Robust Geotag Generation Algorithms from Noisy Loran Data for Security Applications

Di Qiu, Sherman Lo, Dan Boneh, Per Enge

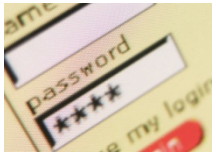
Stanford University

ILA 2009

Sponsored by FAA Loran Program CRDA 2000-G-028

Geo-security: Why?

- Doesn't require memorization



  - Users don't have to memorize location dependent signal characteristics.

- Resistant to misplaced tokens

  - Users won't forget to bring location.

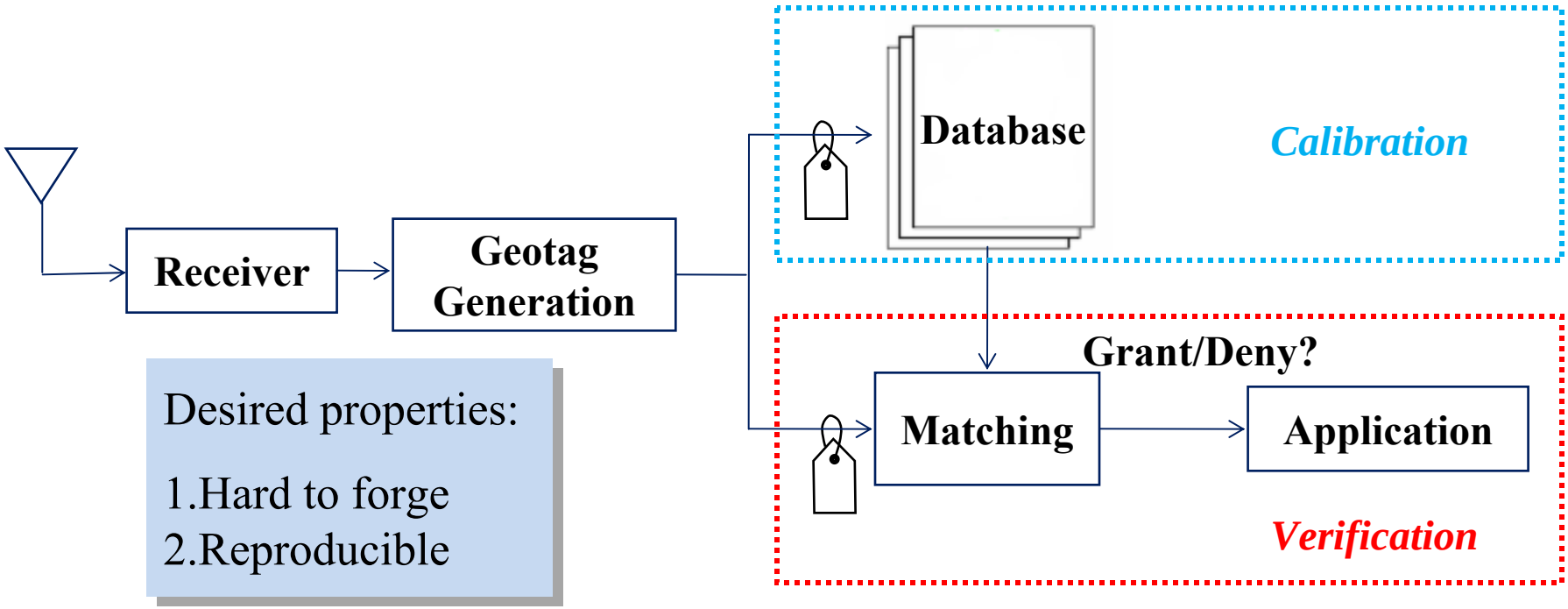
- Can't be delegated
  - Users can't lend location to someone else.
- Low awareness
- Don't need physical access to attack

"On the Internet, nobody knows you're a dog."

