NOV 9 2000

SANDIA REPORT

SAND2000-2467
Unlimited Release
Printed October 2000

Approximate Public Key Authentication with Information Hiding

E. V. Thomas and T. J. Draelos

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/ordering.htm

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

RECEIVED
NOV 2 8 2000
OSTI

SAND2000-2467
Unlimited Release
Printed October 2000

Approximate Public Key Authentication with Information Hiding

E. V. Thomas
Statistics & Human Factors Department

T. J. Draelos
Cryptography and Information Systems Surety Department

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0449

Abstract

This paper describes a solution for the problem of authenticating the shapes of statistically variant gamma spectra while simultaneously concealing the shapes and magnitudes of the sensitive spectra. The shape of a spectrum is given by the relative magnitudes and positions of the individual spectral elements. Class-specific linear orthonormal transformations of the measured spectra are used to produce output that meet both the authentication and concealment requirements. For purposes of concealment, the n-dimensional gamma spectra are transformed into *n*-dimensional output spectra that are effectively indistinguishable from Gaussian white noise (independent of the class). In addition, the proposed transformations are such that statistical authentication metrics computed on the transformed spectra are identical to those computed on the original spectra.

Acknowledgments

The authors express appreciation to Amy Johnston, Department 9211, for her contributions in developing an attack against the PTP algorithm. The authors also thank Cheryl Beaver and Rich Schroeppel, both of Department 6234, and Ross Lippert, Department 9214, whose review and analysis of the PTP algorithm helped enhance the security and utility of the algorithm. Finally, the authors also thank John Cochran, a University of New Mexico graduate student, for developing software for the PTP algorithm.

Contents

Figures

Tables

Approximate Public Key Authentication with Information Hiding

1. Introduction

In many situations, it is desirable to authenticate data without revealing the data in detail. For instance, a party to a multilateral treaty might want to convince monitoring inspectors of the treaty that collected data represents a particular weapon type without revealing a detailed gamma spectrum of the weapon, which may in fact be classified. Even if the classified data can be kept private via an alternative representation, consistent correspondence between the classified and unclassified representations may reveal too much information about the weapon. As another example, the government may wish to utilize an information hiding mechanism to mitigate the concern of the private sector in providing proprietary information for national infrastructure protection.

Encryption alone cannot solve this problem. Additionally, certain data are prone to statistical variation, thus creating difficulties for consistent authentication results using standard digital authentication techniques. Gamma spectra are also examples of statistically variant data where measurements of the same sample with the same equipment will result in different spectra due to Poisson noise conditions.

Finally, public key cryptographic techniques are often useful in situations where one authenticating party seeks to convince multiple verifying parties or when the origination of data must be verifiable, thus providing non-repudiation. The purpose of this LDRD will be to investigate digital public key mechanisms that can be used to authenticate data prone to statistical variation and to investigate techniques for hiding data details while still proving the authenticity and integrity of the data.

1.1 Problem Description

Although the weapon inspection problem will be used as the primary application of this work, other applications exist with a similar problem set. For example, the use of biometrics to enable or authorize a certain function such as entrance into a building faces similar challenges. A biometric reading from the same individual using the same equipment will likely be slightly different each time. Moreover, the use of biometric information may have privacy implications that drive the need for hiding the detailed biometric information itself.

Generally speaking, any authentication process will have two steps. The first step is to initialize the authentication system by acquiring a reliable template of the item in question. In the weapon inspection application, this will be a representative weapon from the class of treaty-limited items. In the biometric application, initialization requires verification of the individual using information such as a birth certificate, driver's license, fingerprint, or DNA sample, and acquisition of the initial biometric. The initialization step requires that the representative item

(e.g., weapon or person) be certified to truly be a member of the said class. This generally requires additional off-line inspection processes that will not be discussed in this paper.

With the acquisition of an authentic template of monitored items, the system has been boot-strapped and can be used for subsequent inspections in the second part of the process. In the weapon authentication application, the basic problem is to make a class association as opposed to differentiating between individual weapons of the same class. In the biometric application, the original biometric is used as a template for subsequent authentication of the individual.

1.2 Related Work

An approach developed for use in biometric identification utilizes error correction coding techniques [DFM98]. It uses majority coding to construct a template of a biometric that is known to vary between measurements. Majority coding takes a number of measurements (preferably odd) and assigns each bit of the template to the value that is most often represented in the measurements using a majority rule. The template is then encoded into a code vector with a specified amount of redundancy. The amount of redundancy and the encoding technique used determines how many bits can be corrected in the template. In other words, if a vector does not perfectly match any codeword (template), then the closest codeword (in a Hamming sense) is generally assigned. The distance between codewords is representative of the number of correctable bits as well.

During verification, the same majority coding technique is used to acquire a biometric representative from a number of measurements. Since majority coding is a bit-oriented technique, the idea is to use it to acquire a representative test biometric. The hope is that it is within a specified Hamming distance of the original biometric template. If the representative is close enough to the template, it can be decoded into the exact biometric using bounded distance decoding.

A speech scrambling concept [SDM97] uses a data hiding technique that is very similar to the method described herein. The difference is that we constrain the input signal via scaling and centering prior to permuted transformation and they propose no authentication of the output signal. Scaling and centering of the input signal allows some very strong statements about the security of the algorithm to be proven. We show that independent of the permutation key and input spectrum, the output of the constrained permuted transform is consistent with a realization of Gaussian white noise. Hence, the distribution of the components of the output signal is substantially non-informative about the input signal.

1.3 Organization of the Paper

The remainder of this paper provides details of the approximate authentication with information hiding method referred to as the PTP algorithm, and illustrates how it might be applied to the weapon authentication problem. Section 2 provides an overview of the weapon authentication process and establishes the need for the PTP algorithm. Section 3 describes the PTP solution in detail, providing a numerical example of its use in gamma spectrum authentication. Section 4 discusses the efficacy of the proposed procedure for data hiding and provides some operational

guidelines for using the proposed method. Section 5 discusses its efficacy in preserving the ability to discriminate across classes. Section 6 provides a summary and gives information about the software available to utilize the PTP algorithm.

2. System Solution

The goals of a system solution for the problem of ensuring integrity of statistically variant data while maintaining privacy of the original data are twofold. First, the system must allow authentication (ideally, integrity, identification, and non-repudiation) of the source of the data. Secondly, the system solution must not reveal any usable information about the original source data.

The approach was developed by the LDRD team and involves the retention of information in the original signal in a statistical sense while provably hiding the original data. This approach is very non-invasive in terms of allowing users to utilize the same statistical authentication measures and evaluations on the measured signal that are used without any information hiding. The algorithm for hiding the original data involves permuting the original signal, applying a linear transformation, and then permuting the transformed signal. Through this process, the output signal becomes essentially indistinguishable from Gaussian white noise. The individual elements of the output signal or vector are completely uncorrelated.

2.1 Overview of the Weapon Authentication Process

In the weapon monitoring application, the host country is responsible for the inspected item while other parties to a particular treaty act as inspectors. The goal of weapon monitoring is to distinguish between different classes of weapons so that certain treaty-limited items can be tightly controlled.

In the two-step authentication process, the first step is to acquire a reliable template of the item under inspection. This step is performed only once for each weapon class and results in a template or reference signal. Step two of the authentication process involves acquiring a measurement of the inspected item for comparison to the reference signal. In the weapon authentication application, a statistically variant signal is recorded and measured in similarity to the prerecorded template. For example, the weapon's gamma spectrum must be statistically similar enough to a weapon-class template to be considered a member of the class. This process is shown in Figure 2.1.

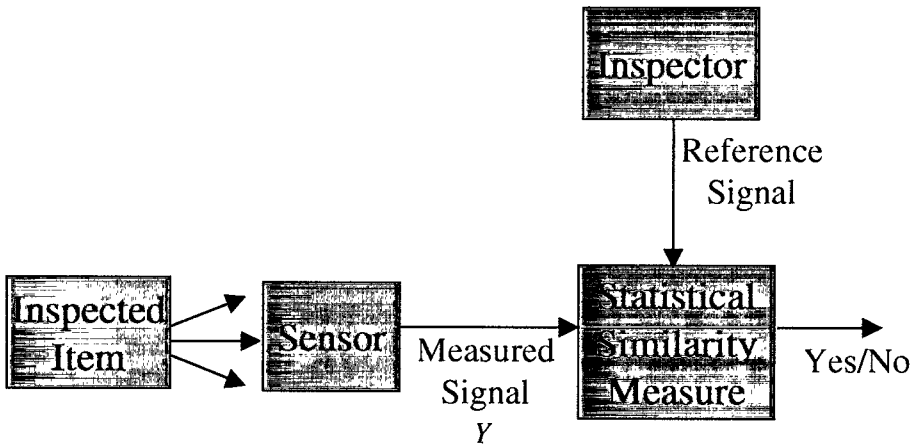

Figure 2.1 Authentication of a statistically variant signal

Given that the process shown in Figure 2.1 provides authentication of the inspected item, information hiding of the original data signal must be added to qualify as an acceptable, secure system solution. The system shown in Figure 2.2 provides information hiding of the original signal and outputs a signal that can be handled in much the same way as the original measured signal without information hiding. This is important because it may be possible to accommodate already familiar techniques for measurement of statistical similarity.

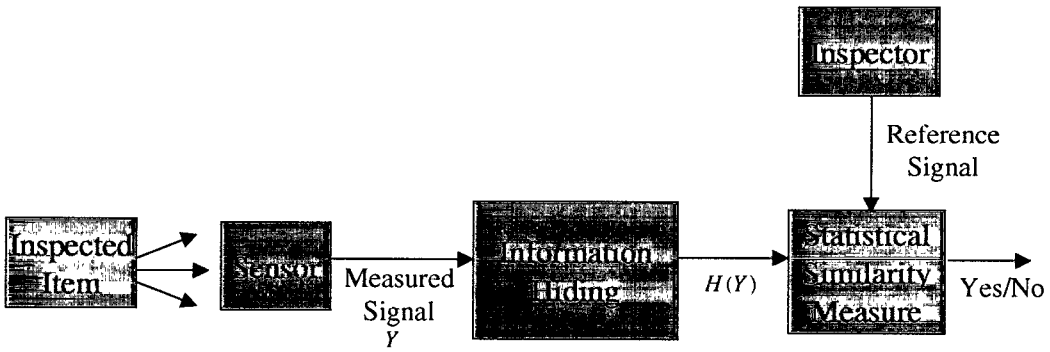

Figure 2.2 Authentication of a statistically variant signal with information hiding

3. The Permutation-Transformation-Permutation (PTP) Solution

The Permutation-Transformation-Permutation (PTP) solution to information hiding, as its name implies, performs 3 operations to the measured signal, after which the signal is completely unrecognizable from its original content. Figure 3.1 shows this system in block diagram form.

