dedicated and error-free communication subsystem is still least favorable when considering either a communication subsystem offering dedicated but error-prone channels, or a communication subsystem offering multiple access channels.

Our results may also be extended to cover sensor detection systems using censoring. The reason we have worked with a composite hypothesis is due to the lack of knowledge about the emitter location distribution. The adoption of the least favorable emitter location distribution made the hypothesis simple. However, even when a distribution for the emitter location is adopted, we can still be left with a composite hypothesis if the signal level is also unknown. In this case, it would be useful to extend our results to the locally optimum case. We have also assumed that the fusion center decides upon a single set of measurements from the sensors. Sequential or change-point detection methods can also be used and it would be helpful to determine whether our least favorable emitter location distributions would still be least favorable when using these methods.

2.2. Information Fusion from Sensor Clusters

Problem Statement

In the previous section, we assumed that each sensor had a dedicated, error-free channel to the fusion center. In this section, we consider a situation in which sensors are placed in randomly located clusters to detect the potential presence of a signal emitter. Furthermore, the locations of sensors within a cluster are also random. If a sensor detects a signal (which may be a local false alarm), the sensor transmits a known waveform $v(t)$ to a common fusion center. Hence, the fusion center receives a superposition of delayed, attenuated copies of the waveform $v(t)$. How can we make statistical inferences based on the arrival times, and possibly the angles of arrival, of these waveform copies?

Main Results

Since the foregoing description of the waveform at the fusion center is similar to what would be received over a multipath channel, our initial investigation has focused on understanding such channels. Unfortunately, most work on multipath channels is based on simulation. Our work in [8] shows that a certain amount of analysis can actually be carried out. In [8] we develop a generalization of existing models and show how to derive closed-form and nearly closed-form expressions for quantities such as the excess-delay moments, the power-delay profile, and the power-delay-angle profile.

Future Work

Our work in [8] has improved our understanding of cluster point processes. We would like to leverage this knowledge to develop tractable models for communication and fusion of information from sensor clusters.

2.3. Optimization of Exponential Error Rates for a Suboptimum Fusion Rule in Wireless Sensor Networks

Problem Statement

Consider a parallel-architecture wireless sensor network used to make a decision about an M-ary hypothesis testing problem. In this system, the kth sensor (out of n) uses its measurement to generate an M-ary message U_k to be sent over its own channel to the fusion center, where U_k is decoded as L_k. In contrast to the common assumption that the data is conditionally independent *and identically distributed*, we assume only conditional independence under each hypothesis. This allows us to model situations in which different sensors have different local detection probabilities as well as situations in which different sensors have communication links of different qualities or signal-to-noise ratios (SNRs). We study the performance of a simple fusion rule that compares the numerical average of the decoded messages to a sequence of thresholds. In general, the joint optimization of multiple thresholds is a challenging problem.

Main Result

In [3] we prove a theorem that says we can select the thresholds independently in a manner that maximizes the asymptotic decay rate of the average probability of error. Furthermore, it is easy to compute these individual thresholds numerically. This is illustrated with an example.

Extensions

The results in [3] require knowledge of the moment generating functions of the channel outputs given the sensor outputs for each class of sensor/channel pair. We conjecture that by developing upper and lower bounds on these functions, we can obtain suboptimal thresholds that require less information about the sensors and channels. This could lead to results that are more useful in practical applications.

2.4. Resolving Hypotheses with Multiple Sensors: The Tradeoff between SNR, Resolution, and Probability of Error

It is commonplace to resolve questions at increasingly fine levels of resolution. But this raises the question of how finely hypotheses can be resolved when there is a signal-to-noise ratio (SNR) budget and a constraint that the probability of error not exceed a specified value P_e. We investigate the tradeoff between resolution level, SNR, and P_e in an n sensor array.

Problem Statement

Consider a sensor suite in which each measurement consists of a linearly transformed state variable plus additive, possibly non-Gaussian noise. We are interested in how finely the state variable can be resolved subject to a constraint on the probability of error. We reformulate this as a family of hypothesis-testing problems in which each family member contains a different number of hypotheses. We then investigate the tradeoffs among signal-to-noise ratio (SNR), desired error probability, and the number of hypotheses to be resolved.

Main Results

We give formulas that show how the SNR required to resolve M hypotheses at a specified probability of error scales with M. We give formulas that show how the number of resolvable hypotheses M at a given error probability scales with the available SNR.

Publication [6]

Our initial results were reported in [6]. This work considered only scalar state variables and scalar measurements. The exact formula for the error probability was not considered, rather only the high SNR/fine resolution regime was considered using a quadratic approximation of the large-deviation rate function.

