Sensor Fusion for Intrusion Detection Under False Alarm Constraints

Matthew Pugh[1]
Jerry Brewer[1]
Jacques Kvam[2]

[1]Sandia National Laboratories [1]

[2]Verdigris Technologies

SAS 2015

[1]Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energys National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND2015-2533 C

Table of Contents

Introduction

What are we doing differently?

Introduction

What are we doing differently?

- Trying to design algorithms with a prescribed false alarm rate

Introduction

What are we doing differently?

- Trying to design algorithms with a prescribed false alarm rate

How is this different than past work?

Introduction

What are we doing differently?

- Trying to design algorithms with a prescribed false alarm rate

How is this different than past work?

- We do not understand the statistics of the events we are trying to detect
- No ROC curves!

Introduction

What are we doing differently?

- Trying to design algorithms with a prescribed false alarm rate

How is this different than past work?

- We do not understand the statistics of the events we are trying to detect
- No ROC curves!

Why is this important?

Introduction

What are we doing differently?

- Trying to design algorithms with a prescribed false alarm rate

How is this different than past work?

- We do not understand the statistics of the events we are trying to detect
- No ROC curves!

Why is this important?

- Mostly focused on detectability
- False alarms cost money

Motivational Questions

How confident can we be in a decision?

Motivational Questions

How confident can we be in a decision?

- Decision theory

Motivational Questions

How confident can we be in a decision?

- Decision theory

What do we have to know to make good decisions?

Motivational Questions

How confident can we be in a decision?

- Decision theory

What do we have to know to make good decisions?

- The more we know the better
- What can be done when very little is known?

Motivational Questions

How confident can we be in a decision?

- Decision theory

What do we have to know to make good decisions?

- The more we know the better
- What can be done when very little is known?
  - No signal model
  - Try to manipulate into something that is known

Motivational Questions

How confident can we be in a decision?

- Decision theory

What do we have to know to make good decisions?

- The more we know the better
- What can be done when very little is known?
  - No signal model
  - Try to manipulate into something that is known

How do design constraints change the system?

Motivational Questions

How confident can we be in a decision?

- Decision theory

What do we have to know to make good decisions?

- The more we know the better
- What can be done when very little is known?
  - No signal model
  - Try to manipulate into something that is known

How do design constraints change the system?

- Detectability versus false alarm

Motivational Questions

How confident can we be in a decision?

- Decision theory

What do we have to know to make good decisions?

- The more we know the better
- What can be done when very little is known?
  - No signal model
  - Try to manipulate into something that is known

How do design constraints change the system?

- Detectability versus false alarm

How to distinguish between noise and not noise?

Motivational Questions

How confident can we be in a decision?

- Decision theory

What do we have to know to make good decisions?

- The more we know the better
- What can be done when very little is known?
  - No signal model
  - Try to manipulate into something that is known

How do design constraints change the system?

- Detectability versus false alarm

How to distinguish between noise and not noise?

Assumption: Components function properly

Test Bed

Sensor Module

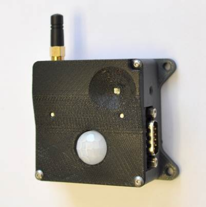
- Tri-axis accelerometer
- Photo-detector
- Passive infrared sensor

Instrumented Room

- Placed 8 sensor modules along walls
- Modules connected via CAN bus

Objective

- Collect background data
- Collected data during entry
- Develop algorithm to detect entry given a false alarm rate
  - **Binary decision problem**



Unknown Everything?

Binary Decision Problem: Intrusion?

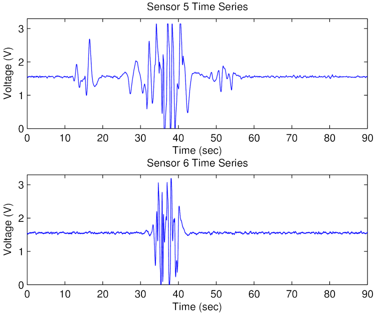
- What are the null and alternative hypotheses?
- What is the distribution of the background noise data?
- What is the structure/distribution of the signal?