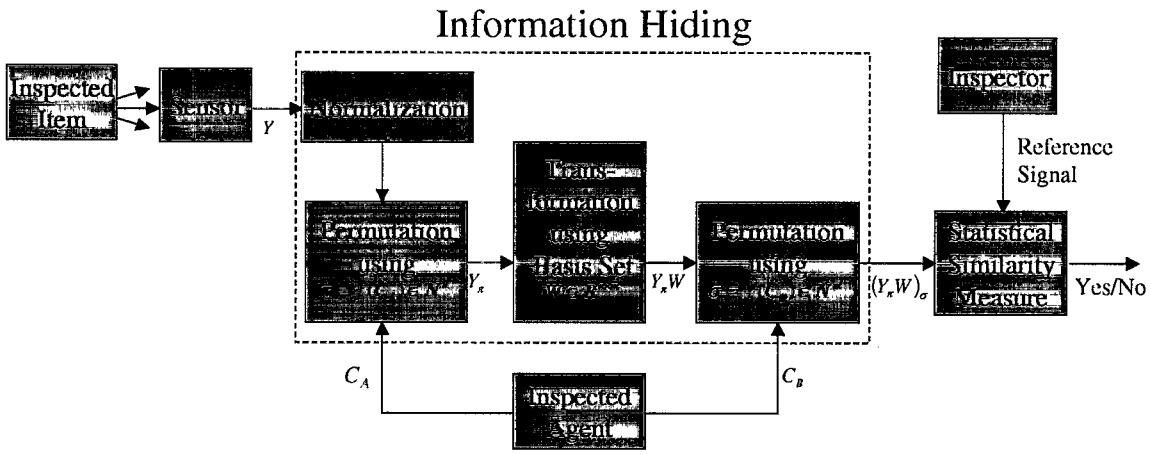

Figure 3.1 Authentication of a statistically variant signal with information hiding using the PTP solution

3.1 Permutation

The purpose of the permutation functions π and σ is to scramble the signal Y such that $(Y_\pi W)_\sigma$ can be made public without risk of revealing Y. Permutation operations are easily reversible if the permutation is known. Therefore π and σ must be kept secret. However, the root permutation function f() can be a standard secure hash function which need not be kept secret. Passcode C_A and C_B, uniquely associated with the Inspected Item, is entered by the Inspected Agent. Passcode C_A (C_B) is hashed and separated into n(m) equally sized pieces, each piece representing a numeric value. The pieces are ranked numerically and the ranking becomes the permutation π (σ).

For example, let a passcode of 1234 hash to the following 32-bit digest.

f (1234) = 9 f 2735a7

If the size of the permutation is 8, then each hex digit of f(1234) is ranked as follows to form the permutation. Note that ties can be handled in a predetermined or random manner.

f (1234)	=	9	f	2	7	3	5	a	7
Permutation	=	6	8	1	4	2	3	7	5

3.2 Transformation

The first task of the monitoring equipment is to produce a digital representation of the inspected item. This information, referred to as the Measured Signal (*Y*), is private and not to be released to or derivable by the Inspector. The Measured Signal is transformed via a linear transformation matrix, *W*, such that the new signal is *YW*. If the transformation matrix *W* were kept private as well, then this step could accomplish the complete information hiding solution. However, if *W* is

made public, then Y could be derived from YW. Therefore, permutations are applied before and after the transformation of Y.

3.3 Details of Permutation-Transformation-Permutation (PTP) Method

$$Y \rightarrow Y_{\pi} \rightarrow Y_{\pi} \cdot W \rightarrow (Y_{\pi} \cdot W)_{\sigma}$$

Definitions:

- Y is the n-dimensional row vector $\{Y_1, Y_2, \ldots, Y_n\}$ of measurements (typically comprising a spectrum).
- π is a *permutation* of the integers from $1{:}n$ that is unique to a particular verification class. A verification class consists of 1 or more physical units/items/individuals. For example, in the degenerate case, a verification class could be a single individual.
- W is an n x m transformation matrix with orthonormal columns that transforms the vector of measurements to $m \leq n$ latent variables.
- σ is a *permutation* of the integers from $1{:}m$ that is unique to a particular verification class.

Step 1: Center and scale-transform Y such that the mean of Y is 0 and the standard deviation of Y is 1. The scale transformation provides data normalization that renders the shape of the spectrum as being the sole identifying characteristic of a class.

Step 2: Permute the elements of Y : $Y \rightarrow Y_{\pi}$. The idea is to permute the elements of Y before applying the linear transformation (W) so that each latent variable is constructed/composed differently for each verification class. The elements of Y are randomly re-ordered.

Step 3: Linearly transform the permuted spectrum via W: $Y_{\pi} \rightarrow Y_{\pi} \cdot W$. The orthonormality of W implies that

$$\sum_{i=1}^{n} w_{ij}^2 = 1, \forall_j \tag{1}$$

Other characteristics of the columns of W are assumed as follows:

$$w_{i1} = \frac{1}{\sqrt{n}}, \forall_i \tag{2}$$

$$\sum_{i=1}^{n} w_{ij} = 0, \forall_{j>1} \tag{3}$$

For a given spectrum, we now have a sample space of $n!$ unique sets of equally likely *m-vectors* (latent variables) that comprise $T = Y_\pi \cdot W$. The particular realization of T that arises (at random) depends on the distribution of intensities within Y and the permutation π (not, however, on the natural ordering of Y [Step 1 took care of that]). Due to (2) and the fact that that the mean of Y is zero, $T_1 = 0$. In general, due to characteristics (1) and (3), we can claim that for the j^{th} element in T (T_j),

$$E(T_j) = 0 \,\forall_{j>1}$$

$$Var(T_j) = 1 \,\forall_{j>1}{}^{**}.$$

At this point, there is an association of the latent variable with its particular basis (a column of W). Depending on the size of m and the particular basis set that is used (e.g., $m = n$), T may be used to extract information about Y_π (e.g., via W^{-1}). From Y_π, one might obtain information about Y. Thus, one final step is needed to completely hide the original spectrum.

Step 4: Permute T: $T \rightarrow U = T_\sigma$ or $Y_\pi \cdot W \rightarrow (Y_\pi \cdot W)_\sigma$. Permute the latent variables. This step hides the association of a specific latent variable with a column of W. At this point we have broken the association between each column of W and its corresponding latent variable. For this step alone the sample space is of size $m!$. The whole process (Steps two to four) defines a sample space of *up to* $n! \times m!$ equally likely sets of permuted latent variables for each spectrum. The actual number of distinct sets of latent variables depends on W. The random permutation (σ) renders the distributions of the elements of U as *mutually indistinguishable* or *interchangeable*. Thus, over the class of possible permutations (π and σ) for a particular spectrum, the elements of U are identically distributed. Note that as an alternative to this second permutation, it has been suggested that the Ts be *sorted* rather than *permuted*. Admittedly, this will result in a simpler procedure. The problem is that such a procedure will allow many other input spectra to be incorrectly authenticated (many to 1 mapping) as only the *distribution* of Ts is authenticated (as opposed to the distribution and *order* of Ts).

3.3.1 Candidate Transformation Matrices (*W*)

The restrictions on W are that its columns and rows must be mutually orthonormal. In addition, we require

$$\sum_{i=1}^{n} w_{ij} = 0, \forall_{j>1} \text{ with } w_{i1} = \kappa, \forall_i .$$

There are a large number of candidates for W. Two possibilities for W are discussed in detail here.

** See [L75] for details.

3.3.1.1 Normalized Hadamard Matrix

A Hadamard matrix, H, has elements $H_{ij} \in \{-1,+1\}$. The rows and columns of H are orthogonal.

The order of a Hadamard matrix, n, is restricted to 1, 2, or $4n$ where $n \in Z^+$. Thus, for example, the dimension of the gamma spectra (n = 128) is compatible with this restriction. H_n has the property that

$$H_n^{-1} = \frac{1}{n} \cdot H_n^T .$$

For example, a Hadamard matrix of size 4 x 4 is

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

In order to obtain columns with length 1, we normalize

$$H\text{: } W = \frac{1}{\sqrt{n}} H_n .$$

In order to achieve our target statistical properties of the transformed spectra, we replace the condition that

$$\sum_{i=1}^{n} w_{ij} = 0, \forall_j$$

with $\overline{Y} = 0$. A lower bound for the number of *unique* Hadamard matrices of order n (when they exist for that order) is $n!$. This lower bound can be achieved by simply permuting the n *unique* columns (or rows) of H_n, where H_n is obtained via a standard construction in which all elements of the first column and row of H_n are +1 (or –1). In general, the number of unique Hadamard matrices of order N is unknown [HSS99]. Constructions for Hadamard matrices are discussed in [HSS99].

Note the uniqueness of the first column and row of the standardized Hadamard (all ones [or negative ones]). Thus, the first latent variable will be zero by construction.

For more information, see http://www.astro.virginia.edu/~eww6n/math/HadamardMatrix.html.

3.3.1.2 Fourier Coefficients – Cosine/Sine Basis

Assume that the spectrum is size n, where n is even (there is a similar development when n is odd). One possible basis set consists of

$$f_1(t) = \frac{1}{\sqrt{2}}$$

$$f_p(t) = \cos\left(\frac{2 \cdot \pi \cdot (p-1) \cdot t}{n}\right), \text{ for } p \in \{2, 3, ..., n/2\},$$

$$f_{n/2+1}(t) = \frac{1}{\sqrt{2}} \cdot \cos(\pi \cdot t), \text{ and}$$

$$f_{n/2+p}(t) = \sin\left(\frac{2 \cdot \pi \cdot (p-1) \cdot t}{n}\right), \text{ for } p \in \{2, 3, ..., n/2\},$$

all defined on $t = \{0, 1, 2, ..., n-1\}$. Let the j^{th} column of F be defined by $\{f_j(0), f_j(1), ..., f_j(n-1)\}$. The columns of F are orthogonal with

$$\sum_{i=1}^{n} f_{ij}^2 = \frac{n}{2}, \forall_j \text{ and } \sum_{i=1}^{n} f_{ij} = 0, \forall_{j>1}.$$

In order to obtain columns with length 1, we normalize

$$F\text{: } W = \frac{1}{\sqrt{\frac{n}{2}}} F_n.$$

As in the case of the Hadamard basis set, note that the elements of the first column are constant. Since, $\overline{Y} = 0$, one latent variable will be zero by construction.

3.4 Example - Weapon Monitoring Application

In the weapon monitoring application, the host country is responsible for the inspected item and plays the role of the inspecting agent. The other parties to a particular treaty act as inspectors. The goal of weapon monitoring is to distinguish between different classes of weapons so that certain treaty-limited items can be tightly controlled. Therefore, there is no need to distinguish between individual weapons.

The monitoring system is initialized once for each class of weapon. During this step, a single weapon representing the entire weapon class is inspected using out-of-band means to acquire trust in the monitoring system from this time forward. If the initialization weapon is not a trustworthy representative of the weapon class, then subsequent inspections cannot be trusted either. During initialization, a class-specific passcode is entered to form the secret permutations within the monitoring equipment. The same passcode must be entered at all subsequent inspections of the same weapon class and the passcode must be kept secret. After initialization, the monitoring equipment can erase the passcode, the permutations, and the classified data from the weapon measurement so that it no longer holds any secrets. The output of the initialization

process is an unclassified reference signal that the inspector can use for subsequent inspections of the weapon class.

During a routine inspection, the monitoring equipment acquires a measurement of the weapon, accepts a passcode from the host country, and outputs an unclassified inspection signal. The inspector can make a statistical similarity measurement between the inspection signal and the reference signal to arrive at an authentication result.

In the nuclear weapons verification area, gamma-ray spectroscopy can be used to uniquely identify weapon classes. The basis for this is the unique radio-isotopic/structural configuration of each weapon class. This gives rise to a characteristic gamma spectrum for each class. Spectra vary within a class due to manufacturing variation across units, the random nature of radioactive decay and measurement error. To illustrate the PTP method, 30 artificial gamma spectra were created. The spectra simulate the measurement (including Poisson counting errors) of 5 different gamma-emitting materials. Each material consists of a mixture of several radionuclides. Two counting times are assumed for each material. For each combination of material/counting time, there are 3 replicate spectra. The spectra, which have dimension $n = 128$, are displayed in Figure 3.2.