The scaling results show that the required SNR grows with the quadratically in the number of hypotheses M, and the number of resolvable hypotheses grows with the square root of the SNR.

Publication [11]

Our latest results are reported in [11]. Here we allow the state to be p dimensional and the measurements to be d dimensional with $d \geq p$. Exact probability of error analysis reveals that the SNR needed to resolve M hypotheses grows proportionally to $M^{p/2}$, and M is bounded by a constant times $\text{SNR}^{p/2}$. This generalizes our $d = p = 1$ result in [6]. In addition to studying the error probability itself, we also do a large-deviation analysis to show that under fairly general conditions, the error probability tends to zero exponentially fast as the number of sensors increases. We characterize the error exponent, and we give examples where it can be computed in closed form. We also derive a d-dimensional quadratic approximation for the error exponent in the low SNR/fine resolution regime.

Applications of Results

Our results establish the tradeoff among SNR, probability of error, and resolution level for a large class of multi-sensor detection problems. To this extent, our results comprise design rules for the number of sensors, dimensionality of the sensor measurements, and the SNR required to resolve hypotheses about a state variable at a specified probability of error. The sensor suite may be a radar suite, a suite of hyper-spectral imagers, a suite of acoustic sensors, etc.

Extensions and Future Work

To this point, the linear map is given and known. However, we would like to consider other possibilities. For example, we could view the map as a precoding matrix to be designed. How should we do this for minimum probability of error? Alternatively, the linear map could be viewed as a channel matrix to be estimated. How would this impact the probability of error or the scaling laws?

2.5. Mutual Information-Based Sensor Suite Selection

Problem Statement

In many applications, a sensor suite can be used to accomplish different tasks; e.g., target detection followed by target tracking. The best sensor suite for detection may be different from the best one for target tracking. How can a sensor suite be selected so as to accomplish multiple tasks with a prescribed level of performance and cost?

Main Results

To compare sensor suite configurations, we must account for both performance and cost. The cost of the ith sensor in a sensor suite is taken as

$$C_{i,\text{inst}} + C_{i,\text{LTOC}},$$

where $C_{i,\text{inst}}$ is the cost of purchasing and installing sensor i, $C_{i,\text{LTOC}}$ is the lifetime operating cost of the ith sensor.

Because we envision a sensor suite being used for multiple tasks, we propose using a generic performance measure based on mutual information. Suppose that a particular sensor suite has n

sensors providing measurements $Y_1, \ldots, Y_n$ and X denotes the underlying system state. Then our measure of the sensor suite's performance is the average mutual information $I(X \wedge Y_1, \ldots, Y_n)$.

To combine the cost with the mutual information, we introduce a conversion factor F having units of dollars-per-bit. Our goal is to identify a sensor suite that maximizes the "net value"

$$F \times I(Y_1, \ldots, Y_n \wedge X) - \sum_{i=1}^{n} \{C_{i,\text{inst}} + C_{i,\text{LTOC}}\}.$$

In [5] we performed numerical experiments in which three possible sensor suites were indexed by a common parameter. For each parameter value, one of the three sensor suites has the optimal "net value," but which suite is optimal varies with the parameter value. The first suite consisted of a single sensor with measurement Y_1. The second suite consisted of a single sensor with measurement Y_2, and the third suite consisted of two sensors with measurements Y_1 and Y_2. Several scenarios were considered. We first considered the case when Y_1 and Y_2 were conditionally independent given the state and the conditional variance of Y_1 was fixed. The parameter that varied was the conditional variance of Y_2. The second case we considered fixed the conditional variances of both Y_1 and Y_2, and the parameter that was varied was the conditional correlation of Y_1 and Y_2 given the state. In both cases, we did some experiments in which the state variable was a discrete random variable and some experiments in which the state a continuous random variable (normal). In all scenarios, the measurements were conditionally jointly normal given the state.

The results in both the discrete and continuous scenarios show that in the fixed conditional variance case, using both sensors has better overall net value when the conditional correlation coefficient is close to 0 or 1. Specifically, when the two sensors are nearly uncorrelated, the combined information gain is bigger as there is not much redundancy in their individual information regarding the state of interest. On the other hand, if the conditional correlation coefficient is close to 1, the two sensors confirm each other, and therefore it helps by having an additional sensor. In the conditionally independent observation case, using both sensors is a better choice when sensor Y_2 has a high precision but not too high compared with Y_1 . When Y_2 loses precision, Y_1 is preferred because it was the less expensive one, and as Y_2 gets better, the information gain gets bigger, with the overall net value increasing. As the conditional variance of Y_2 decreases, using both is first preferred because if Y_2 is not too much better than Y_1 then the information of Y_1 is also helpful. And if the variance of Y_2 gets too small, then using only Y_2 is preferred because Y_1 can no longer help much.