Geo-security: Big picture

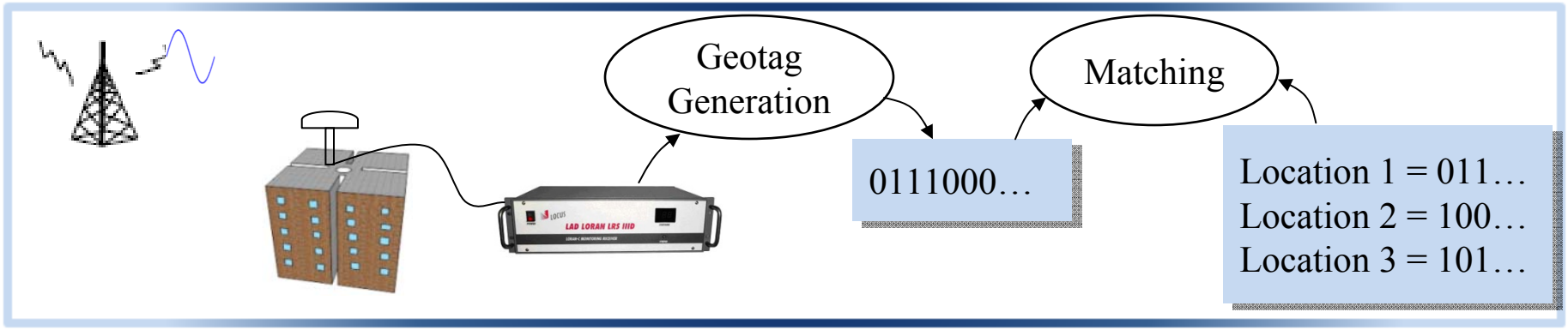
- Use location-dependent signal characteristics from multiple transmitters.
- Restrict access of information content or electronic devices.

Geo-security: How?



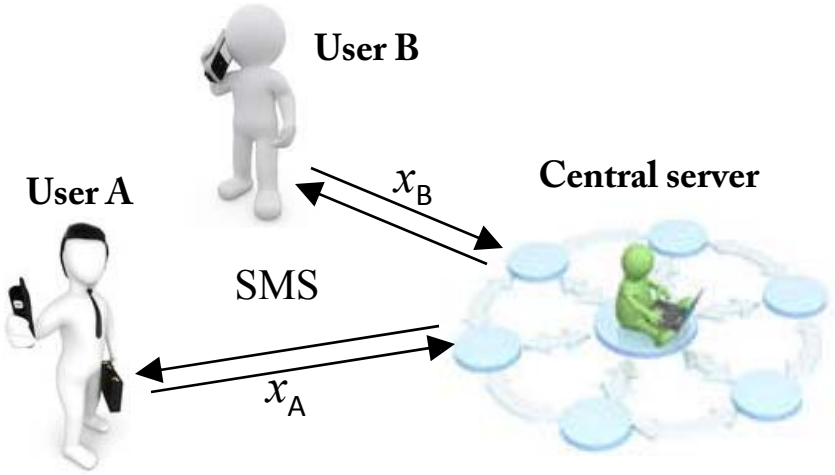
Desired properties:

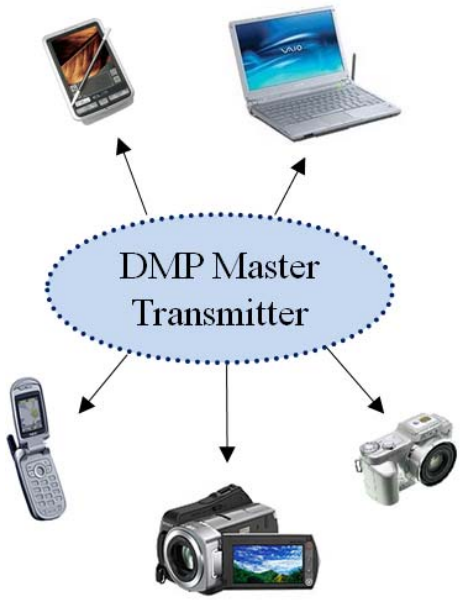
1.Hard to forge
2.Reproducible



Geotag applications

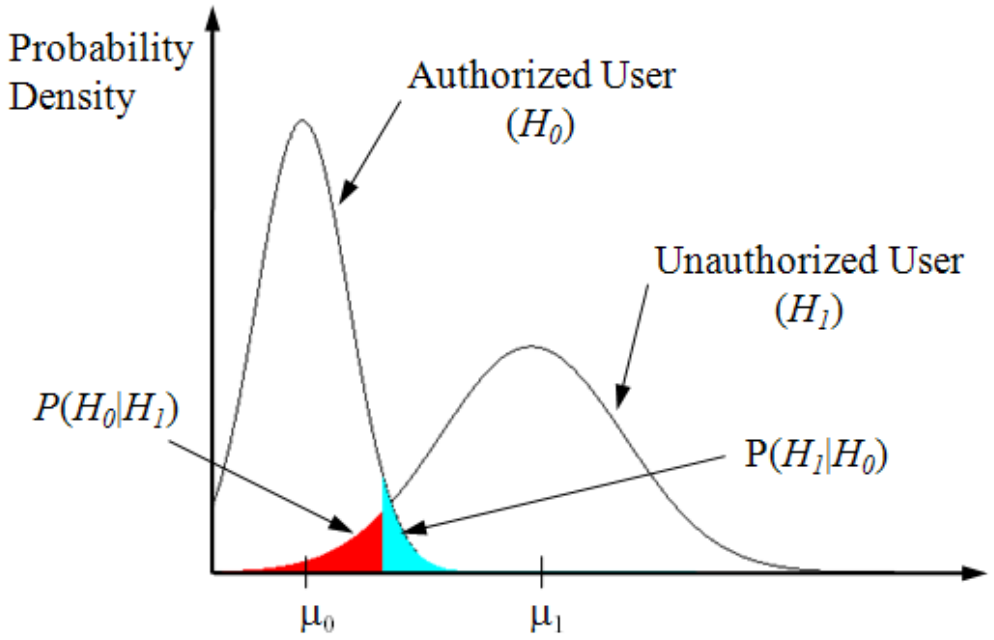
- Loopt
  - Social networking
- Digital Manners Policy
  - Microsoft
- Data access control
  - Geo-encryption
- Geo-fencing
  - Laptop anti-theft technology





Performance standards

- Two hypotheses
  - H_0: accepting as authentic user
  - H_1: rejecting as an attacker
- Two errors
  - False reject: accepting hypothesis H_1 when H_0 is true
  - False accept: accepting hypothesis H_0 when H_1 is true



Geotag generation

- Quantization-based

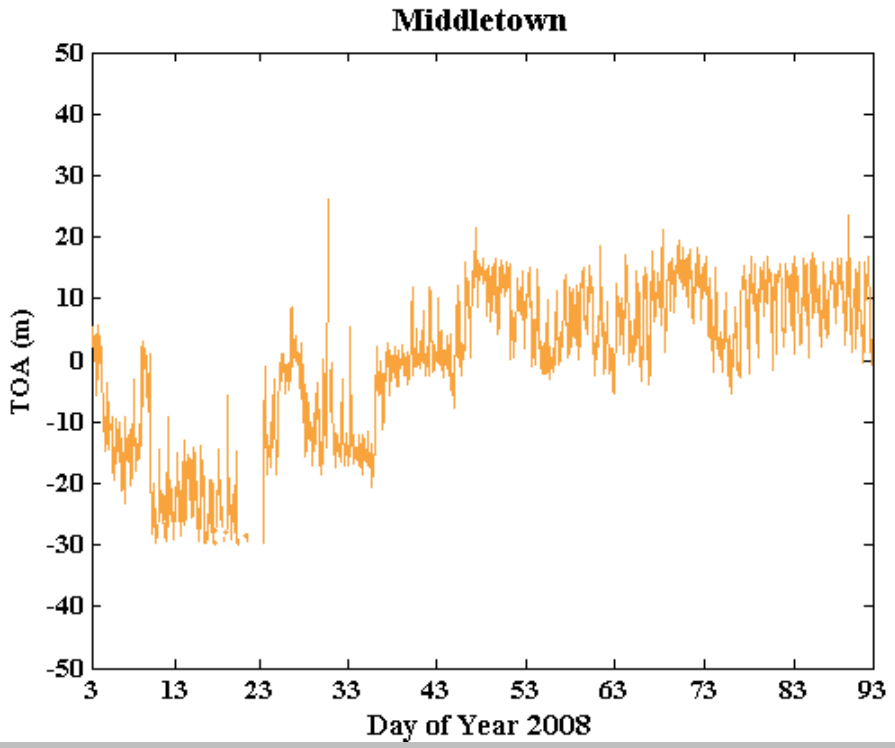
$$T(i) = k; x_i \in S_k = [k\Delta_i, (k+1)\Delta_i)$$