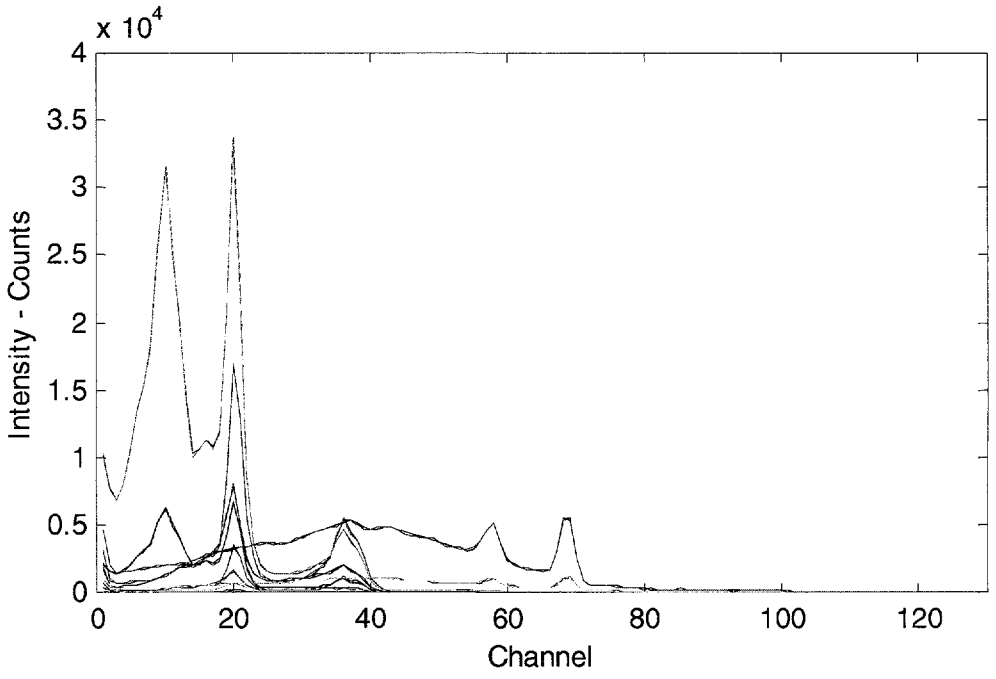

Figure 3.2 Gamma Spectra

The complete data-hiding mechanism in this example is as follows. First, a spectrum is square-root transformed on a pixel-by-pixel basis. (Here, the square-root transformation is variance-normalizing.) Next, the square-root transformed spectrum is centered (translated) and normalized such that its average value is 0 and standard deviation is 1. The resulting spectrum is then permuted (via a random permutation, π) and transformed via a normalized Hadamard matrix

$$(W = \frac{1}{\sqrt{n}} H_n).$$

The first latent variable is identically zero by construction, because the first column of W is a constant and the average spectrum is zero. Since there is no information in the first latent variable it is deleted. The remaining latent variables (m=127) are permuted via a random permutation, σ. Figure 3.3(a.–g.) illustrates the step-by-step metamorphosis of an individual spectrum from its original state (Y) to its final PTP-state (U). Notice that, as expected from theory, the public version of the spectrum (Figure 3.3g) is effectively indistinguishable from Gaussian white noise.

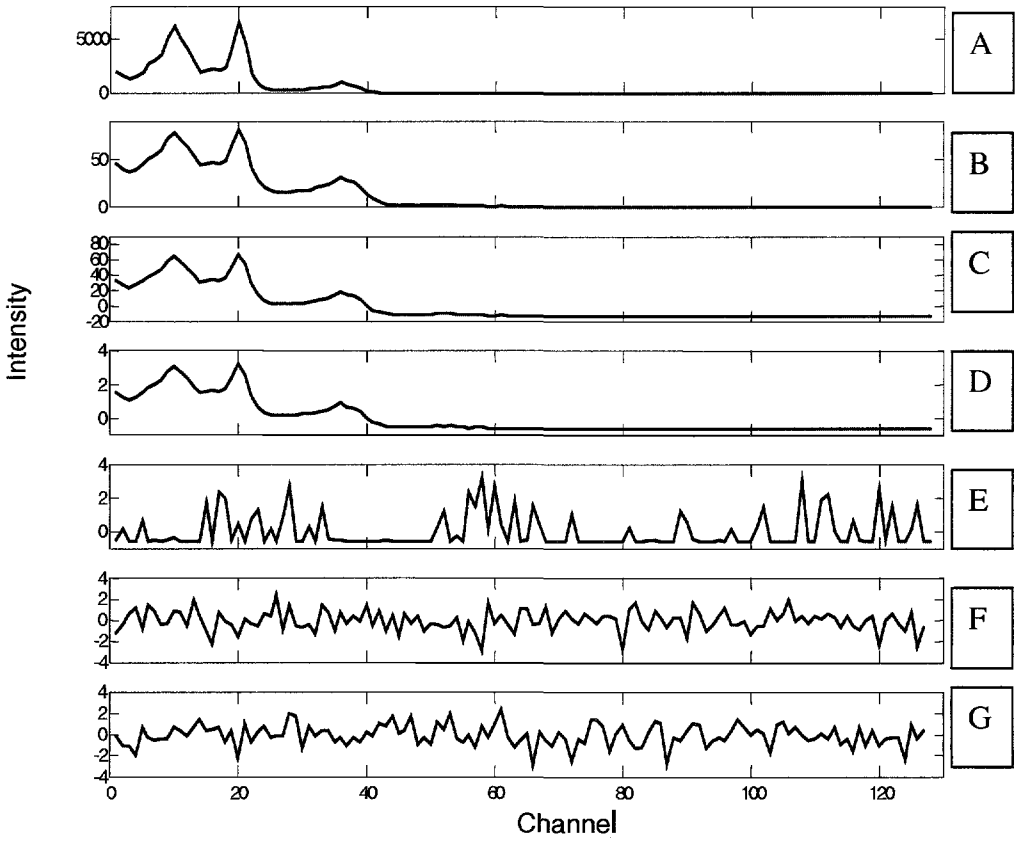

Figure 3.3 a-g Metamorphosis of Gamma Spectrum

A. Original Spectrum, B. Square Root Transformed Spectrum, C. Centered Spectrum (from B.), D. Scaled (Centered Spectrum), E. Permuted (Scaled/Centered Spectrum), F. Latent Variables Obtained Via Hadamard Transformation, G. Permuted Latent Variables.

Figure 3.4 displays 5 output spectra that are the result of varying the permutation set $\{\pi, \sigma\}$ when constructing the output spectrum. These radically different output spectra were constructed by applying different random permutation sets (and W) to the spectrum in Figure 3.3d.

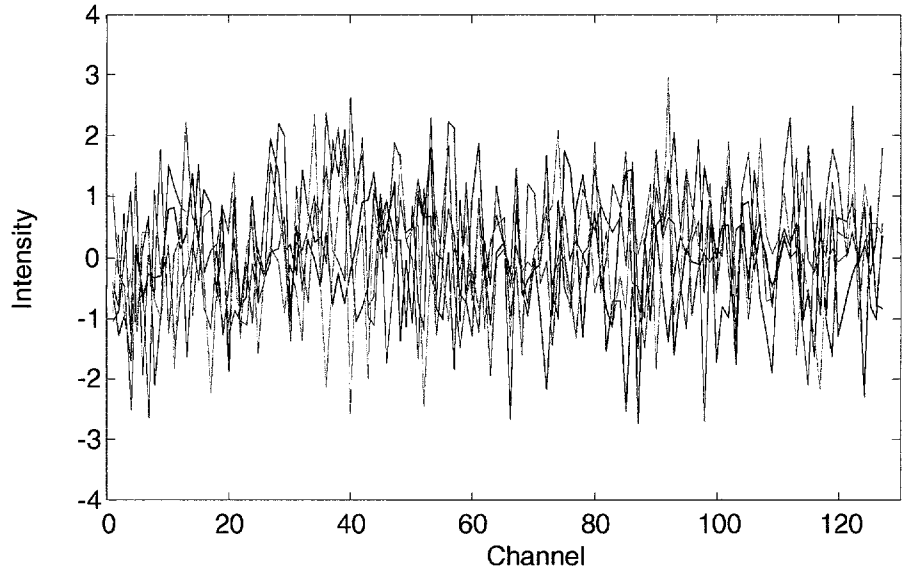

Figure 3.4 Multiple Output Spectra From Single Input Spectrum

Figure 3.5 illustrates the effect of applying different random permutations (in conjunction with the fixed *W*) to input spectra from different classes. A comparison of Figure 3.4 and Figure 3.5 shows that there is as much diversity within a class (using different permutation sets) as there is across classes. The spectra in both figures are essentially indistinguishable from independent Gaussian white noise processes.

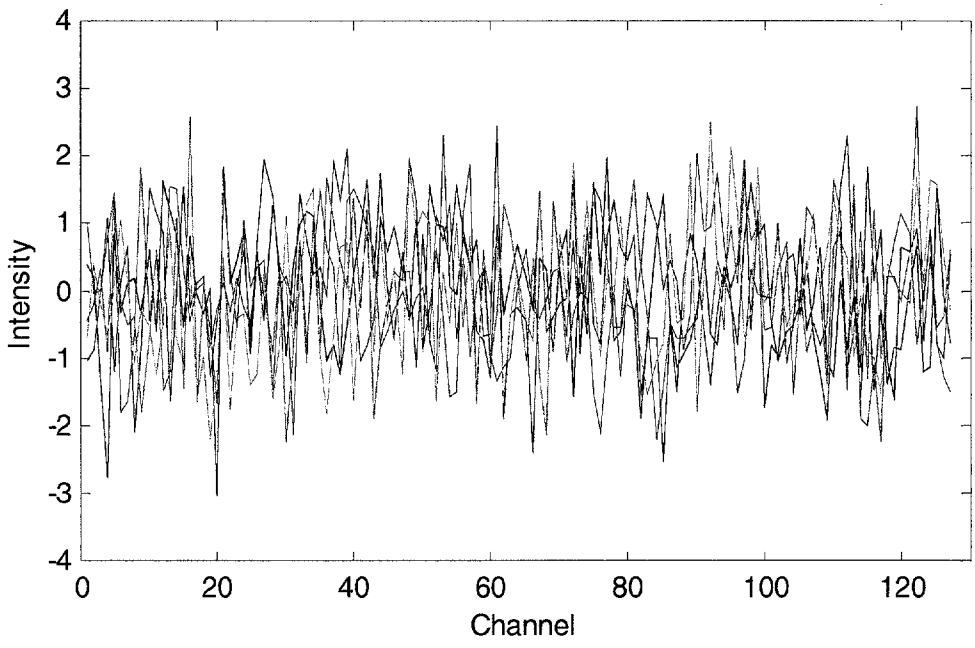

Figure 3.5 Output Spectra From Different Input Classes

Figure 3.6 illustrates the effect of applying the same transformation (W and $\{\pi, \sigma\}$) to the 2 sets of replicates of a single class. As is evident, there are only very minor differences across the spectra.

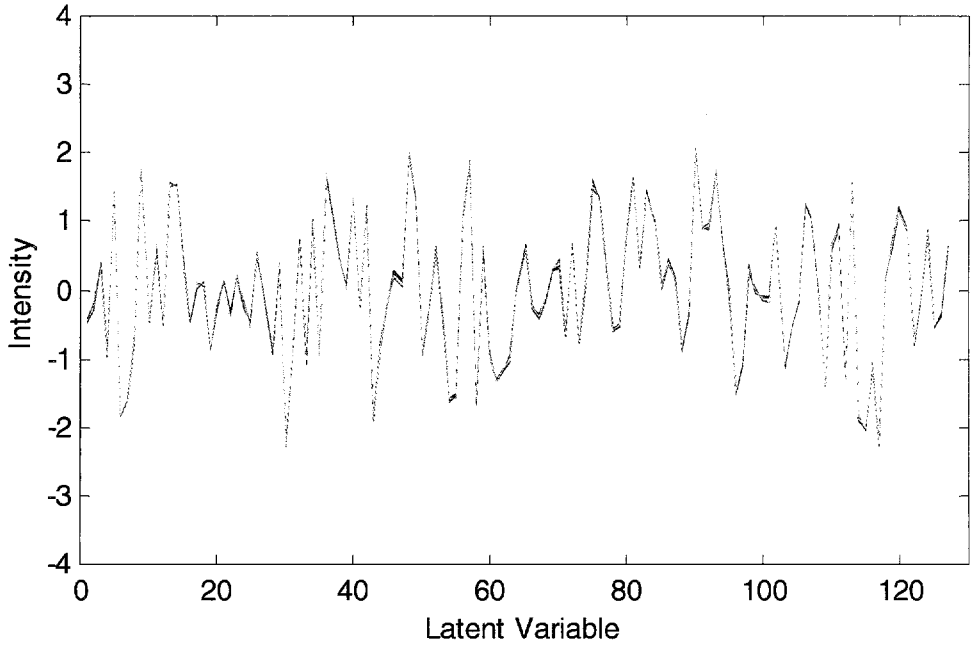

Figure 3.6 Output Spectra From Different Input Spectra of Same Class

3.5 Example - Biometric Application

In the biometric application, individual people represent the inspected agent and the inspected item is some biometric (e.g., fingerprint, retinal scan, hand geometry, etc.) of the inspected agent. The passcode is analogous to a personal identification number (PIN) that is entered by the individual or is read from a badge held by the individual. The inspector in this application is in control of both the monitoring equipment and the reference templates.

