Specific Emitter Identification and Verification

Kenneth I. Talbot, Paul R. Duley, and Martin H. Hyatt
Northrop Grumman Mission Systems, Command, Control and Intelligence Division

Northrop Grumman Mission Systems has developed the specific emitter identification *(*SEI*) process, a unique, timely method for identifying specific mobile transmitters. SEI has been used by the U.S. government to identify emitters of interest for intelligence gathering or countermeasure activities. SEI designates the unique transmitter of a given signal, using only external feature measurements, by comparing those features with a* library of clusters *(feature sets that uniquely identify a signal) and selecting the cluster that best matches the feature measurements.*

A complementary process, specific emitter verification *(*SEV*), confirms the identity of the unique transmitter of a signal, using only external feature measurements where the identity is known a priori or established by collaterally collected information. The SEI/SEV combined techniques and processes require a set of signal features that will uniquely identify an emitter, as well as a robust collection system that supports both precision measurements and a classification system. SEI and SEV entail precise measurement of signal features that are consistent from one transmission to another for a given emitter.*

Mission Systems has built a proprietary SEI workstation that incorporates the development tools for the feature extraction, measurement, and classification systems described in this article. SEI and SEV techniques and processes have been used successfully in both government and commercial applications.

Introduction

In the mid-1960s, a high-priority problem of the U.S. government was to identify and track unique mobile transmitters for targeting. The government asked Northrop Grumman Mission Systems to develop a unique, timely method for identifying and tracking specific mobile transmitters. *Specific emitter identification* (*SEI*) was the result: SEI designates the unique transmitter of a given signal, using only external feature measurements, by comparing those features with a *library of clusters* (feature sets that uniquely identify a signal) and selecting the cluster that best matches the feature measurements.

We initially defined, built, and delivered a processing system. Over the next four decades, we have successfully applied SEI techniques to signals from communication devices in all radio frequency (RF) bands, as well as those from acoustic and other sources, such as magnetic strips, credit cards, and checks. Although most of our projects were feasibility studies, a high percentage resulted in successfully deployed SEI systems delivered to the respective customers. In one program, a specific mobile emitter was tracked more than 15 years, using only SEI techniques.

Nearly all SEI analysis projects have involved government or military customers. However, since 1994, the complementary *specific emitter verification* (*SEV*) technology has also been applied to commercial customers. SEV confirms the identity of the unique transmitter of a signal, using only external feature measurements where the identity is known a priori or established by collaterally collected information.

SEI/SEV techniques and processes are used in two Mission Systems–built commercial systems to identify whether transmitters are valid—the Inmarsat system and PhonePrint®, the Advanced Mobile Phone System (AMPS) analog cellular system. Deployed in North and South America, as well as Taiwan, PhonePrint terminates nearly one million fraudulent telephone calls every month.

This article presents an overview of a typical SEI/SEV system, the development methodology, and the key processes for system development and operation.

SEI System Design: The Core Concepts

An SEI/SEV system is driven by four key concepts:

- Accurately measuring signal *features* that are consistent from one transmission to another for a given emitter but differ from emitter to emitter
- *Clustering* the *features* by emitter, so that different emitters can be identified or verified
- Entering the *cluster* information into a database and maintaining the *clusters* as the features age in time
- Providing *ground truth* (the correct identification of the emitters being evaluated) for the naming of the clusters and evaluation of the clustering process

The key concepts and principles of SEI and SEV are detailed in this section. We begin with a short discussion and examples of feature clusters, followed by feature measurement, feature set development, identification and verification, and finally classifier management. Later, we describe the elements of a typical SEI/SEV system and how we have developed the SEI/SEV processes and techniques. The major terms used to characterize the SEI/SEV process are defined in the sidebar (page 115).

Feature Clusters

The key to an effective SEI/SEV system is the set of clusters formed from the signal features that make identification and verification possible. Selected SEI features must form natural clusters. Visually, if two features are plotted against each other, then the features of one emitter must group together in an area different from that occupied by the features of another emitter. When two features are plotted against each other in a two-dimensional (2-D) scatter plot, a cluster is formed, as Figure 1 demonstrates. Figure 1a shows 159 intercepts of two separate emitters (different colors) that occupy two separate areas in the feature space. The plot also shows that features can have multiple values—for example, feature 3 for the blue cluster (multimodal feature).

Superior features create wide separation between clusters. If clusters overlap, additional features are required for separation. The other three plots (Figures 1b, 1c, and 1d) show similar clusters for different features with many more intercepts and emitters (36 emitters, 5151 intercepts). Each color in the plots represents a different emitter. Figures 1b, 1c, and 1d demonstrate that many emitters can be separated by just two

Definition of Terms

Class	A *cluster*, or set of clusters, that represents one emitter.
Clone	An emitter modified to imitate another emitter.
Cluster	A collection of the *feature* values of multiple intercepts from the same emitter that group together, with essentially the same mean value. The number of clusters required for a given emitter depends on how the features behave and whether there is *clone* activity. For example, if a feature has two different modes, then at least two clusters are required for the given emitter. If a clone exists, its features will be different from those of the true emitter. Thus, an additional cluster will be required for the clone so that it can be identified.
Clustering	The collection of *features* from the same emitter that tend to cluster or group together in a feature space. Each feature corresponds to a dimension in the feature space.
Feature	A *measurement* of a signal characteristic that is consistent from one transmission to another for a given emitter, but differs from emitter to emitter.
Feasibility	Demonstrates whether a *feature* set can be found that can differentiate emitters. Data are collected at a high signal-to-noise ratio, so that a high-quality SEI feature set can be identified. Data are also collected with real-world conditions, including noise and interference, to determine the effect on the feature set. Five to ten transmissions from each emitter are required to verify consistency. *Ground truth* is required during this phase, in order to determine whether the feature set is sufficient.
Ground truth	The true emitter identity, used to name the cluster and used in the clustering process to determine the effectiveness of the features and to score the effectiveness of the identification and verification results. Ground truth is usually determined from collaterally collected information.
Measurement	An estimate of a specific characteristic of a signal.
Outlier	A *measurement*, or a set of measurements, that does not fit in with the normal set of measurements.
Parameter	An internal coefficient of the insular isometric net.
Perishability	Ascertains by collecting data whether, after a significant amount of time—2 to 3 months—the selected *features* have drifted. The analysis determines how often emitters must be intercepted to update the library in order to keep the feature statistics current.
Signal understanding	Guides the analyst in selecting *measurements* that are possible *features*: • Knowing how the signal is used and modulated • Knowing how different sections of the signal relate to one another • Knowing details of the internal structure and external characteristics