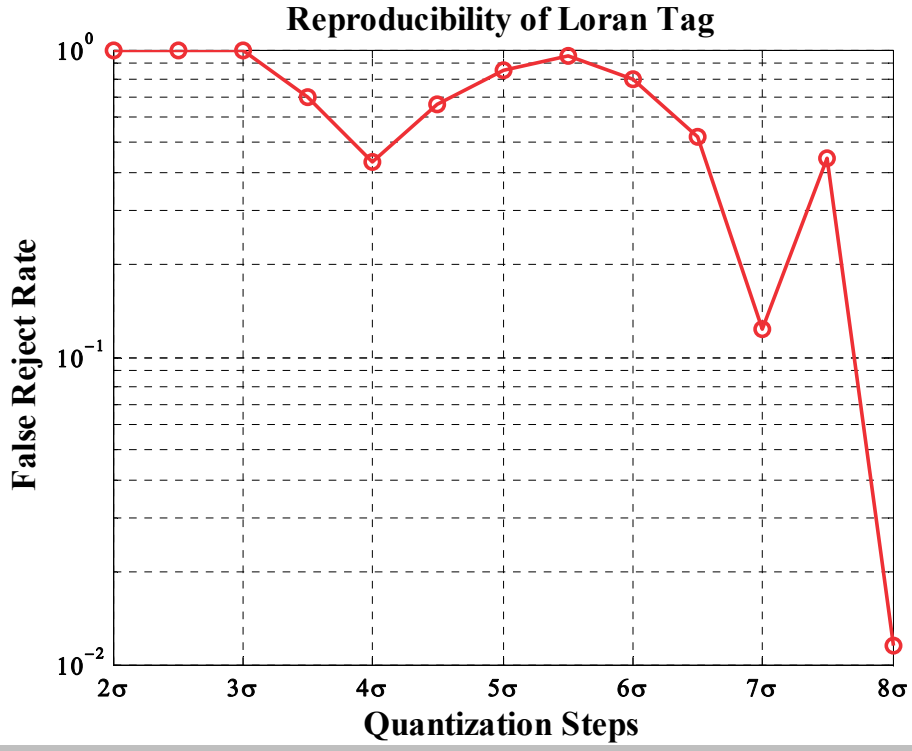
$$M(\widetilde{T}, T') = \begin{cases} 1 & \text{if } \frac{1}{n}\sum_{i=1}^{n} \widetilde{T}(i) \oplus T'(i) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

- Fuzzy extractor-based
- Pattern classification-based
  - k-Nearest Neighbor (kNN)
  - Support Vector Machines (SVM)

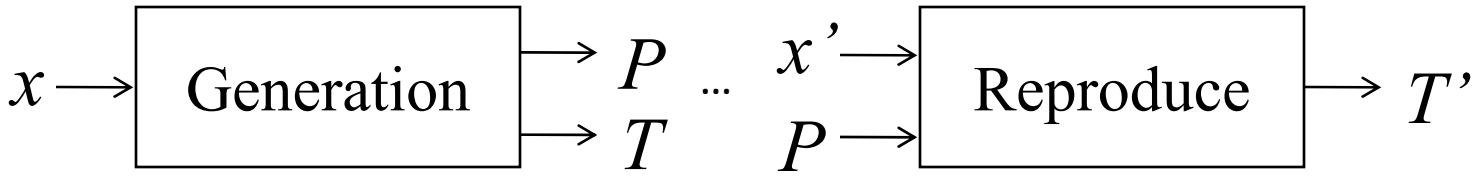
Quantization-based geotag reproducibility