Initialization of the biometric monitoring system occurs once for each individual person. During this step, trust in the individual is acquired using out-of-band means. Once trust is established, the individual's biometric is measured, the PIN is acquired and a reference signal is computed for use during subsequent authentication of the individual.

When a PIN is entered by an individual or via a badge, the monitoring equipment retrieves the appropriate reference signal, measures the biometric and tests the "hidden" biometric with the reference signal as a test of authentication. The biometric is hidden so a collection of reference signals can be stored on a server while maintaining privacy of the associated biometrics.

3.6 Distribution of PTP Output Elements

This section provides a discussion of the distribution of the PTP output elements both within and across verification classes. This discussion is important because it provides a foundation for demonstrating the difficulties in distinguishing verification classes via the PTP output. The first part of the section deals with the limiting normal distribution of the PTP output as $n \to \infty$. The second part of this section illustrates that the normal distribution suggested via the asymptotic analysis is a very good approximation for $n \geq 100$.

First, assume a fixed input vector $\boldsymbol{Y}$, such that its elements have mean 0 and variance 1. Recall that $T_j = w_{1j}Y_{\pi_1} + w_{2j}Y_{\pi_2} + \ldots + w_{nj}Y_{\pi_n}$, where Y_{π_i} is the i^{th} element of Y_π and w_{ij} is the ij^{th} element of W. In cases where the first column of W is a constant (e.g., see Hadamard or cosine/sine constructions for W), $T_1 = 0$ since the elements of the first column of W are identical and $\overline{Y} = 0$. Thus, T_1 is not informative, so we really have only n-1 informative outputs. However, for

$$j > 1,\ \overline{w}_{\cdot j} = 0 \text{ and } \sum_{i=1}^{n}\left(w_{ij} - \overline{w}_{\cdot j}\right)^2 = 1.$$

Over the sample space populated by the permutation π on the single spectrum Y, $E(Y_{\pi_i}) = 0$ and $Var(Y_{\pi_i}) = 1$. First, fix j. Let

$$T_j^n = \sum_{k=1}^{n} X_k \text{ and } X_k = w_{\rho_k j}Y_{\pi_k}.\quad \boldsymbol{X} = \{X_k; k = 1:n\}$$

is a set of random variables derived from the sample space populated by the permutation $\boldsymbol{\pi}$ on $\boldsymbol{Y}$ and the permutation ρ on the j^{th} column of W. Note that the elements of $\boldsymbol{X}$ are exchangeable [KT81]. Note that

$$E(X_k) = 0 \text{ and } E(X_k^2) = \frac{1}{n}.$$

Suppose that $\chi_n = \{X_k,\ k=1:n\}$ defines a sequence indexed by n. Assume the *regularity conditions*

$$E|X_1|^3 < \infty,\ E(X_1 \cdot X_2) = o(n^{-2}),\ E(X_1^2 \cdot X_2^2) \to n^{-2}, \text{ and } E|X_1|^3 = o(n^{-1}).$$

Due to the exchangeability of the elements in X,

$$\boldsymbol{X},\ Z_j^n = \frac{T_j^n - E(T_j^n)}{\sqrt{Var(T_j^n)}}$$

converges to a Gaussian (0,1) random variable as $n \to \infty$ [CT97].

Furthermore, since $E(T_j^n) = 0$ and $Var(T_j^n) = 1 \forall_{j=1,\ldots,m}$, T_j^n converges to a Normal (0,1) random variable. So in the limit for a fixed spectrum, the T_j are identically and normally distributed over the sample space of the permutations of $\boldsymbol{Y}$. Furthermore, the T_j are uncorrelated since they are derived via an orthogonal basis ($\boldsymbol{W}$). Uncorrelated random variables that are Gaussian distributed are independent. Thus, the vector $\boldsymbol{T} = [T_1, T_2, \ldots, T_n]$ converges to Gaussian white noise.

With regard to satisfying the regularity conditions, one has to consider both the transformation matrix ($\boldsymbol{W}$) and the input data vector ($\boldsymbol{Y}$). For example, assume that $\boldsymbol{W}$ is a Hadamard matrix. In this case, for $j>1$:

1. $E_{i \neq k}(w_{ij} \cdot w_{kj}) = \dfrac{2}{n \cdot (n-1)} \cdot \sum_{i>k} \sum w_{ij} \cdot w_{kj} = -\dfrac{1}{n \cdot (n-1)}$ and

2. $E_{i \neq k}(w_{ij}^2 \cdot w_{kj}^2) = \dfrac{2}{n \cdot (n-1)} \cdot \sum_{i>k} \sum w_{ij}^2 \cdot w_{kj}^2 = n^{-2}$.

Thus, in this case, the regularity conditions

$$E(X_1 \cdot X_2) = o\left(n^{-2}\right) \text{ and } E(X_1^2 \cdot X_2^2) \to n^{-2}$$

imply that

1. $E(Y_1 \cdot Y_2) = o(1)$ and

2. $E(Y_1^2 \cdot Y_2^2) \to 1$.

The other regularity conditions

$$E\left|X_1\right|^3 < \infty \text{ and } E\left|X_1\right|^3 = o(n^{-1})$$

require a certain amount of variety in the input spectrum. For example, in the gamma spectrum example, the spectral mass should be well spread out and not concentrated at a few channels.

The above result holds for any input spectrum Y and transformation matrix W consistent with the above regularity conditions and the specifications for W. The result does not depend on the verification class. Thus, in the limit for such spectra, T converges *in distribution* to Gaussian white noise as $n \to \infty$. The noise realization depends on Y and the permutation, π. Thus, as $n \to \infty$, the *distributions* of T are indistinguishable across input spectra.

The second permutation, σ, simply permutes the elements of T and does not affect its distribution. Thus, the final PTP output (U) converges *almost surely* to Gaussian white noise as $n \to \infty$ so that the distributions of U are indistinguishable across input spectra within and across classes.

Next, we will demonstrate the distributional similarity of PTP output across permutations and a variety of simulated gamma spectra with n=128 channels. Figure 3.7 displays the PTP output from a single input spectrum with 5 different sets of permutations. The upper portion of Figure 3.7 displays the 5 realizations of output while the bottom portion of the figure displays the empirical cumulative distribution functions of the PTP elements associated with each realization. The standard Normal (mean=0 and standard deviation=1) cumulative distribution is superimposed in black for comparison.

Figure 3.8 displays summary of PTP outputs derived from 5 different input spectra, each associated with unique sets of permutations. A comparison of Figure 3.7 and Figure 3.8 shows that the distributions for a fixed input (but different permutations) differ by as much as the distributions associated with a variable input. At a macro-level the distributions of PTP elements are similar and are each indistinguishable from a standard Normal distribution.

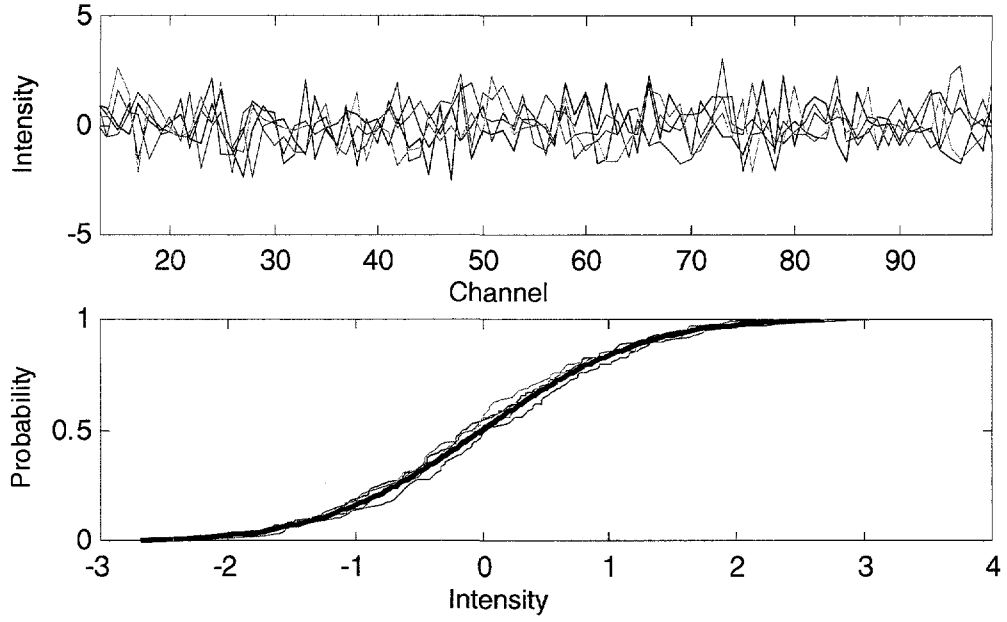

Figure 3.7 Summary of PTP Realizations Derived from Single Input Spectrum

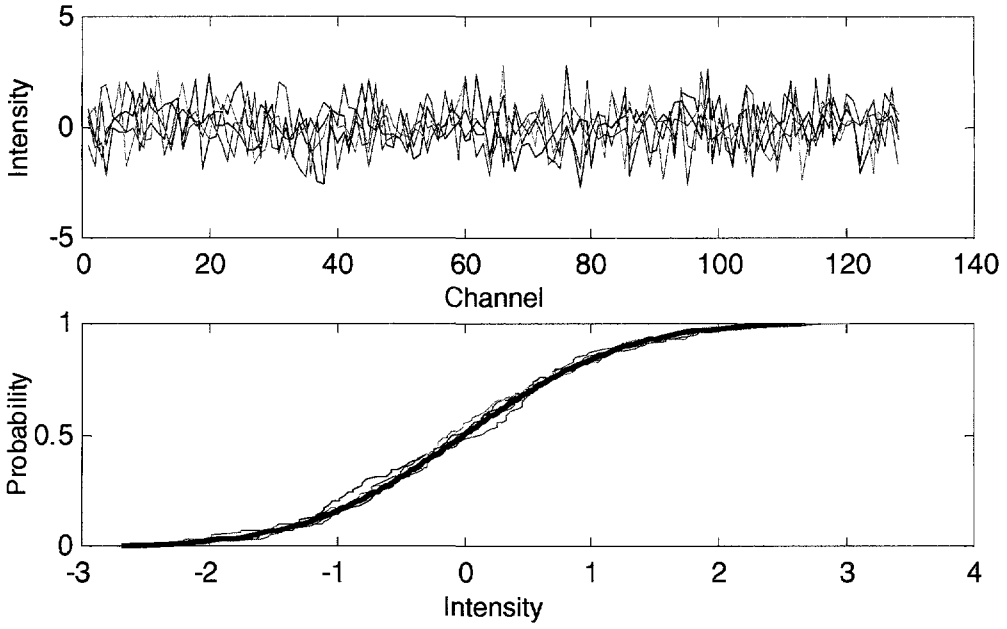

Figure 3.8 Summary of PTP Realizations Derived from Different Input Spectra, each spectrum with a unique permutation set

For a more formal assessment of normality, PTP output spectra were computed based on 1000 permutation sets for each of the 30 simulated gamma spectra from Section 3.4. This gives rise to 30000 PTP output spectra, each of dimension n=128. The Kolmogorov-Smirnov statistic,

$$D_n = \max\left|F_n(x) - F(x)\right|,$$

was computed for each output spectrum [DS86]. This statistic measures the maximum distance between the empirical c.d.f. of the elements of a specific PTP output spectrum ($F_n(x)$) and the c.d.f. of a target distribution ($F(x)$). Small values for D_n are indicative of a good match between $F_n(x)$ and $F(x)$. In this case, the target distribution is the standard normal distribution. The distribution of D_n was computed for each of the 30 simulated gamma spectra. Various percentiles of D_n are summarized below in Table 3.1. If the distribution giving rise to $F_{n=128}(x)$ is $F(x)$, then the expected values of the 50^{th}, 75^{th}, 90^{th}, 95^{th} , and 99^{th} percentiles of $D_{n=128}$ are 0734, .0902, .1078, .1202, and .1441 respectively. In general, the observed percentiles of $D_{n=128}$ are less than their expected values, indicating a better than expected match between $F_n(x)$ and $F(x)$. The exception to this is the 99^{th} percentile of $D_{n=128}$. This suggests that the difference between $F_n(x)$ and $F(x)$ is occasionally larger than would be expected if the distribution giving rise to $F_{n=128}(x)$ is $F(x)$. *Almost all* PTP outputs derived from the simulated gamma spectra are indistinguishable from *Gaussian white noise*.