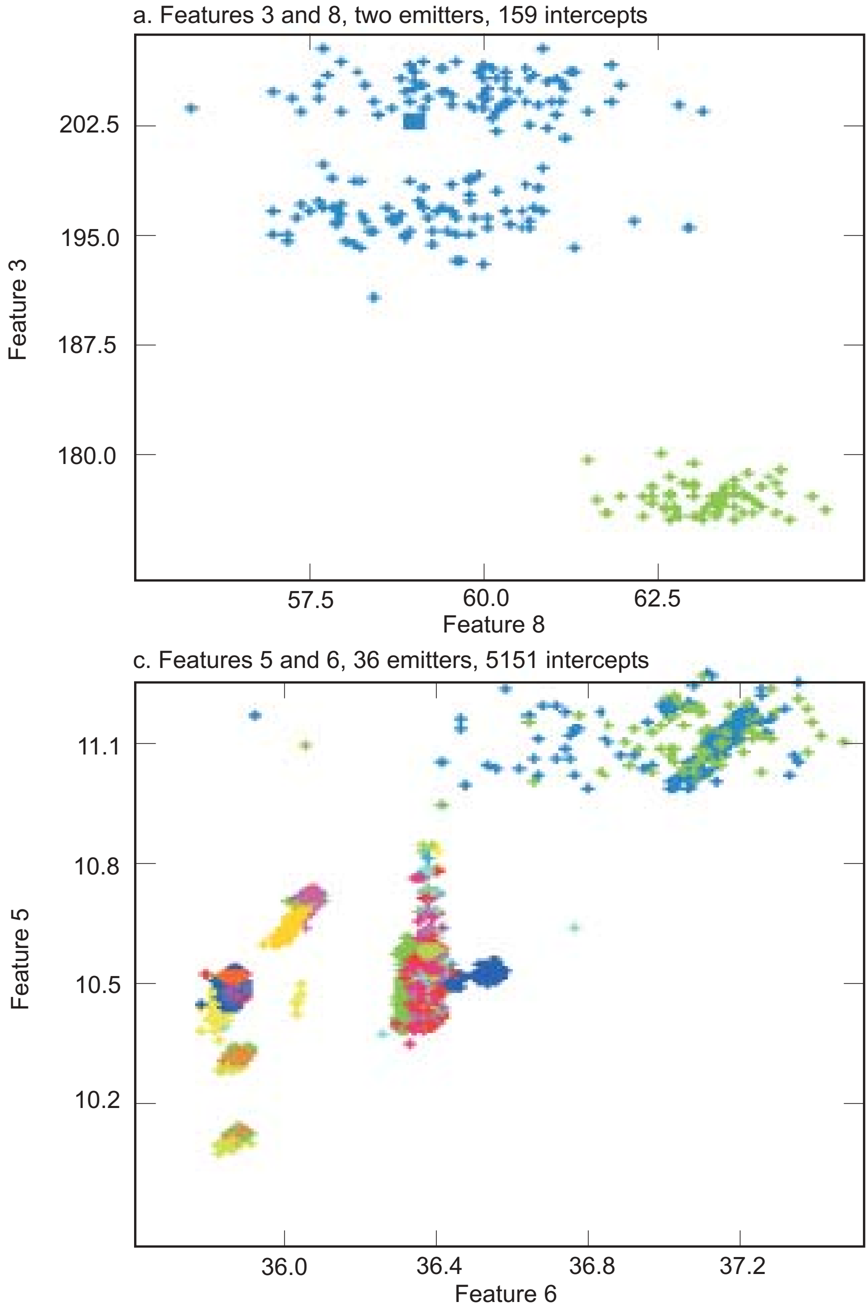


Figure 1. Natural clustering for two emitters and many emitters, where colors represent different emitters

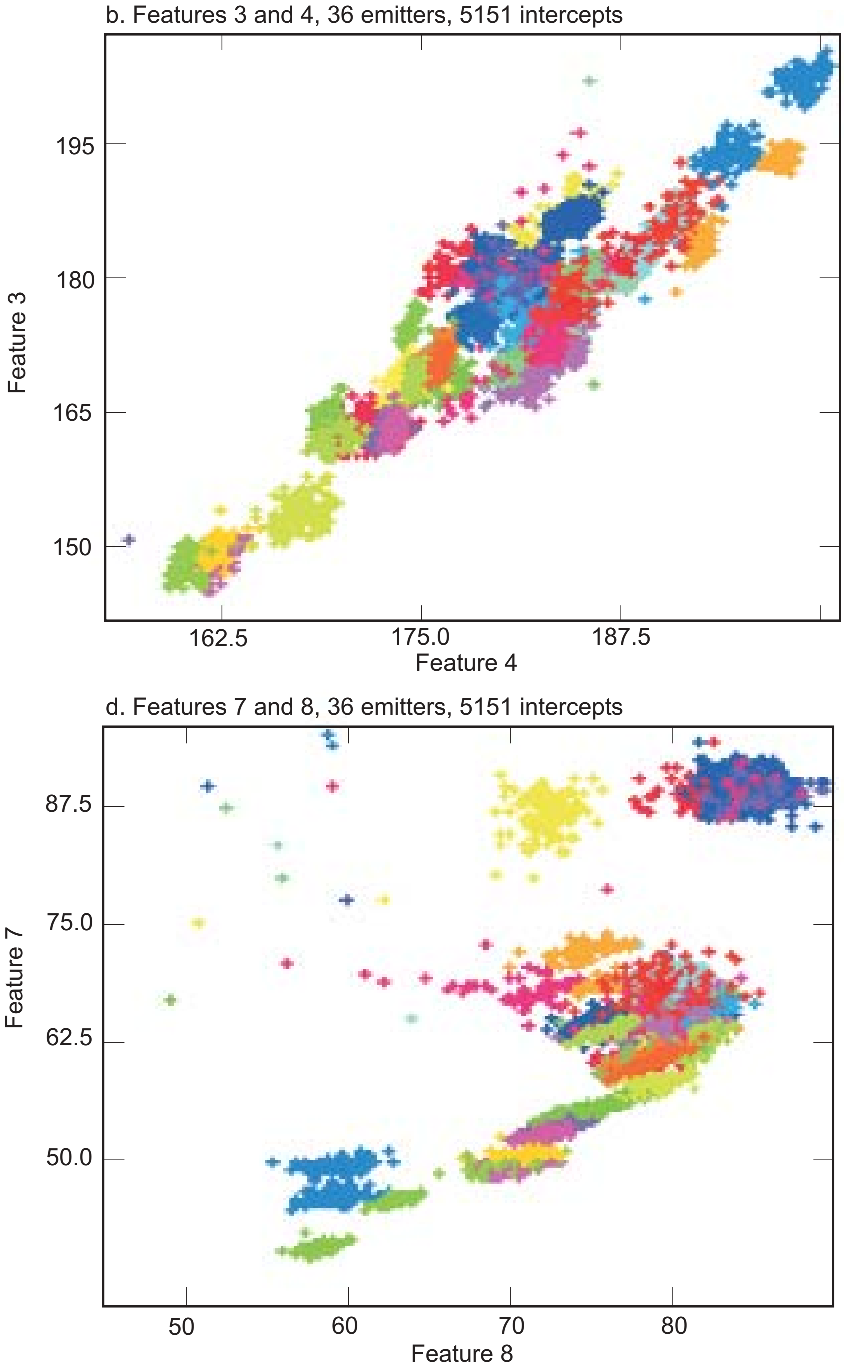


Figure 1. Concluded

features. However, when overlapping clusters occur, additional features are required to achieve separation in a multidimentional space.

A library of clusters is formed from the statistics of the feature estimates. A cluster library entry consists of the class name (obtained from ground truth—a naming action that is performed just once), the class statistics, and any other information that may be necessary. As time passes and the emitter ages, its features usually change. Therefore, timely intercepts of the emitters are essential, so that the changing feature statistics can be updated. We have defined procedures for that process, which are discussed below under “Cluster Growth Management” (page 127).