Unknown Everything?

Binary Decision Problem: Intrusion?

- What are the null and alternative hypotheses?
- What is the distribution of the background noise data?
- What is the structure/distribution of the signal?

Unclear how to model PIR Sensors



Classic Example: Detection Theory

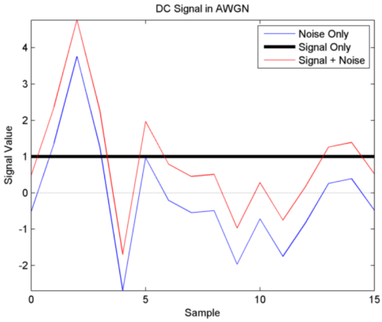
Deciding whether or not a DC signal is present in AWGN

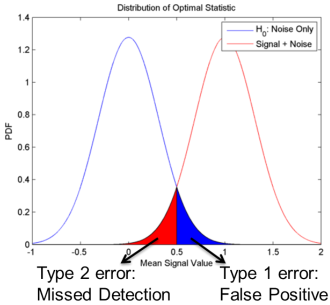
- H_0: noise only
- H_1: Known DC signal + noise
- **Note**: Signal and noise models are known!

Classic Example: Detection Theory

Deciding whether or not a DC signal is present in AWGN

- H_0: noise only
- H_1: Known DC signal + noise

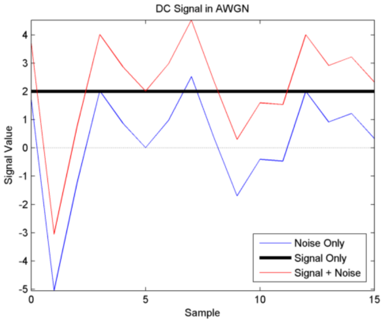


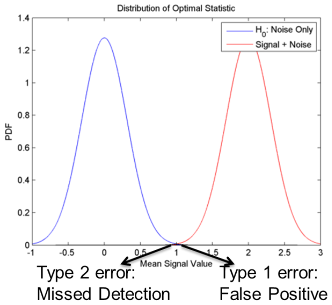


Classic Example: Detection Theory

Deciding whether or not a DC signal is present in AWGN

- H_0: noise only
- H_1: Known DC signal + noise

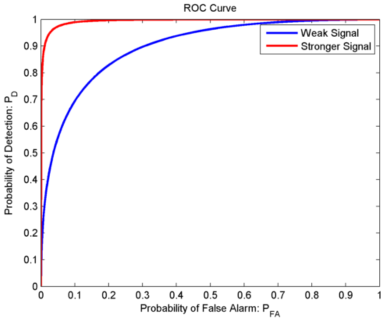


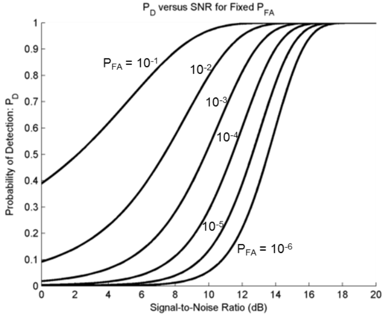


Classic Example: ROC Curves

Error probabilities depend on Signal-to-Noise Ratio (SNR)

- Signal power
- Signal length
- Noise variance



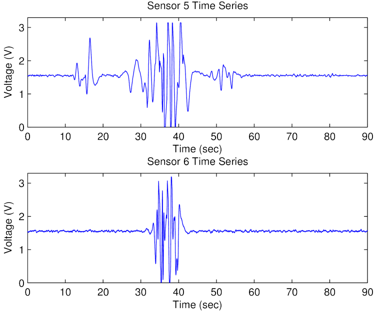


Unknown Everything - Revisited

Binary Decision Problem: Intrusion?