2.6. Information Fusion and Communication Systems

Problem Statement

Consider a communication system whose transmitter has been built to send a message $X \in \mathbb{C}^N$ over a noisy, possibly fading channel. Now suppose there is an additional signal $c \in \mathbb{C}^N$ that is to be combined with X to form a new message $\widetilde{X}$. To use the same transmitter, the new message $\widetilde{X}$ must also lie in $\mathbb{C}^N$. In this problem, we must design a fusion rule to combine c and X into $\widetilde{X}$, and we must design a receiver to recover c and X.

A Special Case

To have a concrete setting in which to study our problem, we consider an orthogonal frequency division multiplexing (OFDM) system. In such a system, it is desired to send a signal

$$x(t) = \sum_{n=0}^{N-1} X[n]e^{j2\pi nt/T}, \quad 0 \le t \le T,$$

over a fading channel with frequency response $H(f)$. The channel output (neglecting receiver noise) is

$$y_0(t) = \sum_{n=0}^{N-1} H(n/T)X[n]e^{j2\pi nt/T}.$$

The receiver must both estimate the channel $H(f)$ and the messages $X[n]$, which belong to a finite signal constellation. The typical approach is to compute the inverse discrete Fourier transform (IDFT) of the samples $y_0(mT/N)$ for $m = 0,\dots,N-1$ to get

$$Y_0[n] = NH(n/T)X[n].$$

To estimate the channel, certain values of n are reserved (pilot tones) for which $X[n]$ is set to a fixed value known at the receiver. For such n, $H(n/T) = Y_0[n]/(NX[n])$. These samples can then be used to estimate $H(n/f)$ for other values of n. For these values of n, $X[n] = Y_0[n]/(NH(n/f))$.

Unfortunately, the signal $x(t)$ often exhibits a high peak-to-average power ratio (PAPR). This means that unless the transmitter amplifier has a large linear operating range, the signal will undergo distortion that changes its bandwidth and makes it difficult to decode at the receiver. One way around this is to transmit an alternative signal

$$\xi(t) = \sum_{n=0}^{N-1} c[n]X[n]e^{j2\pi nt/T},$$

where the coefficients $c[n]$ have unit magnitude and are chosen to reduce the PAPR. In other words, PAPR reduction suggests that we fuse c and X by computing their componentwise product. When $\xi(t)$ is sent over the channel, the receiver sees (again neglecting receiver noise)

$$y(t) = \sum_{n=0}^{N-1} H(n/T)c[n]X[n]e^{j2\pi nt/T}.$$

At the receiver, we compute the inverse discrete Fourier transform (IDFT) of the samples $y(mT/N)$ for $m = 0,\dots,N-1$ to get

$$Y[n] = NH(n/T)c[n]X[n]. \tag{1}$$

Again, certain tones n are reserved for which $c[n]$ and $X[n]$ have known values and the channel at those tones is estimated by $H(n/T) = Y[n]/(Nc[n]X[n])$. There are two problems with this approach. First, because $c[n]$ is constrained at the pilot tones, there is less flexibility to reduce the PAPR. Second, researchers typically assume the existence of a reliable side channel that is used to tell the receiver the sequence $c[0],\dots,c[N-1]$ that was used. For example, typically, the choices for the values of $c[n]$ are quite limited, and so one approach is to reserve additional pilot tones and use the formula $c[n] = Y[n]/(NH(n/T)X[n])$. The values of $c[n]$ for these additional pilot tones can serve to identify the entire vector c. The downside of this approach is that it further limits the number of tones available for message symbols.

Is there a way to avoid reserving extra pilot tones and to avoid sending knowledge of c to the receiver?

Main Results

Our method in [10] is based on looking at (1) and viewing the $c[n]$ as samples of a virtual channel $C(f)$ that is cascaded with the fading channel $H(f)$. We then use the original pilot tones to estimate the cascade $H(n/T)C(n/T) = H(n/T)c[n]$ at the non-pilot tones. Our initial work in [10] chooses the $c[n]$ to have magnitude one and piecewise-linear phase. We then estimate the messages by $X[n] = Y[n]/(NH(n/f)c[n])$. In this context, we do not estimate the $c[n]$, and we do not need a reliable side channel.