Feature Measurement

SEI and SEV are both art and engineering. The art involves the selection of features that make identification and verification possible. The engineering is the robust calculation of the features and pattern recognition algorithms. The key to any SEI/SEV system is a set of well-defined features. Not all measurements of a signal are suited to be SEI features. A good SEI feature is a measurement that is consistent for one emitter and different for another.

SEI/SEV systems will always be hampered by feature measurement issues, such as noise, interference, multipath, lost data, and false data. It is best to resolve such issues as early as possible, as they will affect the choice of the measurement techniques. In most cases we have dealt with, noise and transmission multipath are the largest problems.

Generally, many different types of measurements are made and, during the clustering process, those that produce good features are established. Measurements of signal characteristics must be as robust as possible. For example, typical problems are noise and transmission multipath. The algorithms must mitigate such effects, especially for signals in the higher frequency bands. The algorithms must also produce a quality estimate of the characteristic that is being measured, such as a standard deviation. During feature set development, such quality estimates can be used to determine whether the feature measurements may be invalid as a result of high noise or high multipath levels. The quality estimates can then be used in the identification and verification algorithms to reduce false alarms.

Push-to-talk communications typically produce good features, and transient patterns can provide unique identification. For example, Figure 2 shows amplitude transient patterns from five push-to-talk transceivers from three different manufacturers, including two samples from two units manufactured by the same company—Sears Roadmaster and Pace 8047. Not only were the transients distinguishable for individual makes/models, but they also appeared to be unique and sufficiently consistent to separate individual units manufactured by the same company over a 30-day period. Figure 2 displays the two types of turn-on transients: those with lots of spikes and those that are relatively smooth. The spikes may be due to key bounce, whereas the smooth turn-on may result from an electronic keying circuit.

Feature Set Development

The next phase is to identify the best feature set from the set of measurements. If a feature set can be found that will separate different emitters using high signal-to-noise ratio (SNR) measurements, then SEI or SEV is *feasible* for the particular set of emitters.

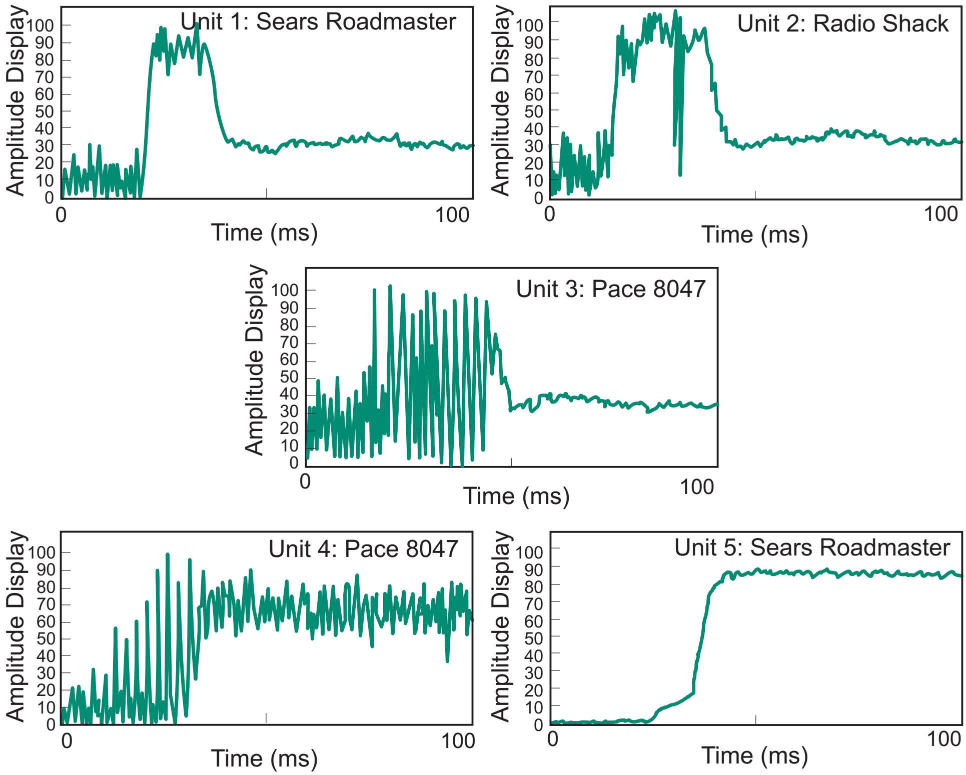


Figure 2. Turn-on transient patterns from five push-to-talk transceivers

Once a candidate feature set is selected, the identification performance is evaluated with the varying noise and interference conditions. Typically, the feature set will require modification to optimize performance. After feasibility has been demonstrated, more extensive collection and evaluation of data are required.

During feature development, it is important to know the correct identification of the emitters being evaluated, called *ground truth*. In cases with few emitters and no ground truth, a system may still be developed, provided features can be designed to give natural clustering. In cases with many similar emitters and no natural clustering, ground truth is necessary to ascertain system performance.

Two processes are used to demonstrate whether a feature set can be obtained from the measurement set. First, the measurement set can be presented in scatter plots marked by ground truth to show whether natural clustering occurs. Second, clusters can be formed according to ground truth and prove whether the emitters can be identified using a pattern recognition algorithm.

Scatter Plots. To evaluate the possibility that a measurement will become a feature, we examine the measurement data. One approach is to visually plot the measurement data using 2-D scatter plots. Figure 3 shows a 2-D scatter plot of two features each of four different emitters. As can be seen, both features of three emitters are multimodal, and three of the four clusters partially overlap. By using projection pursuit methods, we have been successful in visualizing the characteristics of sets of more than 30 features.

Scatter plots become very powerful when the plotting symbols are color-coded according to ground truth. The color is the best visual indicator to identify the measurements that will make SEI features.

Figure 4, taken from PhonePrint, is a projection on a plane of a three-dimensional (3-D) cluster plot with both isolated and overlapping clusters of more than 4000 intercepts, where the clusters are color-coded by emitter identification.

Figure 5 is a scatter plot of two measurements that show little or no discrimination. The measurements from the different emitters overlay one another and provide little separation. Therefore, they should not be considered SEI features.

If the emitter manufacturer is known, the clusters in scatter plots can be color-coded by manufacturer, in order to visually ascertain whether the manufacturer can be determined from the features. Since each manufacturer builds emitters differently, certain features may group themselves, thus identifying different manufacturers, as shown in Figure 6. If such groupings occur, then the features can be exploited and the size of the databases that must be searched during classification can be significantly reduced. In Figure 6, the colors equate to unique manufacturers as follows: dark blue is Marconi, green is Sperry Marine, orange is Nera, light blue is Magnavox, red is Japan Radio, and purple is Toshiba.

Feature Correlation. Correlated features can cause errors in the identification or verification process. The correlation causes a decrease in the calculated distances, which, in turn, may cause identification errors. The correlation should be removed before the features are used in a pattern recognition algorithm. The parameters shown in Figure 1b (page 117) are all correlated. The correlation should be removed for improved identification accuracy.