Table 3.1 Distribution of $D_{n\bar{s}}$ by Input Spectrum

Spectrum	50th Percentile	75th Percentile	90th Percentile	95th Percentile	99th Percentile
1	0.0629	0.0763	0.0944	0.1076	0.1760
2	0.0644	0.0779	0.0934	0.1043	0.1358
3	0.0635	0.0779	0.0928	0.1046	0.1782
4	0.0632	0.0762	0.0918	0.1048	0.1931
5	0.0622	0.0757	0.0903	0.1034	0.1870
6	0.0630	0.0769	0.0936	0.1087	0.1806
7	0.0542	0.0649	0.0809	0.0962	0.2064
8	0.0542	0.0649	0.0810	0.0988	0.2210
9	0.0556	0.0665	0.0816	0.1009	0.2215
10	0.0555	0.0663	0.0851	0.1048	0.2882
11	0.0534	0.0640	0.0779	0.0936	0.2146
12	0.0536	0.0651	0.0810	0.0985	0.2862
13	0.0567	0.0705	0.0952	0.1396	0.1875
14	0.0571	0.0696	0.0897	0.1136	0.1912
15	0.0574	0.0711	0.0905	0.1222	0.1730
16	0.0565	0.0695	0.0890	0.1121	0.1969
17	0.0577	0.0710	0.0912	0.1136	0.1706
18	0.0568	0.0698	0.0903	0.1195	0.1813
19	0.0565	0.0667	0.0842	0.0986	0.1905
20	0.0578	0.0691	0.0883	0.1097	0.2620
21	0.0568	0.0689	0.0879	0.1091	0.1935
22	0.0573	0.0704	0.0860	0.0992	0.1640
23	0.0569	0.0695	0.0885	0.1059	0.2094
24	0.0569	0.0692	0.0837	0.0961	0.1662
25	0.0700	0.0826	0.0986	0.1086	0.1221
26	0.0695	0.0837	0.0980	0.1070	0.1217
27	0.0686	0.0825	0.0957	0.1027	0.1194
28	0.0689	0.0835	0.0977	0.1059	0.1200
29	0.0697	0.0826	0.0964	0.1054	0.1215
30	0.0686	0.0836	0.0983	0.1106	0.1245

In general, characteristics that influence the degree to which the PTP output elements resemble Gaussian white noise include the dimension of the spectrum n, and the distribution of intensities associated with the input spectrum, Y. For a fixed spectral shape, the PTP outputs tend towards Gaussian white noise as n increases. If the input spectral intensities are well distributed and not concentrated at a few pixels, the PTP output is likely to resemble Gaussian white noise. Such spectra are said to be *well behaved.*

So far we have focused on the distribution of PTP output elements in a *gross sense*. For finite n, in the case of simulated gamma spectra used as input, we have shown that it is difficult to distinguish the resulting PTP output from Gaussian white noise. However, at a *micro level*, the distribution of PTP output elements *does depend* on the input spectrum that is to be transformed.

For a fixed input spectrum over all possible permutations (π and σ), there is a finite set of values possible in the PTP output. In this regard, the PTP output based on a sine/cosine construction is superior to output derived from a Hadamard-constructed PTP because the finite set of values in the former case is larger than the set of possibilities derived from the later case. The more limited set of possibilities in the case of the Hadamard construction is due to the restricted set of coefficients available in a Hadamard matrix (+1 or –1). The set of possible PTP output values varies from spectrum to spectrum. However, the variation of spectra within a class (e.g., due to unit-to-unit variation and measurement error) broadens the set of possible output elements considerably within a class.

4. Efficacy of the PTP Method for Data Hiding

The efficacy of the PTP method for data hiding is described in 2 contexts. In the first context, we discuss the ability of the PTP method to hide Y given that an adversary has a *single* output spectrum from a given class. In the second context, we discuss some possible vulnerabilities of the PTP method when an adversary has *multiple* output spectra from the same class.

4.1 Single Output Spectrum

It is clear that the second random permutation (σ) destroys any structure in the PTP output (U) so that the *order* of the elements of an output spectrum provides no information regarding the characteristics of the input spectrum. We argue that all residual information about Y within U is localized to the distribution (as a whole) of values within U. Thus, the information-containing aspect of U is limited to the *distribution* of its elements.

However, assuming that the candidate spectra are well behaved, the distributions of the elements of the U are *approximately the same* regardless of the particular spectrum that is transformed. In the limit, as $n \rightarrow \infty$, the elements of U *are* independently and identically distributed according to a Normal distribution with zero mean and variance equal to 1 (*Gaussian white noise*). Thus, in this limiting case and without knowledge of π and σ, *a single* PTP-output is completely uninformative about the character of the input spectrum (Y). This is the fundamental basis for claiming that the PTP method is an effective data-hiding mechanism.

In summary, the PTP transformation makes the distribution (in a statistical sense) of possible output vectors from one class indistinguishable from the distribution of possible output vectors from another class. The PTP transformation produces output that is essentially indistinguishable from Gaussian white noise. For a particular input spectrum there is a huge number of possible realizations of this Gaussian white noise process (e.g., when W is a normalized Hadamard matrix obtained by a standard construction there at least $n!$ possible realizations). The uniqueness of the output for a particular class is provided by the combination of the input spectrum, the input permutation, and the output permutation.

Despite the evidence about the lack of information concerning Y via U, it is interesting to consider possible ways to attack the PTP scheme (i.e., gain information about Y). A brute force attack would be to invert U via all possible permutations in conjunction with W^{-1}. There are at

least 2 problems with this attack. First, the number of possible permutations could be enormous. It is believed that it would be computationally infeasible to compute all permutations for reasonably large n and m (e.g., $128! > 10^{200}$).

The second problem is more subtle. Suppose one were able to compute all possible versions of Y given U. Out of the huge number of possibilities for Y, we conjecture that just by chance there could be a very large number of feasible solutions. We conjecture that an attacker would have great difficulty in identifying the true value of Y from the feasible solutions. For example, Figure 4.1 displays the first 64 channels of a simulated gamma spectrum. Superimposed on this spectrum is a fixed line at 260 counts. Let S1 be the sub-spectrum from the pixel denoted by a '*' to the pixel denoted by a 'o'. Let S2 be the sub-spectrum from the pixel denoted by a 'o' to the pixel denoted by a 'x'. Let S3 be the sub-spectrum from the pixel denoted by a 'x' to the pixel denoted by a '+'. The count values of the 4 highlighted pixels are the same except for measurement error. The 3 sub-spectra (S1, S2, and S3) can be interchanged and reversed to form $2^3 \cdot 3!$ spectra that have an underlying *smooth nature* (hence possibly feasible) and would be equally likely solutions for Y given U that is derived from the true Y. Figure 4.2 displays one such result obtained by exchanging S1 and S3 and reversing S1. Note that the construction (e.g., isotopic composition) of systems giving rise to these spectra would be significantly different. There are numerous other ways to develop feasible spectra.

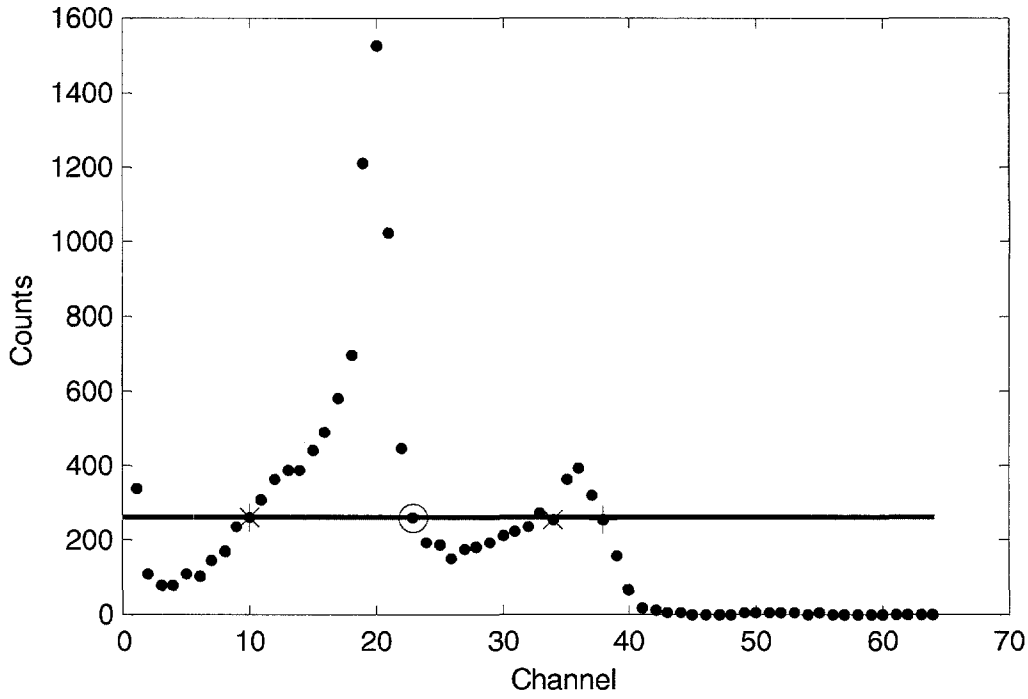

Figure 4.1 Simulated Gamma Spectrum

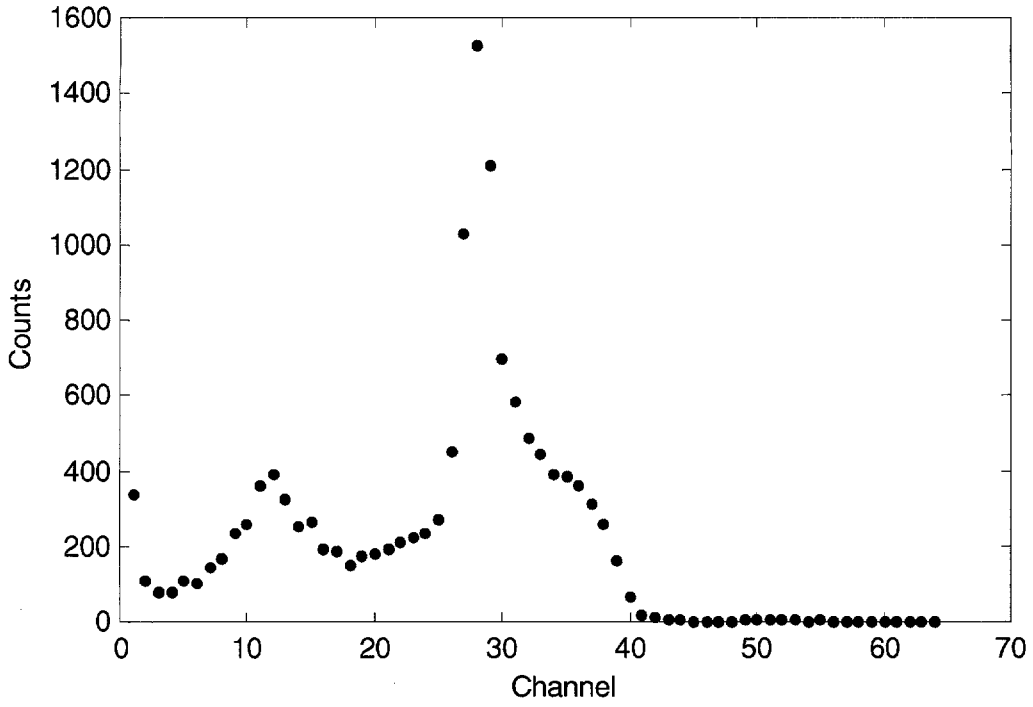

Figure 4.2 Feasible Solution

Continuing with the notion that one could compute all possible versions of Y given U, a proposed attack involves finding the "smoothest" version of Y and using that version as the solution. To formalize, suppose that we have been able to reduce the candidate solution set to X which is a permutation of the true input vector. Consider the *objective function*:

$$D = \sum_{i=1}^{n} (Y_{i+1} - Y_i)^2 ,$$

where $Y = X_\rho$ is a permutation of X. The permutation of X that minimizes D produces a monotone non-decreasing (or non-increasing) set of values for Y (e.g., $Y_i = X_{(i)}$). This can easily be proved by induction. The point is that the use of *smoothness* as a singular criterion for finding the true input vector may not be useful.