- What are the null and alternative hypotheses?
- What is the distribution of the background noise data?
- What is the structure/distribution of the signal?

Unclear how to model PIR Sensors

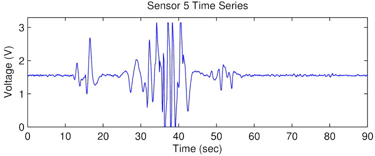


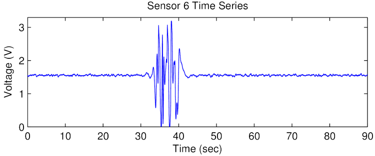
Unknown Everything - Revisited

Binary Decision Problem: Intrusion?

- What are the null and alternative hypotheses?
- What is the distribution of the background noise data?
- What is the structure/distribution of the signal?

Unclear how to model PIR Sensors





Approach

- Model background "noise"
- Declare an event when signal deviates from the background by a specified amount
- Threshold determined by false alarm constraint
- Theoretical ROC curves not possible

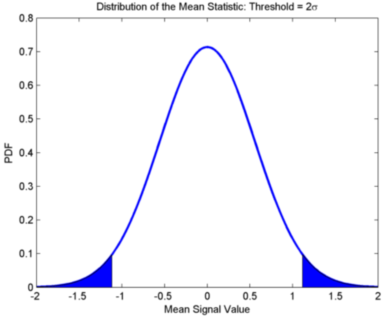
Matching the Noise Distribution

Statistical Model of Noise Distribution → Problem Solved

- Compute threshold to meet false alarm requirement
- Declare an event when signal metric exceeds threshold

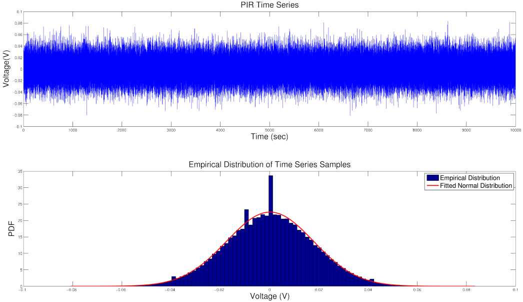
Example

- H_0: Noise only
- H_1: Not noise



- Selected threshold s.t. probability of false alarm is 5%
- Threshold computed from distribution of noise metric
- What is the distribution of the noise metric?

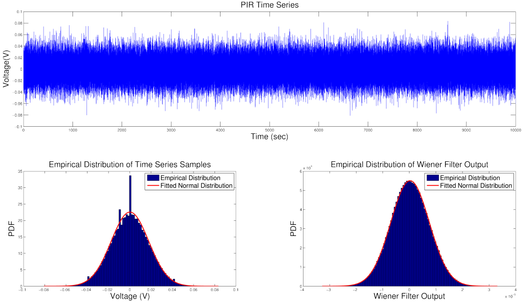
Time Domain Approach



Looks "close" to a Gaussian marginal distribution

- Need to be confident otherwise false alarm constraint is meaningless
- How to have confidence?
    - Match data to theoretical model
    - Gather large amounts of data for empirical estimates

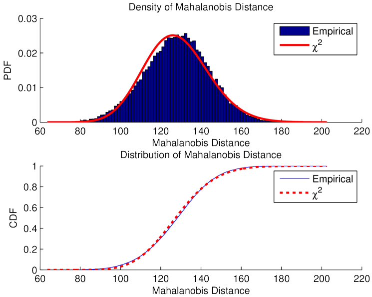
Time Domain Approach



Looks "close" to a Gaussian marginal distribution

- Need to be confident otherwise false alarm constraint is meaningless
- How to have confidence?
  - Match data to theoretical model
  - Gather large amounts of data for empirical estimates