Cluster Formation. A cluster is a collection of the features that represent the same emitter and are similar in value. We have developed five algorithms that automatically examine the features representing one emitter and specify the required number of clusters. Our algorithms use spanning trees, morphology clustering, nearest neighbor, *normjugation* (i.e., clusters arranged so that, pairwise, the distances separating them will be maximized), and expectation maximization [1,2]. Some of the algorithms operate without knowing much about the data, whereas others require the typical cluster variance to be known.

Clusters are built by establishing the statistics of each feature that belongs to a given emitter's cluster. Once the cluster has been formed, an outlier algorithm is used to remove *outliers* that contaminate the statistical values. After the outliers have been removed, the final cluster mean and variance can be calculated from the remaining data and entered into the cluster library.

Identification and Verification

For *identification*, given a feature set, the question is, "To which class does a signal belong?" For *verification*, given a feature set, the question is, "Does this signal belong to a specific class?" Several existing algorithms can answer these questions. We have found that, for nearly all cases, a distance-ranking algorithm works well, so long as there is little cluster overlapping. However, in cases with considerable overlapping, other algorithms perform better. One that works well for the latter case is the insular isometric net (IIN) classifier. We favor distance ranking for simple cases and, for more difficult cases, the IIN.

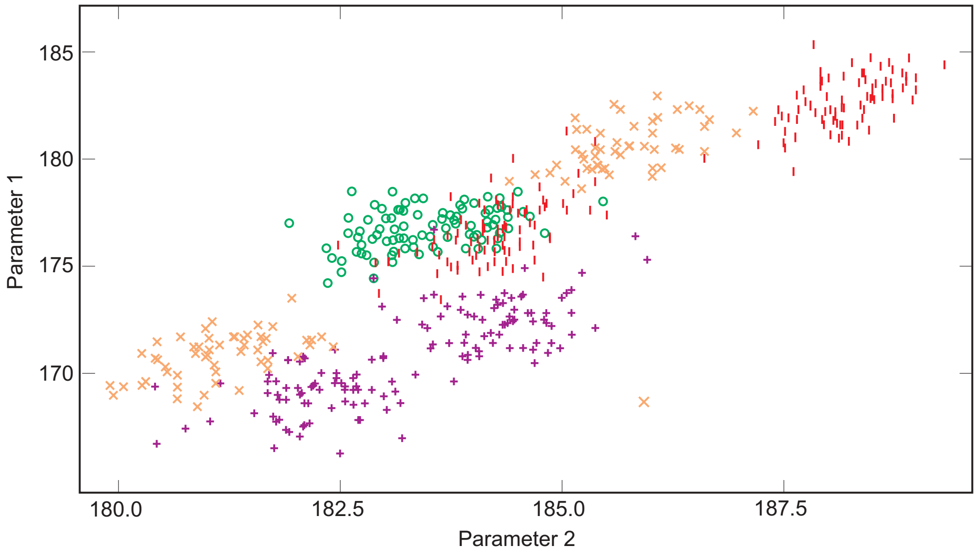


Figure 3. Partially overlapping clusters of two features with three multimodal clusters for four different emitters, where the clusters are color-coded according to ground truth and each color represents a different emitter

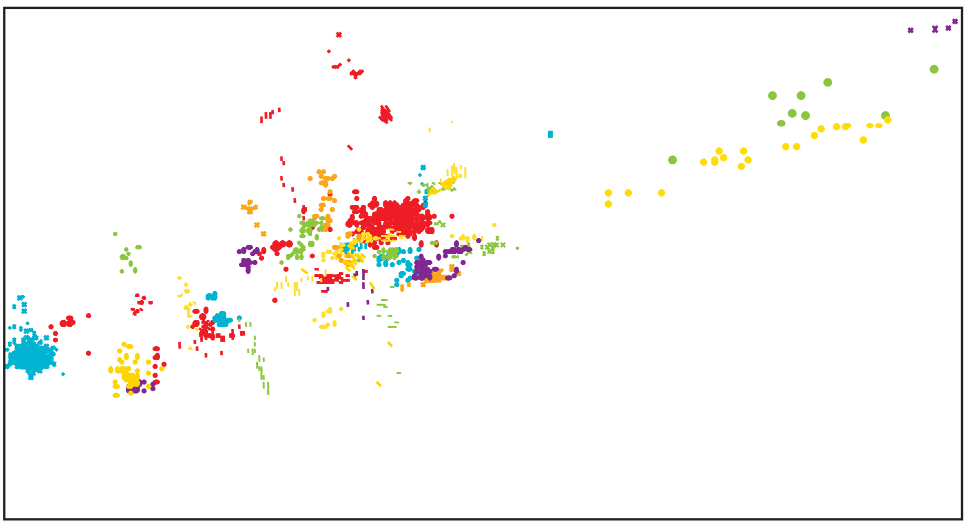


Figure 4. Projection on plane of 3-D cluster plot showing isolated and overlapping clusters of more than 4000 intercepts, where each color and symbol represents a different emitter