4.2 Multiple Output Spectra

Suppose that multiple output spectra from a single class are available to an adversary who is trying to obtain class-specific information about the input spectra. To formalize, consider the following 2 measurement error models.

Model 1: $Y = y + \delta$ and $U = u + \varepsilon$

1. Y is the input spectrum as measured (not accessible by host or adversary).

2. y is an idealized input spectrum that is perfectly repeatable within a class.

3. δ is the vector difference between the actual input spectrum and the idealized class-specific spectrum.

4. $U = (Y_\pi \cdot W)_\sigma$ is the public output of the PTP procedure applied to Y.

5. $u = (y_\pi \cdot W)_\sigma$ is the hypothetical output of the PTP procedure applied to y.

6. $\varepsilon = (\delta_\pi \cdot W)_\sigma$ is the hypothetical output of the PTP procedure applied to δ.

Note that the only *public observable* is U (the output spectrum).

For this discussion, assume that Y is centered and scaled such that $YY^T = 1$ and $\overline{Y} = 0$. Also assume that W is n by n (i.e. there is no dimension reduction). Other assumptions are made concerning the elements of δ, denoted by $\delta_\bullet$ where i=1, 2, ..., n. These assumptions are:

$$E(\delta_\bullet) = 0 \text{ and } Var(\delta_\bullet) = \sigma_\delta^2,$$

where the sample space spans all pixels of spectra within the same class. This broad sample space is considered due to the permutation, π. It follows that $E(\delta_\bullet \delta_\bullet^T) = n \cdot \sigma_\delta^2$. Note that σ_δ^2 could depend on the class and/or the measurement conditions.

In certain limits (e.g., well-behaved Y with large n), U is indistinguishable from Gaussian white noise (independent Gaussian elements with mean 0 and variance 1).

Via similar conditions on δ, one can argue that ε is indistinguishable from Gaussian white noise (elements with mean 0 and variance,

$$\sigma_\varepsilon^2 = \frac{1}{n} \cdot \delta\delta^T).$$

By difference, it follows that u is indistinguishable from a Gaussian white noise process with zero mean and variance, $1 - \sigma_\varepsilon^2$.

Multiple observations of U for a certain class might be used to estimate σ_ε^2 which can in turn be used to estimate σ_δ^2. For example a reasonable estimate of σ_ε^2 is

$$\hat{\sigma}_\varepsilon^2 = \sqrt{\frac{1}{(T-1) \cdot n} \cdot \sum_{t=1}^{T} \sum_{i=1}^{n} (U_{ti} - \overline{U}_{\cdot i})^2},$$

where T is the number of observations and $\overline{U}_{\cdot i}$ is the average value of the ith pixel over the T observations. Based on $\hat{\sigma}_\varepsilon^2$, a good estimate of σ_δ^2 is $\hat{\sigma}_\delta^2 = \hat{\sigma}_\varepsilon^2$. Thus, multiple observations of

U can provide information on Var(δ_i), which is a gross measure of the repeatability of Y across pixels and spectra.

In the case of the example involving gamma spectra, a relatively large value for $\hat{\sigma}_\delta^2$ might suggest that the underlying gamma spectrum (prior to scaling) has relatively poor signal-to-noise quality. For example, consider the 6 simulated gamma spectra shown in Figure 4.3. These spectra represent repeat measurements of the same material. Three of the spectra are associated with a simulated counting time of X, while the other 3 are associated with a counting time of 5X and thus have better signal-to-noise quality. These spectra (not square-root transformed) are centered and scaled prior to applying the PTP. (Also see Figure 4.4. Clearly the spectra associated with a counting time of X are noisier.) Following the PTP operation, we observe $U_{ti} - \overline{U}_{\cdot i}$ (for each set of 3 spectra. Figure 4.5 displays $U_{ti} - \overline{U}_{\cdot i}$ associated with each counting time case. In the case of relatively poor signal to noise (counting time is X), $\hat{\sigma}_\varepsilon = .1283$, whereas in the case of relatively good signal-to-noise, $\hat{\sigma}_\varepsilon = .0621$.

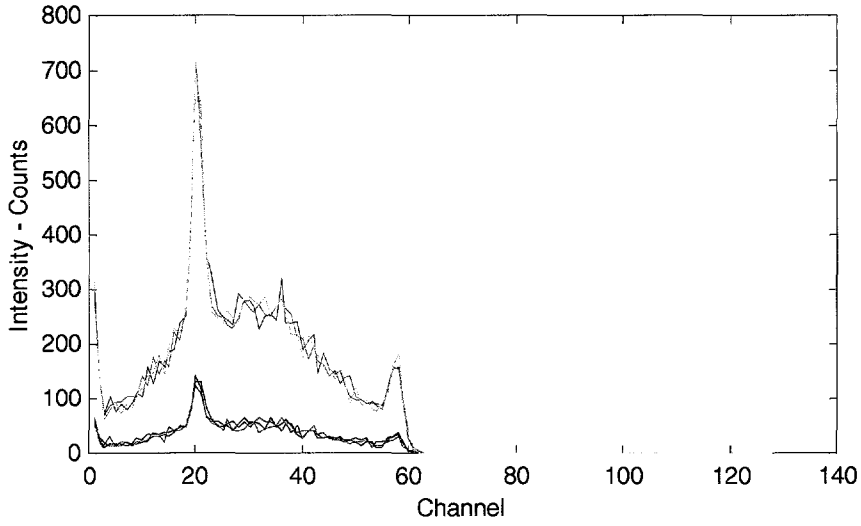

Figure 4.3 Simulated Spectra (Counting times of X and 5X)