Future Work

Future work will include examining and improving existing channel estimators. We would also like to do time-domain simulation to take into account the effects of nonlinear amplifiers on the error probability. It would also be useful to extend our one-shot analysis to an iterative scheme over several symbol times.

3. Personnel Supported or Associated with this Research

- John A. Gubner, Professor of ECE and PI, University of Wisconsin–Madison (supported).
- Louis L. Scharf, Research Professor of Mathematics, Colorado State University, Fort Collins, CO (associated).
- Edwin K. P. Chong, Professor of ECE, Colorado State University, Fort Collins, CO (associated).
- Kei Hao, Honorary Fellow in ECE, University of Wisconsin–Madison (associated).
- Badri Narayan Bhaskar, ECE Ph.D. student, University of Wisconsin–Madison (associated).
- Jittapat Bunnag, ECE Ph.D. student, University of Wisconsin–Madison (supported).
- Benedito J. B. Fonseca Jr., ECE Ph.D. student, University of Wisconsin–Madison (supported).
- Luyu Yang, ECE M.S. student, University of Wisconsin–Madison (supported).

4. Interactions and Transitions

Participation and presentations at conferences and workshops are indicated in the Publications section below by having the presenting authors' names in bold. No transitions to report.

5. New Discoveries, Inventions, or Patent Disclosures

The new discoveries from this program are reviewed in the Accomplishments section above. There have been no inventions for which patents have been disclosed or filed.

6. Publications

The following publications were supported wholly or in part by AFOSR. Those publications that were presented at a conference or workshop have the presenting author's name in bold.

[1] **B. J. B. Fonseca Jr.** and J. A. Gubner, "Analysis of randomly deployed sensor detection systems under least favorable distributions," in *Proc. 48th Annual Allerton Conf.*, Allerton, IL, Sep. 29–Oct. 1, 2010, pp. 325–332.

[2] **B. J. B. Fonseca Jr.** and J. A. Gubner, "Least and most favorable distributions for the design of randomly deployed sensor detection systems," in *Proc. 14th Int. Conf. Inform. Fusion. (Fusion 2011)*, Chicago, IL, July 5–July 8, 2011.

[3] J. A. Gubner, **L. L. Scharf**, and E. K. P. Chong, "Optimization of exponential error rates for a suboptimum fusion rule in wireless sensor networks," in *Proc. Asilomar Conf. Signals, Syst., Comput*., Pacific Grove, CA, Nov. 6–9, 2011, pp. 1655–1658.

[4] B. J. B. Fonseca Jr., "On the use of least favorable distributions to facilitate the design of randomly deployed sensor detection systems." Ph.D. dissertation, University of Wisconsin–Madison, 2012. [Online]. Available: `http://depot.library.wisc.edu/repository/fedora/1711.dl:52JDV4PPQY3IL8V/datastreams/REF/content`

[5] L. Yang, "Particle filtering method for object tracking and sensor management using mutual information," M.S. thesis, University of Wisconsin–Madison, 2012. [Online]. Available: `http://minds.wisconsin.edu/bitstream/handle/1793/62125/thesis_Luyu%20Yang.pdf?sequence=1`

[6] **J. A. Gubner** and L. L. Scharf, "Resolving a variable number of hypotheses with multiple sensors," in *Proc. IEEE Statistical Processing Workshop (SSP 12)*, Ann Arbor, MI, Aug. 5–8, 2012, pp. 504–507.

[7] **B. J. B. Fonseca Jr.** and J. A. Gubner, "Least favorable distributions for the design of sensor detection systems in non-circular regions of interest," in *Proc. IEEE Statistical Processing Workshop (SSP 12)*, Ann Arbor, MI, Aug. 5–8, 2012, pp. 516–519.

[8] **J. A. Gubner**, B. N. Bhaskar, and K. Hao, "Multipath-cluster channel models," **invited paper** in *Proc. IEEE Int. Conf. Ultra-Wideband (ICUWB 2012)*, Syracuse, NY, Sep. 17–20, 2012, pp. 292–296.

[9] B. J. B. Fonseca Jr. and J. A. Gubner, "Least favorable distributions for the design of randomly deployed sensor detection systesm," submitted to *IEEE Trans. Inform. Theory*, 2012, revised 2013.

[10] J. Bunnag, "Minimizing peak-to-average power ratio in an orthogonal frequency division multiplex system," Ph.D. prelim, University of Wisconsin–Madison, 2013. Available from the PI upon request.

[11] J. A. Gubner and L. L. Scharf, "Resolving hypotheses with multiple sensors: The tradeoff between SNR, resolution, and probability of error," in preparation, 2013.