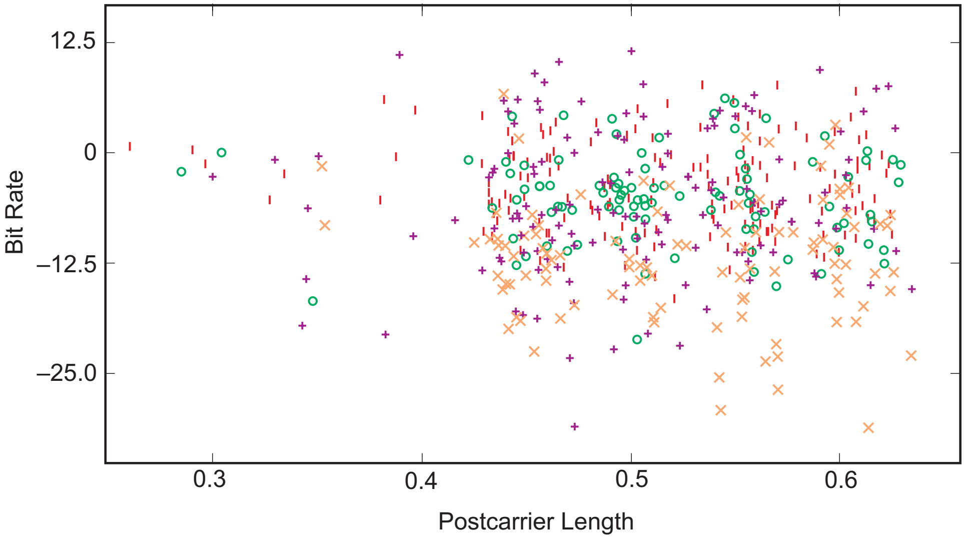


Figure 5. Two different measurements that are poor SEI features

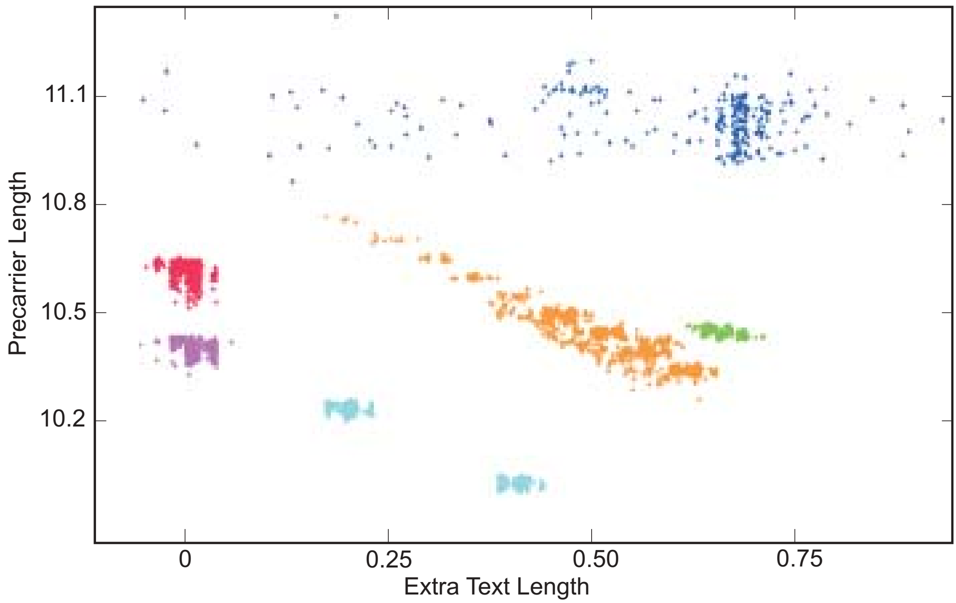


Figure 6. Two features identify manufacturer of emitter through color-coded clusters

Specific Emitter Identification. SEI designates the unique transmitter of a given signal by comparing all clusters in the library with the intercept features. The process returns the cluster names and the normalized distance from the cluster to the intercept of the three best results, as shown in Table 1. The best result is rank 1, the second best is rank 2, and the third best is rank 3. An operator makes his selection based on his knowledge and possibly collateral information. In automatic systems, rank 1 is usually selected if its distance is less than a specified distance and the rank 2 distance is greater than a specified distance. If not, then the intercept is either an outlier or an unknown emitter.

In 1994, a case occurred in which multimodal features and overlapping clusters caused the distance algorithm to produce false identifications. The IIN classifier, developed at Mission Systems in 1988, was then used, greatly reducing the error rate. The IIN is easy to train, and its internal coefficients equate to the typical feature statistics. In terms of processing, the IIN requires only slightly more computing power and disk storage than distance ranking.

Specific Emitter Verification. The method of verification is to compare specific clusters in the library with the intercept features. The specific clusters are determined from collateral information, such as an identification decoded from the signal. For example, in the PhonePrint system, the telephone number is decoded and used to identify the specific clusters. Distance ranking is used to calculate the distance from the intercept to the clusters. If the distance is less than a specified threshold, then the intercept is verified. Otherwise, it is probably a clone.

Those concepts readily transfer to problems outside communications, such as medical analysis of tests, chemical signature analysis, automating credit checks, biometrics, speech, and image analysis. For example, the above concepts readily transferred to a check verification system for identifying counterfeit checks. The check is read by the system before it is issued and the feature set is cataloged in the library by checking-account number. When presented for cashing, the check is read again and its feature values compared with its pre-issue values. If they agree, the check is valid; if not, it is invalid. During tests, the system was 100% effective; it verified every valid check and exposed every counterfeit check presented to it.

Classifier Management

Once an SEI or SEV system has been implemented, the clusters must be maintained with current intercept data to be effective. As emitters age, their feature values will change (drift), so the clusters in the library must be updated. In 1996, we considered the

Table 1. Typical output of identification process, indicating class name and normalized distance from intercept to cluster

	Rank 1		Rank 2		Rank 3	
Intercept	Class	Normalized Distance	Class	Normalized Distance	Class	Normalized Distance
A	C1523	0.83	C1043	1.93	C1425	2.45
B	C1403	1.23	C1352	2.05	C1356	2.98
C	C1435	0.94	C1223	1.84	C1043	2.75

problem and developed the concept of *classifier management shells*. Our multishell architecture combines a number of different methods or algorithms to ensure that the clusters are kept current, cluster quality is maintained, and the proper feature set is used. The architecture also describes a method of integrating knowledge fusion into the process to allow collateral data to aid in identification.

Significant to the entire management process is component structure. The key structuring principle is that some processes work within others and require the outer processes to configure the classifier to work properly. Figure 7 illustrates the relationships among the shells.

A low-maintenance environment for a classifier is necessary because of the overwhelming number of intercepts that can be collected. The clusters must be maintained, because aging causes the signal features to drift over time. We have conceptualized a hierarchy of maintenance software shells that would increase the longevity of the classifier's applicability and decrease its maintenance. Although the management shells were developed for use with the IIN, they can be applied to other classifiers.

Insular Isometric Net. The foundation of the approach is to select a classifier that can be integrated with management shells around it. The IIN is a good choice as a classifier because its two fundamental properties facilitate the management shells. First, its isometric property means that the IIN's internal parameters map into characteristics of the features. Thus, various management tasks can be performed by manipulating the net's parameters. Second, its insular property means that training is localized (i.e., there is no cross training). Hence, the net can be trained independently over input classes and time without retraining on all other data. That property also makes IIN training several orders of magnitude faster than training of more mainstream neural nets with fewer required net variables.

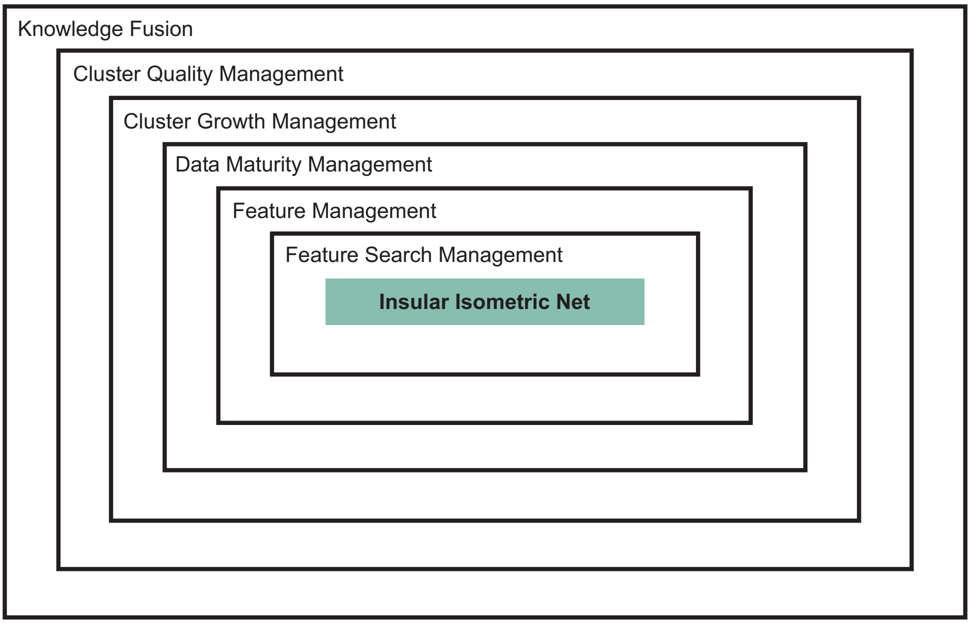


Figure 7. Relationships among classifier management shells

The IIN is a semiparametric algorithm that capitalizes on the advantages of both parametric and nonparametric approaches. Parametric approaches are best when there exists an easily understood model. The model is used to partition the feature space with estimates of the internal parameters. Nonparametric approaches do not use a model but can be effectively applied when data are abundant. For nonparametric classifiers, the feature *space partitioning* is derived from the data. However, many signal classification problems involve a combination of limited model understanding and minimal data. Such problems require the strength of both approaches and are therefore ideally dealt with by the semiparametric approach of the IIN. Further, the IIN can adapt to the best balance between parametric and nonparametric approaches.