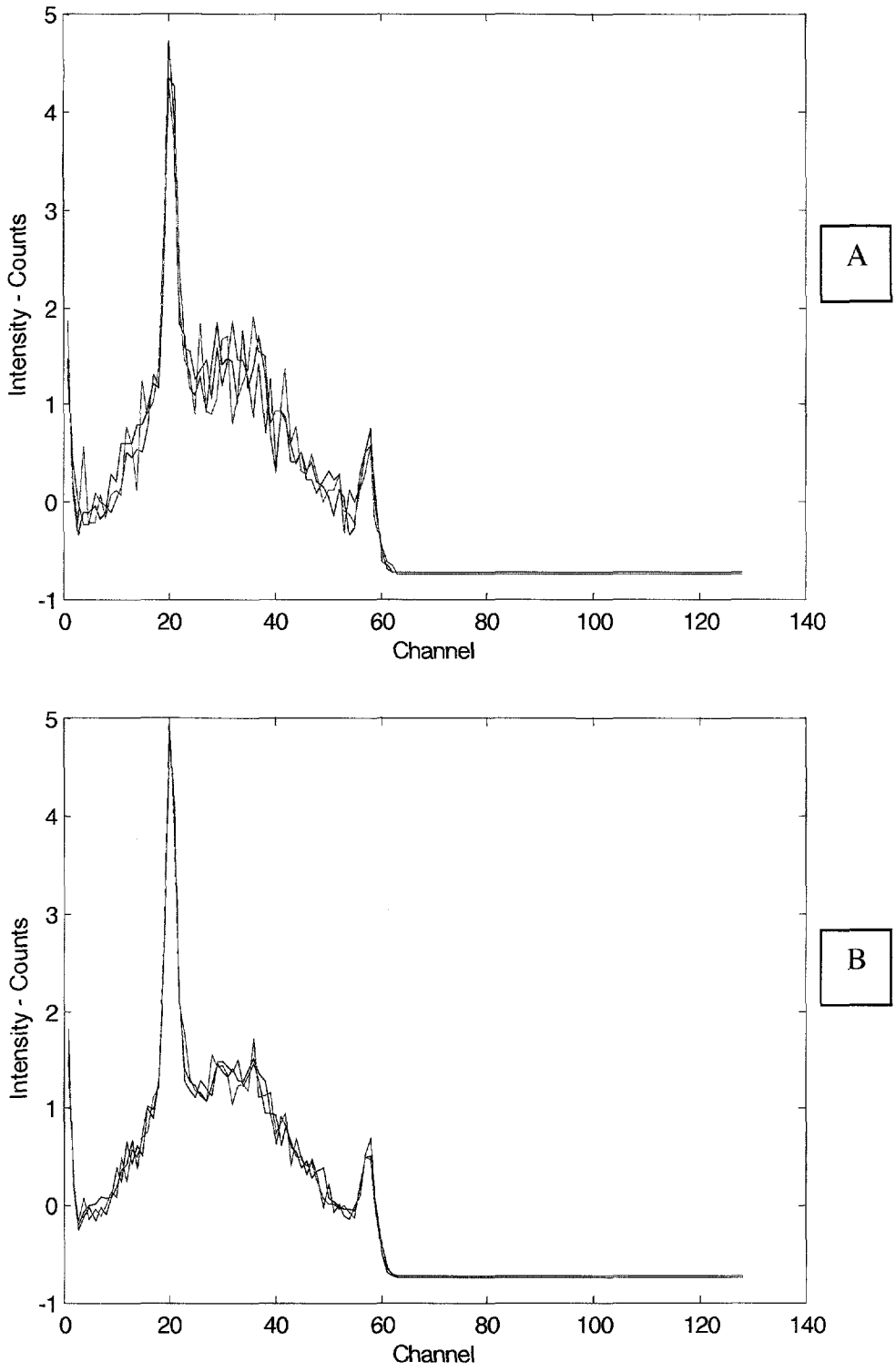

Figure 4.4 Normalized Spectra

A. With X Counting Time, B. With 5X Counting Time.

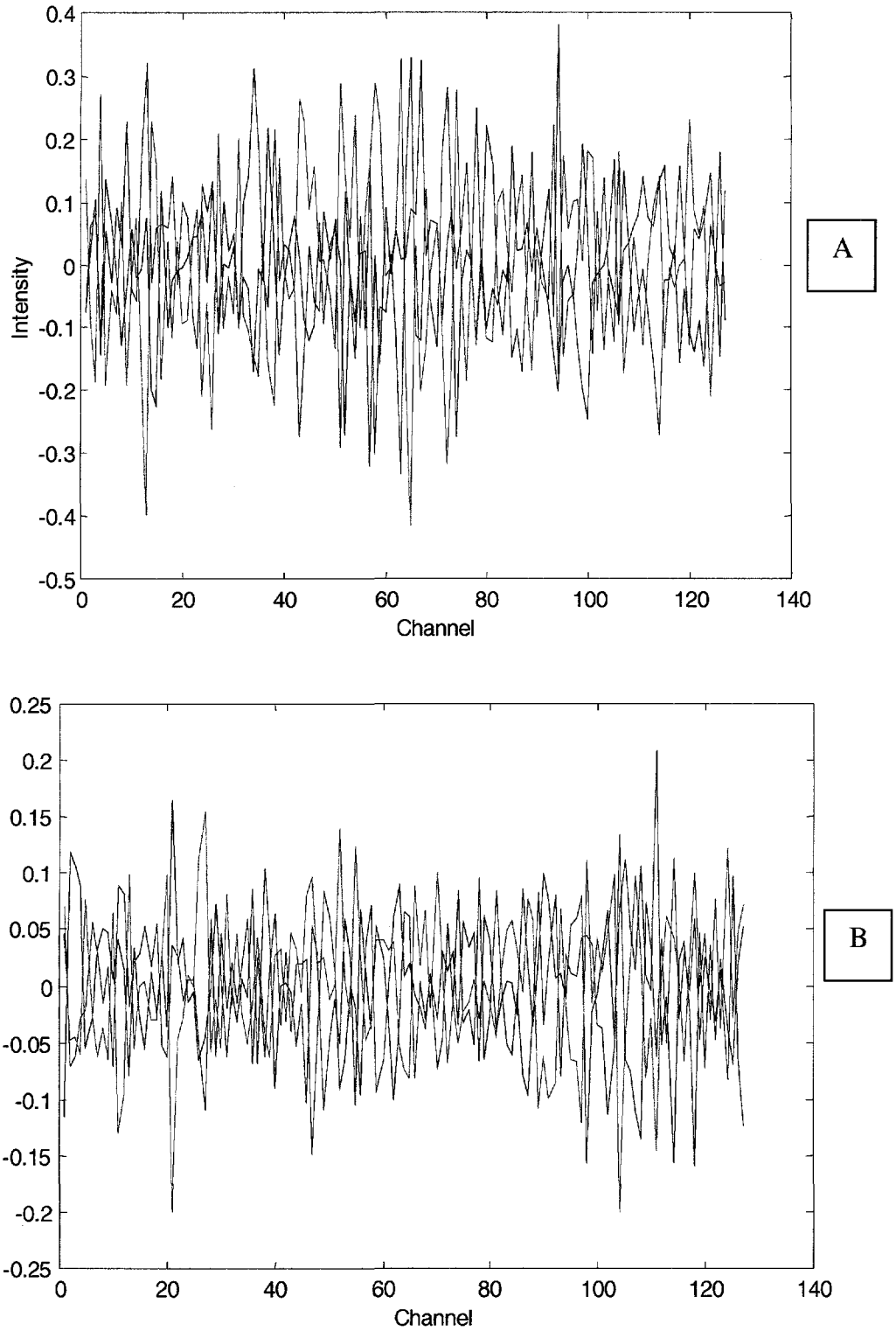

Figure 4.5 $U_{ti} - \overline{U}_{\cdot i}$

A. With X Counting Time, B. With 5X Counting Time.

Here we have used simulated measurements of the same class at different counting times to vary signal-to-noise quality. From repeated observations of U we were able to ascertain the relative signal-to-noise of the inputs. In practice, an adversary could analyze the repeated outputs from a fixed class to ascertain the relative signal-to-noise in the inputs. From that, an adversary might be able to deduce something about the magnitude of the underlying class-specific signal y (e.g., yy^{T}) if there is a relationship between σ_δ^2 and yy^{T}. However, even if ε was available to the adversary, nothing about the shape of Y (or y) would be revealed, since mapping ε to δ is as difficult as mapping U to Y.

So far, we have assumed that δ is unknown to the adversary. However, if δ is known or if an adversary had the ability to adjust Y (e.g., by adding a known perturbation) and observe the corresponding change in U, then the adversary could determine π and σ even without knowledge of Y. Potential adversaries must not have the ability to affect Y. One way to defend against the adversary would be to vary the signal to noise in the original input spectra (prior to the normalization that results in $YY^T = 1$) from measurement to measurement. For example, one could accomplish this by varying the counting time. An assumption here is that the original input noise is only small part of the original total input signal. If that assumption holds, the magnitude of the resulting idealized spectrum is, after normalization, relatively unaffected.

Model 2: $Y = y + \delta$ and $U = u + \varepsilon$, the same as Model 1 except that the individual elements in δ have different variances. This attack assumes that the measurement-error variances of the input channels are unique. The basic strategy of the attack is to do an eigenanalysis of a set of output spectra (U_1, U_2, ..., U_N) from a single class. *If certain conditions are satisfied*, then each sign-transformed eigenvector is equivalent to a row of $\pm 1 \cdot_\pi W_\sigma$. Given that there are n rows, one can guarantee that a row permutation of the matrix $_\pi W_\sigma$ exists within the set of 2^n possible candidates (the set of possible candidates is due to the sign ambiguities of the eigenvectors). The search space is thus reduced from more than $n!$ candidates to one with 2^n candidates. The search space can be further reduced greatly by utilizing the known form of the Hadamard matrix. Assume that the row permutation of the matrix $_\pi W_\sigma$ (say $V =_{\omega(\pi)} W_\sigma$) can be identified from the possible candidates. Now given a specific output spectrum (U), V can be used in conjunction with U to obtain a permuted version of the associated input spectrum (Y). That is $Y_{\omega(\pi)} = U \cdot V^T$. Thus, this attack can produce a permuted version of the input spectrum.

The following MATLAB code provides an example of how this might be done.

```
% This attack obtains ROWS of the transformation matrix
% It uses many repeat observations (n=100)
% Start with Hadamard matrix (forget about permutations for
now)
w=hadamard(8);
% Create input data (actually perturbations) such that each
input variable has a % unique standard deviation, sd
sd=[100 1000 10000 10 1 .1 .01 .001];
x=randn(100,8)*diag(sd);
%Transform to output (y)
```

```
y=x*w;
my=y-ones(100,1)*mean(y);
% Do eigen-analysis of y
[c,n]=size(y);
ay=ones(1,c)*y/c;
[u,s,nev]=svd(y-ones(c,1)*ay,0);
nev=nev(:,1:8);
score=y*nev;
ev=sign(nev);
w =
     1     1     1     1     1     1     1     1
     1    -1     1    -1     1    -1     1    -1
     1     1    -1    -1     1     1    -1    -1
     1    -1    -1     1     1    -1    -1     1
     1     1     1     1    -1    -1    -1    -1
     1    -1     1    -1    -1     1    -1     1
     1     1    -1    -1    -1    -1     1     1
     1    -1    -1     1    -1     1     1    -1
ev =
     1    -1     1    -1     1    -1     1    -1
     1     1     1     1     1     1     1     1
    -1    -1     1     1     1    -1    -1     1
    -1     1     1    -1     1     1    -1    -1
     1    -1     1    -1    -1     1    -1     1
     1     1     1     1    -1    -1    -1    -1
    -1    -1     1     1    -1     1     1    -1
    -1     1     1    -1    -1    -1     1     1
```

Except for a sign (e.g., columns 2, 4, 6, and 8), the original columns of *EV* form the rows of *W*. Here, due to the wide disparity in variances and relatively small dimension (n=8), we needed relatively few observations to get the rows of *W*. Although this attack does reveal the rows of *W*, it does not (in general) reveal their ordering. Also note that the columns of *EV* are the rows of *W* sorted by the standard deviation of the errors in the input variables (sd = [100 1000 10000 10 1 .1 .01 .001]). For example, the first column of *EV* is the third row of *W* corresponding to the location of the maximum element in sd (10000). The second column of *EV* is the second row of *W* corresponding to the location of the second largest element in sd (1000).

The rough argument behind this attack is that the uniqueness of the measurement-error variances of the input channels induces a correlation structure in the output channels. To illustrate, represent the k^{th} replicate (within a class) of the j^{th} ouput channel as $U_{jk} = w_{j1}Y_{1k} + w_{j2}Y_{2k} + \ldots + w_{jn}Y_{nk}$, where the $w_{..}$ terms are the values in the Hadamard-constructed

$$W(\pm\frac{1}{\sqrt{n}}),$$

the $Y_{\cdot}$ terms are the values associated with the input channels, and $k = 1,2,\ldots,N$. Now, consider a degenerate case where the first input channel ($Y_{1\cdot}$) is the only input channel that exhibits variability over the set of replicates ($k = 1,2,\ldots,N$). Thus, observed variation in the output channels U_{jk} is due solely to the variation in ($Y_{1\cdot}$). The result is perfect positive correlation between output channels j and j' if $w_{j1} = w_{j'1}$ (concordant channels) and perfect negative correlation between these output channels if $w_{j1} = -w_{j'1}$ (discordant channels). In the general case, an eigen-analysis of the sample covariance matrix of the output channels (n by n) could be used to extract the rows of W. The sample covariance matrix of the output channels is decomposed into n n-dimensional eigenvectors. In our example this analysis was accomplished via a singular-value decomposition of the mean-centered output spectra. A sample size (N) that is suitable for the degree of variance similarity across input channels is required (more similarity in variances requires a larger sample size). An element-by-element *sign transform* of each eigenvector will result in a vector that is equal to a column of W (to within a sign). It does not seem possible to resolve the sign ambiguity.

Note that the difficulty in implementing this attack increases as the channel variances become more homogeneous, the sample size (N) decreases, and the dimension of the spectrum (n) increases.

In order to characterize the efficacy of this attack over a broad range of conditions, the following limited study was conducted. The degree of *variance similarity* was controlled by the function $V_i = (1+f)^i, i = 1,2,\ldots,n$, where V_i is the measurement error variance associated with the i^{th} input channel. For this study, the assumed distribution of the measurement errors is Normal. Larger values of f impose greater diversity in variance and hence make the attack easier. When $f = 0$, this attack will not work even for an arbitrarily large replicate sample size. Values of n that were considered are in the set {8, 16, 32, 64, 128}. Replicate sample sizes considered (N) are in the set {10, 100, 1000, 10000}. For each value of $f \in \{.025, .05, .1, .2, .5, 1\}$ and n we identified (in a rough sense) the minimal sample size that would allow a successful attack (Table 4.1). See the sample MATLAB code below.

```
n=128;
f=1
sampsize=1000;
w=hadamard(n);
% Create input data such that each input variable has a
unique variance
dd=(1+f).^(1:n);
x=randn(sampsize,n)*diag(dd);
%Transform to output (y)
y=x*w;
% Do eigen-analysis of y
ay=ones(1,sampsize)*y/sampsize;
[u,s,nev]=svd(y-ones(sampsize,1)*ay,0);
nev=nev(:,1:n);
ev=sign(nev);
```

Table 4.1 Sample Size Required* for Successful Attack

	f = .025	f = .05	f = .10	f = .20	f = .50	f = 1.0
n = 8	10,000	10,000	1,000	1,000	100	100
n = 16	10,000	10,000	10,000	1,000	100	100
n = 32	>10,000	10,000	10,000	1,000	100	100
n = 64	>10,000	10,000	10,000	1,000	100	100
n = 128	>10,000	10,000	10,000	1,000	1000	1000

*Smallest value of {10, 100, 1000, 10000} that will likely facilitate a successful attack. This is just a rough order of magnitude estimate. Note that a hard restriction is that the sample size must exceed n.

For any particular situation, the ratio of the largest variance to the smallest is $(1+f)^{n-1}$. For example, for n = 128 and f = .025, this ratio is larger than 23. Thus, even with the amount of variance disparity in this case, more than 10,000 replicate samples would be required to make a successful attack.

5. Efficacy of the PTP Method for Class Discrimination

Suppose our test statistic is of the form

$$D = \sum_{j=1}^{m} \left(U_j(\text{new}) - U_j(\text{target})\right)^2,$$

where U_j(new) represents the j^{th} element of the PTP-spectrum from the item being evaluated and U_j(target) is the j^{th} element of the target PTP-spectrum. We conclude that the item is authentic if $D < D_{crit}$.