Time Domain Approach

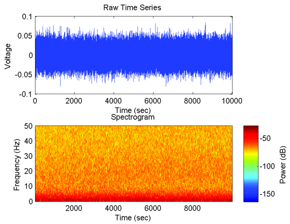


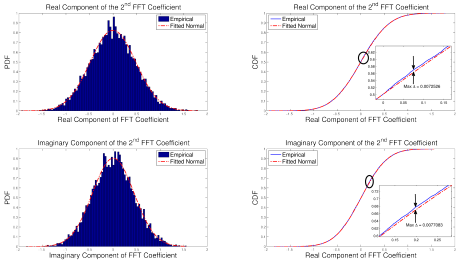
Looks "close" to a Gaussian marginal distribution

- Need to be confident otherwise false alarm constraint is meaningless
- How to have confidence?
  - Match data to theoretical model
  - Gather large amounts of data for empirical estimates

Frequency Domain Approach

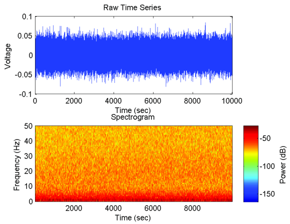
Analyze distribution of frequency components

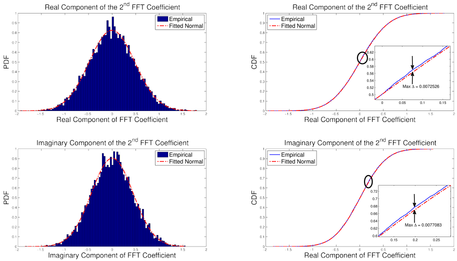




Frequency Domain Approach

Analyze distribution of frequency components

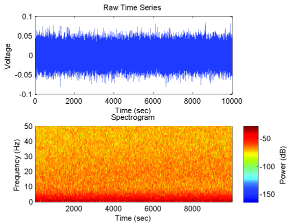


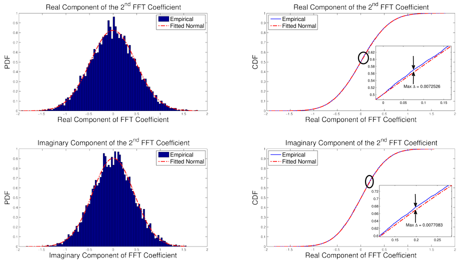


- Distribution of frequency components is <u>not</u> rejected by hypothesis test

Frequency Domain Approach

Analyze distribution of frequency components

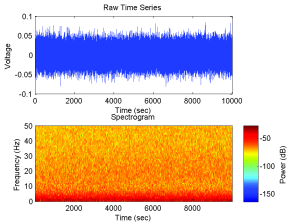


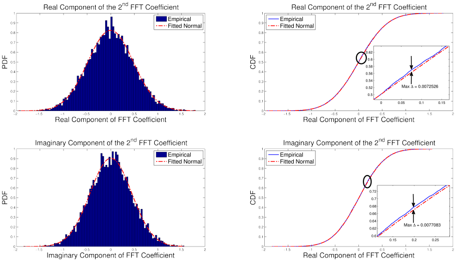


- Distribution of frequency components is <u>not</u> rejected by hypothesis test
- More confidence in match

Frequency Domain Approach

Analyze distribution of frequency components





- Distribution of frequency components is <u>not</u> rejected by hypothesis test
- More confidence in match
- How to combine frequency component information?