We developed the IIN algorithm to capture the advantages of both traditional pattern recognition algorithms (such as model based and statistical) and new approaches (such as the classification and regression tree and the multilayer perceptron neural net). Table 2 compares the capabilities of the various algorithms. The IIN captures the advantages of the traditional and the new, because the semiparametric approach is a hybrid of the two.

An important advantage of the IIN is that it can use different feature distributions. In some cases, histograms of features show that the distributions are not Gaussian; rather, they are logarithmic, beta, or gamma distributions. The IIN can use different distributions to calculate the probability that an intercept is in a particular cluster. Modeling probability correctly is important in reducing classification errors.

Feature Search Management. The feature search management shell automates the process of learning the IIN parameters, which describe the clusters for each class of signals. The large number of similar emitters produces many close clusters. Despite the wide range of clustering algorithms available—from AutoClass (a statistical classifier that determines clusters by optimizing against a probability model of the data, but faces

Table 2. Capability summary: IIN versus existing classifiers

Capability	Existing Classifier				
	Mo [a]	S [b]	C [c]	M [d]	IIN [e]
Includes expertise	X		X		X
Learns from data		X	X	X	X
Requires no model			X	X	X
Produces output in probability form—good for system integration		X		X	X
Adds new signal classes	NM [f]	X			X
Interprets features	X	X			X
Includes feature-quality metric	X	X			X
Accommodates missing measurements	N/A	X			X
Adapts design to meet data/expertise balance	X	X			X

[a] Mo = Model
[b] S = Statistic
[c] C = Classification and regression tree
[d] M = Multilayer perceptron
[e] IIN = Insular isometric net
[f] NM = New model needed

a complex search space when more than a few clusters are involved) to minimal spanning trees (which can handle an unlimited number of clusters)—unfortunately, none of the existing methods are good at handling, simultaneously, close clusters and many clusters. By applying a morphology technique to the derived features, we developed a new, patented morphological clustering algorithm that has been mapped to spanning trees for computational efficiency and will handle the case of similar and close clusters.

Feature Management. The feature management shell performs three primary functions:

- Selecting the best set of features
- Weight-training data according to signal quality
- Masking anomalous features

The feature-selection problem is important because a classifier's performance will degrade in high dimensions, commonly referred to as the curse of dimensionality. The space per sample increases as the dimension increases: For example, 10 features produce 1,024 subsets, but 20 features produce 1,048,576 subsets. At some point, the amount of discriminating information that each sample can provide to the classifier actually decreases. Thus, given a large number of available features with a limited number of samples, it is prudent to try to select an optimal subset. The feature-selection algorithm searches alternative sets, calculating the probability of correct classification and selecting the set that gives the best performance.

When a feature is measured, a quality factor is also estimated. For example, when the average or mean of a data set is estimated, the standard deviation is also estimated. If the standard deviation is large, there is less confidence that the correct mean value has been determined. When clusters are made, the quality features are used to weight the fraction of the feature that is added to the cluster. Data from high-quality signals are given a greater weight. Some features will be affected differently by environmental conditions than other features. Therefore, each feature must be analyzed separately to determine the relationship between signal quality and variation in the feature metric.

Occasionally, a signal is measured properly overall, despite a few problematical features. The poor measurements, called *outliers*, can result in classification errors. Using outlier analysis, such features are detected and declared invalid, so that they do not affect the classifier. (The orange × in the lower right of Figure 3, away from the other orange ×'s, is an example of an outlier.)

Data Maturity Management. If an SEI system is to operate successfully, the cluster statistics must be kept up to date. As emitters mature, the feature values may drift from the originally measured values. Therefore, a procedure must be established that will allow new intercept data to modify the cluster statistics and remove old data. Algorithms have been developed to maintain the cluster statistics over time by combining statistics derived from the current data with previous statistics. When the number of samples in the current set reaches a preset amount, the data statistics are transferred into an archive and a new set is collected. The statistics of the previous data set are combined with the current statistics to provide the cluster statistics.

That design allows vital statistical information to be maintained and just two data sets to be stored for comparison. Such a combination takes into account the quantity of data in each set, the time that the data were collected, and the confidence level of each set. To maintain cluster stability, a typical variance value is used for clusters of small sample size, which gradually transfers to the statistical variance as the sample size increases.

Cluster Growth Management. The cluster growth management shell has several functions that monitor the cluster to ensure that the current configuration is valid and to indicate the need to adapt when the data change in any significant way, such as the following:

- *Cluster alignment.* When statistics are combined, they must be combined with the corresponding matching clusters. The clusters are matched with one another and the files aligned before updating can occur.
- *Cluster alteration.* The IIN is initialized with a specific number of clusters per class and particular basis functions to represent the distribution of samples in each cluster. When changes are required, the clusters are either retrained or updated.
- *Cluster drift.* Similar to the above circumstance (cluster alteration), cluster drift differs because it requires only modifying the features, not changing the number of clusters.

  Outliers. This function removes samples that are not representative of the general population. Outliers distort the specification of the classifier's features.

Cluster diagnostics, common to each of the four issues, must determine

- Number of clusters for each class
- Basis functions to be used
- Proportion of samples in each cluster

Those diagnostics are also important in the cluster quality management shell, which initialized the same factors. Cluster diagnostic methods include density peaks, maximum likelihood measures, Pearson distance, and kurtosis.

Cluster Quality Management. Like the cluster growth management shell, the cluster quality management shell relies heavily on cluster diagnostics, as described above. The difference is that the diagnostics are used for initialization, rather than monitoring.

Knowledge Fusion. The knowledge fusion management shell processes additional external information. The knowledge fusion includes other forms of information that are less amenable to the analysis performed by the IIN. For example:

- Collateral information, such as geographic location, time of day, activities of the transmitted agent
- Conditional factors regarding the measured features that are being analyzed by the IIN, such as modulation and purpose of the signal
- Information from other sources, including that from other systems (such as data fusion), experts, or case reports

A belief net estimates probability among variables. In addition, it incorporates algorithms that propagate evidence among variables. Those representations capture the conditional and unconditional dependencies among variables in a compact form. The algorithm exploits the representation, allowing a significant reduction in the number of computations to solve problems with complex relationships, large numbers of components, and a significant degree of uncertainty. Since the IIN efficiently learns measurement relationships to the signal type, and belief nets efficiently learn other conditional relationships, combining the two tools seems sensible.

Typical SEI/SEV System

Now that the basic concepts have been given, we describe the elements of a typical SEI/SEV system. A typical system generally consists of several subsystems: RF system, data collection, signal processing, feature estimation, identification classifier, cluster management, and database. Figure 8 shows a combined SEI/SEV system. The subsystems shown inside the shaded area are included in the SEI/SEV workstation.