Case 1: No dimension reduction ($m = n$).

Suppose

$$E = \sum_{j=1}^{n} \left(Y_j(\text{new}) - Y_j(\text{target})\right)^2$$

represents the test statistic that is computed when using the original spectra (Y) modified by Step 0. We conclude that the item is authentic if $E < E_{crit}$. For this case we will also assume the following:

Theorem: For any real-valued n-dimensional spectrum (Y) and any permutations (π and σ), $E = D$ if W is symmetric orthonormal with dimensions n x n. Alternatively, we could specify that $W^{-1} = W^T$ (see example below with

$$W = \frac{1}{\sqrt{N}} H_N,$$

where H_N is a Hadamard matrix of order N).

Proof: The permutations only hide the data. They have no effect on efficacy. $U = Y \cdot W$ and $U_{new} = Y_{new} \cdot W$. Thus,

$$D = (U_{new} - U) \cdot (U_{new} - U)^T = (Y_{new} - Y) \cdot W \cdot W^T \cdot (Y_{new} - Y)^T = (Y_{new} - Y)(Y_{new} - Y)^T = E.$$

NOTE: $W^T \cdot W = W \cdot W^T = I$ if W is symmetric orthonormal.

Consequences of Theorem: If $E = D$ (and $E_{crit}=D_{crit}$), then the classification of an item (authentic or not authentic) is the same whether we use E or D and hence the original spectrum or PTP spectrum. So if E provides adequate discrimination, then D will provide adequate discrimination.

Example: Suppose W is a *normalized* Hadamard matrix (see Section 3.3.1.1). This choice for W seems especially attractive due to its simple binary nature that may make it a good candidate for hardware implementation.

It is very interesting that [SDM97] use a very similar approach for scrambling speech. The authors exploit the fact that Hadamard matrices may be transformed into other *H-equivalent matrices* by permuting rows and columns and by multiplying rows and columns by –1. The resultant transformation of the original data is given by $Y \cdot S$, where

$$S = \frac{1}{\sqrt{n}} P_{row} \cdot H_N \cdot P_{col}. \quad P_{row} \text{ and } P_{col}$$

are signed permutation matrices (row and column, respectively). The signed attribute of the permutation matrix makes this transformation somewhat more general than the case without it. The authors point out that there are $(n!2^n)^2$ *H-equivalent matrices*, but some are identical. Also, the authors provide limited crypt-analysis of this scheme. Finally, the authors mention how this scheme could be further generalized by choosing S to be any well-conditioned normalized matrix with fast algorithm for multiplying with S and S^{-1}.

Case 2: Maximum dimension reduction ($m = 1$).

In this case, W is an n x 1 column vector. Therefore, $U = Y \cdot W$ and $U_{new} = Y_{new} \cdot W$ will now be 1 x m row vectors. Thus,

$$D = (U_{new} - U) \cdot (U_{new} - U)^T = (Y_{new} - Y) \cdot W \cdot W^T \cdot (Y_{new} - Y)^T$$

where $W \cdot W^T$ is an n x n symmetric matrix.

Case 3: General dimension reduction ($m < n$).

The efficacy of the method depends on the relationship between W and the data. It is possible that D will be better or worse than E with respect to classification/discrimination.

In the case of the gamma spectroscopy example, discrimination across classes is relatively easy given the relatively large inter-class spectral differences and the relatively good repeatability of spectra. By mean-centering (translating) and scaling the spectra, we have lost the ability to discriminate based on the average value and/or standard deviation within a spectrum. We have, however, maintained the ability to discriminate based on the shape of the spectrum. An accepted test statistic based on the original spectrum (after translation/scaling) is

$$D = \sum_{j=1}^{m} \frac{\left(Y_j(\text{new}) - Y_j(\text{target})\right)^2}{Y_j},$$

where the denominator could be Y_j(new) or Y_j(target). Note that this test statistic can be rewritten as

$$D = \sum_{j=1}^{m} \left[\left(Y_j(\text{new}) - Y_j(\text{target})\right)/\sqrt{Y_j}\right]^2.$$

A nearly equivalent test statistic is obtained by using the square root transformed spectra (followed by translation and scaling),

$$D^* = \sum_{j=1}^{m} \left(Y_j^{1/2}(\text{new}) - Y_j^{1/2}(\text{target})\right)^2.$$

Note that in the case where Poisson counting errors are responsible for the difference between Y_j (new) and Y_j (target), the square-root transformation is *variance normalizing*. That is, the variance of $Y_j^{1/2}(\text{new}) - Y_j^{1/2}(\text{target})$ does not depend on j. This is due to the fact that the variance of Y is Y in the case of *Poisson* counting errors. Note that D is the sum of the normalized squared measurement differences over the m input channels, where the channel specific normalization is the variance associated with the channel. Note that this variance normalizing transformation would significantly reduce vulnerability to attacks that might utilize the differential variances of input channels (e.g., see Section 4.2).

5.1 Dimension Inflation

Suppose that the dimension of the signal of interest is small enough such that it is computationally feasible for an adversary to examine the whole sample space of possible permutations. One way to reduce the computational feasibility of a brute-force examination of the sample space would be for the host to add pseudo-dimensions that would be indistinguishable when compared to the true dimensions (e.g., both Gaussian white noise). This could improve security considerably.

Let:

1. $Y_\pi W$ be $1xm$,
2. $\boldsymbol{E}$ be a m'-dimensional random variable with independent elements having approximately the same distribution as the elements of $Y_\pi W$ (e.g., normal mean zero and variance 1), and
3. ε be a m'-dimensional random variable with independent elements having mean zero and variance σ_ε^2 of the order of the reproducibility of measurements within a class.

Characteristics of $\boldsymbol{E}$ and ε are as follows: $\boldsymbol{E}$ is constant within a measurement class and variable across measurement classes. $\boldsymbol{E}$ could be a consequence of the seeds used to generate π and σ or derived from another seed that is private to the host and specific for a particular class. ε varies within a class from measurement to measurement (perhaps derived from yet another seed).

The net transmitted signal is of the form $Z=[Y_\pi W \;\; E+\varepsilon]_\sigma$. Z has dimension $m+m'$ with approximately identically distributed elements (statistically indistinguishable). Thus, we have intermixed signals that are relevant to the measurements of the class $Y_\pi W$, irrelevant to the measurement but specific to the class ($\boldsymbol{E}$), and irrelevant to the class (ε). Without ε, the inspector could compare transmitted signals $(Z=[Y_\pi W \;\; E])_\sigma$ and determine the positions of the elements of $\boldsymbol{E}$. Of course, with the addition of ε, there is some impact on discrimination performance. Thus, there is an inherent tradeoff between increased security ($m+m'>m$) and the ability to discriminate between classes (when $\sigma_\varepsilon^2>0$). Naturally, the ability to discriminate decreases as σ_ε^2 increases.

6. Summary

Authentication is difficult because there is no inherent trust embodied in any created thing. In addition, approximate authentication is difficult because a consistent representation of the authenticated item is not available. The need to conceal details of the original signal such that an adversary cannot learn useful information about the original signal adds complexity to the underlying authentication objective. The PTP algorithm offers an information hiding technique believed to be usable in high security applications. By design, this algorithm is such that the value of a simple Euclidean-distance authentication metric based on the PTP output will provide results that exactly match the value of the metric that is obtained by using the original input. Hence, authentication of a sensitive input signal can be achieved indirectly by authenticating the "non-sensitive" output signal. The PTP algorithm has been demonstrated analytically and empirically to provide a high level of assurance that details of the original signal remain unknown and authentication is effective. For those interested in utilizing this algorithm in their own applications, MATLAB, Lisp, and C code is available from the authors.

7. References

[CT97] Y. S. Chow and H. Teicher. 1997. *Probability Theory, Independence, Interchangeability, Martingales.* 3d ed., 336. Springer-Verlag.

[DFM98] G. I. Davida, Y. Frankel, and B. J. Matt. On Enabling Secure Applications Through Off-line Biometric Identification. Proceedings of the IEEE Symposium on Security and Privacy, Oakland, CA, May 1998.

[DS86] R. B. D'Agostino and M. A. Stephens. 1986. *Goodness-of-fit Techniques.* Marcel Dekker.

[HSS99] A. S. Hedayat, N. J. A. Sloane, and J. Stufken. 1999. *Orthogonal Arrays, Theory and Applications.* Springer.

[KT81] Samuel Karlin and Howard Taylor. 1981. *A Second Course in Stochastic Processes.* 454. Academic Press.

[L75] E. L. Lehmann. 1975. *Nonparametrics: Statistical Methods Based on Ranks.* 334. Holden-Day.

[SDM97] V. Senk, V. D. Delic, and V. S. Milosevic. 1997. A New Speech Scrambling Concept Based on Hadamard Matrices. *IEEE Signal Processing Letters* 4(6) June: 161-163.

8. APPENDIX—Discussion of Misclassification Rates

For this discussion, assume that we are attempting to verify that a certain weapon is of Type-XX. We will compare the "transformed/permuted spectrum" of the weapon with that of a known Type-XX weapon. Four possibilities describe the decision space:

	Weapon is Classified as XX	**Weapon is Classified as not XX**
True Weapon Type is XX	Correct Classification	Incorrect Classification (Type I Error)
True Weapon Type is not XX	Incorrect Classification (Type II Error)	Correct Classification

We are interested in quantifying:

- The *conditional probability* of misclassification given the true weapon type is XX-"Type I Error", and
- The *conditional probability* of misclassification given the true weapon type is not XX-"Type II Error".

In general, it is straightforward to quantify the conditional probability of Type I error in case A since the truth ("true weapon type is XX") is very well specified. In the case of Type II error in case B, the truth ("true weapon type is not XX") is not very well specified. Hence, quantification of Type II Error is difficult in this case.

It has been proposed that the classification decision will be made on the basis of a scalar "statistic" or "metric" which is a measure of the similarity between the current and target spectra. Small values of this statistic indicate relatively good similarity. (In the limit, a value of zero is indicative of a perfect match.)

A *threshold value* for the statistic will be used for purposes of classification. When the statistic exceeds the threshold value, the weapon will be classified as "not XX." Otherwise, the weapon will be said to match the target. The selection of the particular threshold value will influence both Type I and Type II Errors.

In order to determine an appropriate threshold, measurements of the same weapon type could be repeated over normal variations of test conditions. A distribution of statistical values will follow. It is a straightforward process for using this distribution to select an appropriate threshold value that is consistent with a specified conditional Type I Error. Given this threshold, it is only possible to express the underlying Type II Error for a specific weapon type (say "YY") being sampled. The ability to discriminate between weapon types "XX" and "YY" will obviously depend on "YY" and the threshold value.

In this context, note the relationship between Type I and Type II Errors and the threshold value. As we increase the threshold value, the probability of a Type I Error will decrease and the probability of a Type II Error will increase.

Also note that the discussion to this point has been limited to conditional probabilities of misclassification. Unconditional probabilities of misclassification depend on the population of items to be tested (e.g., how often does the host try to fool the inspector).

Finally, all of this easily extends to other applications (e.g., biometrics).

DISTRIBUTION:

1	MS	0428	D. D. Carlson, 12300
1		0449	C. L. Beaver, 6234
3		0449	T. J. Draelos, 6234
1		0449	V. A. Hamilton, 6234
1		0449	R. C. Schroeppel, 6234
1		0741	S. G. Varnado, 6200
1		0829	F. W. Spencer, 12323
5		0829	E. V. Thomas, 12323
1		1110	A. M. Johnston, 9211
1		1110	C. A. Phillips, 9211
1		1110	R.A. Lippert, 9214
1		1207	D. J. Mitchell, 5914
1		1207	D. R. Waymire, 5914
1		1361	K. M. Tolk, 5323
1		1361	J. C. Matter, 5323
1		0188	LDRD Program Office, 1030 (attn: Donna Chavez)
2		0899	Technical Library, 9616
1		0612	Review & Approval Desk for DOE/OSTI, 9612
1		9018	Central Technical Files, 8945-1