Mahalanobis Distance

Want to combine as much frequency information as possible

Mahalanobis Distance

Want to combine as much frequency information as possible

- Requires sub-sampling of frequency components
  - Parseval's Identity

Mahalanobis Distance

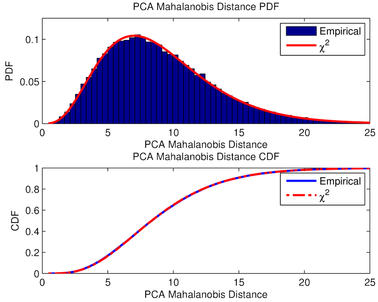
Want to combine as much frequency information as possible

- Requires sub-sampling of frequency components
  - Parseval's Identity
- Use Principal Component Analysis (PCA)

Mahalanobis Distance

Want to combine as much frequency information as possible

- Requires sub-sampling of frequency components
  - Parseval's Identity
- Use Principal Component Analysis (PCA)



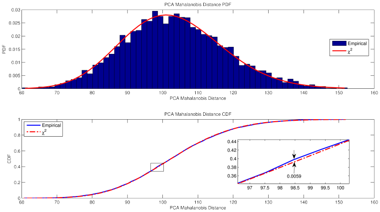
Need metric to combine principal components and sensors

- Mahalanobis distance
- Easily computable
- Known distribution given Gaussian frequency components
- χ^2 distribution for Mahalanobis distance
- Closed-form threshold

Mahalanobis Distance

Want to combine as much frequency information as possible

- Requires sub-sampling of frequency components
    - Parseval's Identity
- Use Principal Component Analysis (PCA)

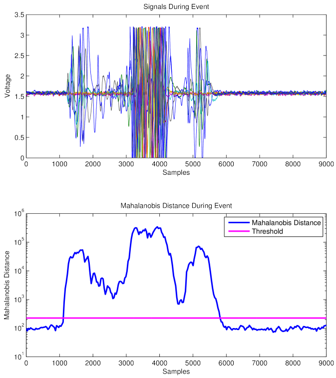


Need metric to combine principal components and sensors

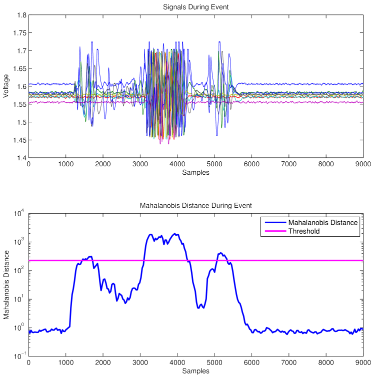
- Mahalanobis distance
- Easily computable
- Known distribution given Gaussian frequency components
- χ^2 distribution for Mahalanobis distance
- Closed-form threshold

Combined Results

- 8 PIR sensors
- False Alarm Constraint: $P_{FA} = 10^{-3}$ per year



Event Data



Scaled Event Data

Future Directions

Adapting Statistical Parameters

- Continuously update estimates of mean and covariance

Future Directions

Adapting Statistical Parameters

- Continuously update estimates of mean and covariance

Optimization of Design Parameters

- FFT length, subset selection method, sample length, new metrics, etc.

Future Directions

Adapting Statistical Parameters

- Continuously update estimates of mean and covariance

Optimization of Design Parameters

- FFT length, subset selection method, sample length, new metrics, etc.

Fully Integrate Sensors

- Combine PIR with photo-detectors and accelerometers

Future Directions

Adapting Statistical Parameters

- Continuously update estimates of mean and covariance

Optimization of Design Parameters

- FFT length, subset selection method, sample length, new metrics, etc.

Fully Integrate Sensors

- Combine PIR with photo-detectors and accelerometers

Sensor Failure Detection

- Current algorithm declares an event when threshold is exceeded
    - Sensor failure could cause algorithm to exceed threshold
- Need to disambiguate between failures and events

Conclusion

Focused on development of detection algorithms with false alarm constraints

- Found metric on background data that matches known closed-form distribution
  - Frequency components
  - Subset Selection: Principal Component Analysis
  - Mahalanobis Distance: χ^2 distributed
    - Combine all PIR sensors into a single metric
- Determine threshold to meet false alarm constraint
- Algorithm performs well on collected data

Still a lot of work to be done

Conclusion

Thank You!

Special Thanks:
Jacques Kvam
Jerry Brewer

Any Questions?