RF Subsystem. The RF subsystem usually comprises a set of RF downconverters to translate the frequency of the signal so that it is compatible with the data collection subsystem. Selected filters should be as linear as possible over the signal bandwidth, and images introduced by the mixers and aliased into the collection system bandwidth should be as attenuated as much as possible. Image attenuations of 60 dB or greater are desirable. Accurate frequency measurements require all oscillators and analog-to-digital (A/D) clocks to be phase-locked to a single common reference that is more accurate than the emitter's frequency accuracy. Receivers with automatic gain control may adversely affect features, particularly turn-on features. In general, fixed-gain receivers are best.

Data Collection Subsystem. Samples of the emitters must be collected with little or no *coloration* (modification of signal characteristics) by a collection system. Proper performance evaluation and feature optimization depend on a wide range of emitters with varying noise levels and interference. High-SNR samples are best for feature development.

To avoid coloring the data with collection system characteristics, special attention must be given to the signal path. In some cases, the SEI system has identified the collection system artifacts, instead of signal characteristics. The collection system must be as pure as possible, because some SEI features measure very subtle differences between signals:

- Local oscillator phase noise must be as low as possible.
- All oscillators in the collection must be phased-locked to a common reference signal.
- The phase response of any bandpass filters must be as nearly linear phase as possible.
- The frequency response of any bandpass filters must be as flat as possible.
- The number of bits in the A/D converter must be maximized to ensure sufficient dynamic range to extract the SEI features. When the signal is being digitized, the signal level should be such that it uses the full dynamic range of the A/D converter minus 6 dB.

Sample rates must be selected carefully. In general, the sample rate should be 5 to 10 times the signal-bit-rate clock frequency. Oversampling is required to ensure accurate length measurements. The signal-bit-rate clock can be measured accurately only if the signal is of sufficient length and the sample rate is not an integer ratio of the signal-bit-rate clock.

A frequency standard should be used to phase-lock all oscillators in the collection system. At times, the carrier frequency of the emitter is one of the best SEI features, and it can be measured accurately only if all oscillators are locked to a common reference. When estimating bit rate, the reference frequency must be more accurate and stable than the signal-bit-rate clock frequency being estimated. For example, if the

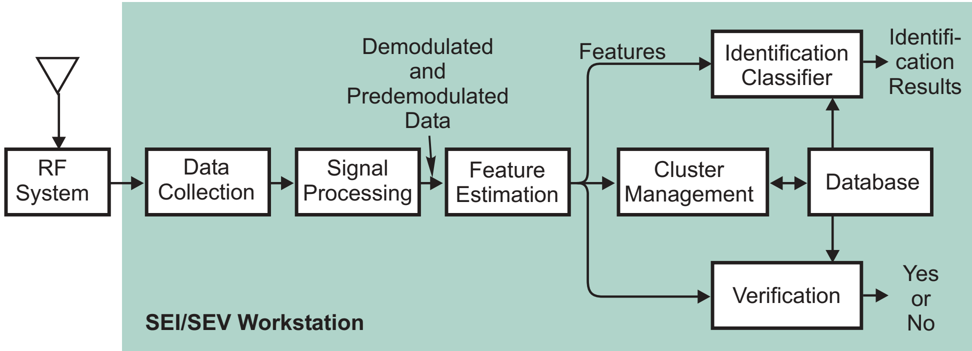


Figure 8. Typical SEI/SEV system

signal-bit-rate clock is derived from a crystal, then the reference frequency should be derived from a standard such as a rubidium or cesium clock.

Signal Processing Subsystem. The data must be properly processed before features can be used in either an SEI or an SEV process. Generally, the signals must be filtered before demodulation and feature measurements. Only linear-phase, finite-impulse response filters should be used. Also, the filters should have little or no amplitude variation over the passband [3].

Phase demodulation requires a constant frequency over the entire signal. Phase-lock-loop demodulators introduce small frequency variations into the signal, which may mask signal variations that a feature may estimate. Normal digital AM and FM demodulators are satisfactory.

Feature Estimation Subsystem. Defined signal features are estimated from the demodulated and predemodulated signal data. Special algorithms have been developed to minimize the noise and multipath effects that contaminate the signal.

Cluster Management Subsystem. This subsystem maintains and modifies the clusters in the database. The algorithms, described earlier, ensure that the clusters correctly represent the characteristics of the signal.

Identification Classifier and Verification Subsystems. These subsystems contain the special classifier algorithm that uses the clusters in the database to either identify the emitter or verify the emitter's identity. In the case of identification, efficient database search algorithms may be required, as all clusters in the database must be examined. Earlier, we have described several classification algorithms, with emphasis on a new algorithm.

SEV requires a collateral method for indicating who the emitter claims to be. Communication systems typically require the decoding of the transmitted data. In PhonePrint, the transmitted telephone number of the cellular telephone was decoded and used as the identify of the telephone. SEV techniques were then used to verify the identity.

Along with selection of the proper SEI features, the classification algorithm is also important. In the past, simple distance-ranking algorithms were sufficient. However, as more precise modulation methods were developed, the measured features became more alike, causing the classification clusters to also become more alike and to overlap with one another. The requirement to classify with little signal model understanding, few data points, and overlapping clusters led us to develop a novel classifier—the IIN. Our classifier can also apply different probability distribution functions that more closely match the feature distributions. Further, it can use multiple clusters for multimodal features. The requirement to work in a continually changing environment has led to a management architecture, briefly described under "Feature Management" (page 126).

Database Subsystems. Typically, any type of database can be used, as long as it will catalog, store, and retrieve data efficiently. For SEI systems with a large number of clusters, a very efficient retrieval system is a necessity, in order to minimize classification time, because so many clusters must be compared with the measured features.

Development Cycle

The SEI/SEV techniques and processes have been developed and continuously refined and improved over many years to identify, process, and extract SEI/SEV features for uniquely characterizing individual transmitters or emitters. The development begins with a signal study, proceeds through data acquisition and technical development, and is completed with the validation and algorithm refinement process summarized in Figure 9. The first three of those processes address the feasibility of finding a feature set that will identify the specific emitters. The last process—validation and algorithm refinement—determines perishability, the validity of the selected features over time. As shown in Figure 9, the overall process is iterative. We elaborate on each of the individual processes in the following subsections.

Signal Study. The signal-generation method is studied and a preliminary set of measurements is defined. Knowledge of the signal-generation method may suggest possible features that may uniquely identify that signal.

Data Acquisition. Data-collection parameters are defined using information from the signal study, such as the sampling rate, required bandwidth, amount of digitization time, etc. Signal data digitized with SNR set as high as possible are used to design the feature estimation algorithms and feature clustering. Lower-SNR data are used to validate performance.

Technical Development. In this phase, a processing system is designed and implemented to estimate the signal features, ascertain whether the features will cluster, and define an optimal feature set. Most design time will be spent on technical development, as it is a highly iterative process.

To develop an SEI/SEV system, first process the collected high-SNR data, so that signal characteristics can be determined. Signal waveforms of different emitters are visually examined to identify possible features. Next, a software program is constructed to estimate the features that have been identified during the signal study and the examination of the data plots. Finally, the collected data are processed and the feature data inserted into a database from which scatter plots are made and feature-selection routines are executed. Ground truth is used to cluster the data, determine the effectiveness of the features, and name the clusters. If the performance is insufficient, signal analysis continues, in order to identify additional features or modify the current features. Additional emitter data may be required for feature refinement or new features. The first three processes constitute the feasibility study, establishing whether or not a feature set can be found that will identify the emitter.

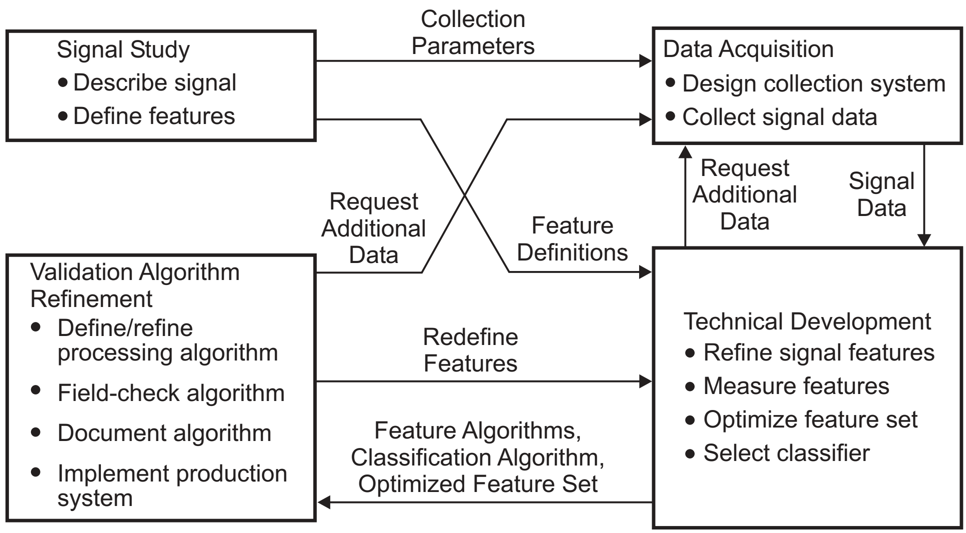


Figure 9. SEI/SEV methodology

The classifier is selected in this phase. For systems with a small number of emitters and reasonable separated clusters, distance ranking has proved very effective. However, with a large number of emitters, our new IIN classifier is generally more effective. Once the classifier has been selected, feature optimization algorithms are then used to determine an optimal feature set. Again, if the performance is insufficient, further refinement of the features is necessary.

The final step is to determine the method of updating the clusters. As emitters age, feature values change and cluster statistics must be updated, so that the clusters reflect the emitter characteristics. We have devised a new cluster management algorithm, described under "Cluster Growth Management" (page 127), that accounts for changing statistics and the elapsed time over which the changes occur.

Validation/Algorithm Refinement. Processes established in the previous phase are validated in this phase. Real-world data (additional data may be necessary) are processed to determine the effectiveness of the algorithms. Usually at this point, the previous phase must be revisited to adjust, refine, and possibly create new features to adjust the overall performance caused by the effects of real-world conditions. Feature sets from two separate collections, usually spaced several months apart, are examined to determine the amount of change that has occurred. That comparison establishes perishability and determines how often emissions must be intercepted to keep the cluster library current. When the system SEI/SEV performance is found to be sufficient, the algorithms and the processes are documented and a production system is implemented, if appropriate.

Summary

SEI and SEV processes have become increasingly important as new communication systems are developed and deployed. Identifying and tracking users as emitters proliferate is increasingly difficult. SEI techniques offer the opportunity to identify or verify specific emitters. The proliferation of emitters also requires processing huge amounts of data to track a few. SEI techniques can be used to filter out most of the uninteresting traffic and concentrate solely on a specific emitter.

The tracking of unknown emitters in a small area, such as a tactical battlefield, can be improved using SEI techniques. When a new signal is received, SEI features are calculated and compared with existing clusters. If a match is obtained, the operator is notified that the emitter has been seen previously. If there is no match, a new cluster is made and assigned an identification number. That procedure allows a number of emitters to be identified and tracked, even if one never determines their identity. Combining line-of-bearing or geolocation techniques with SEI techniques produces a very powerful tracking and identification system. When communication systems transmit an identification number, a verification approach can be used to determine whether cloning is occurring. That information indicates the true number of emitters present, their locations, and the threats they represent.

We have integrated the various elements (digitizing software, signal analysis software, statistical analysis tools, the SEI algorithms, and classification software) into the SEI/SEV workstation that is hosted on a PC using an NT operating system. With the rapid changes in the communication area, the SEI/SEV workstation allows Mission Systems to perform quick-reaction contracts faster, more efficiently, and at a reduced cost to its customers. We have also defined a development methodology that leads from the concepts to an effective operating SEI/SEV system.

References

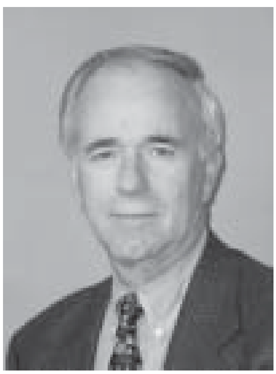
1. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
2. G.A.F. Seber, *Multivariate Observations*, Wiley, New York, 1984.
3. R.E. Crochiere and L.R. Rabiner, "Optimum FIR Digital Filter Implementations for Decimation, Interpolation, and Narrow-Band Filtering," *IEEE Trans. ASSP*, Vol. ASSP-23, No. 5, October 1975, pp. 444–456.

Author Profiles

Kenneth I. Talbot is a Mission Systems Technical Fellow at Command, Control and Intelligence Division's (C2ID's) Electromagnetic Systems Laboratory (ESL) in San Jose, California. He specializes in the application of digital computers to electronic interception and communication systems for signal recognition, analysis, and identification. His work involves both algorithm development (design and testing) and software (design and implementation of signal processing software). Dr. Talbot has been involved with SEI programs since the early 1970s. He developed the special algorithms required for SEI. He also helped design and code the cellular and satellite signal recognizers, signal feature measurement, and fraud prevention systems. That work included analog cellular ("Fraud Prevention System," U.S.

Patent 5905949, for the PhonePrint system) and Inmarsat System A units. Dr. Talbot received Mission Systems' 1994 Chairman's Award for Innovation for PhonePrint. He holds a BS and PhD in electrical engineering, both from the University of Utah. His PhD work focused on microwave power tubes.

ken.talbot@ngc.com



Paul R. Duley currently manages the Signal Exploitation and Research Department in C2ID's ESL in San Jose, California. Since joining C2ID in 1974 as a staff engineer, he has performed and led a wide range of signal processing and signal exploitation analyses. Mr. Duley is recognized as a national expert in specific emitter identification and has been directly involved in all aspects of Mission Systems' SEI development. He has extensive experience in signal search, system technology, and signal acquisition, processing, and analysis through all frequency ranges. He has been responsible for the development and deployment of several successful signal collection and exploitation systems. He received Mission Systems' 1994 Chairman's Award for Innovation for PhonePrint.

rick.duley@ngc.com

Martin H. Hyatt is a senior staff system engineer at C2ID's ESL in Sunnyvale, California. He has 19 years of experience designing advanced algorithms, such as morphological clustering and the IIN. Dr. Hyatt has received several awards for his innovations and a patent for the morphological clustering algorithm. He holds an MS in electrical engineering from the University of Southern California, and two MS degrees and a PhD in decision engineering and management science from Stanford University.

martin.hyatt@ngc.com