- Tamper-resistant device and self-authenticated signals for spoofing attacks.
- Temporal variation degrades performance
- Parameters --TD, ECD,SNR-- from GRI 9940
- Non-monotonic trend comes from quantization





Fuzzy extractor



Definition. A fuzzy extractor is a tuple (M, t_0, Gen, Rep), where M is the metric space with a distance function *dis*, Gen is a generate procedure and Rep is a reproduce procedure, which has the following properties:

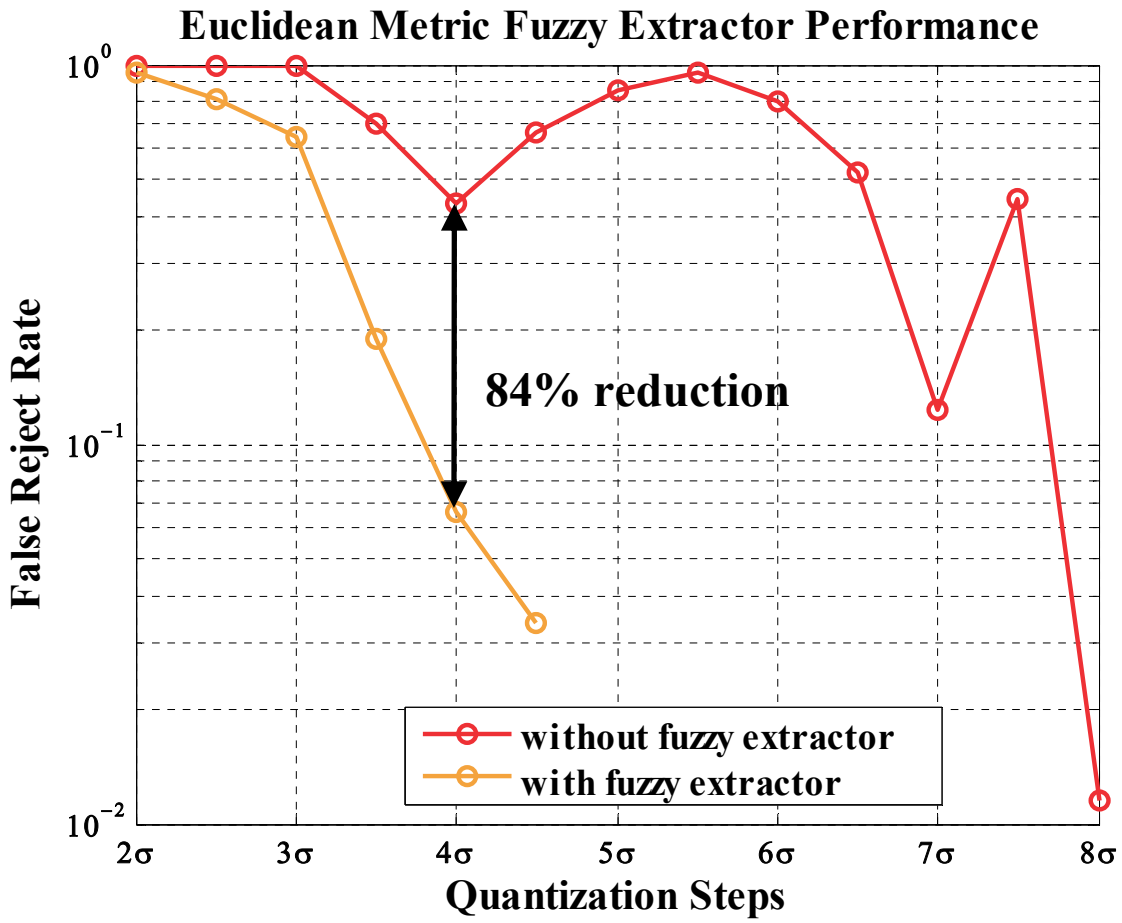
1. If $dis(x, x') \leq t_0$, $T' = T$.
2. If $dis(x, x') \geq t_0$, $T' \neq T$.

Y. Dodis el al., "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," 2004.

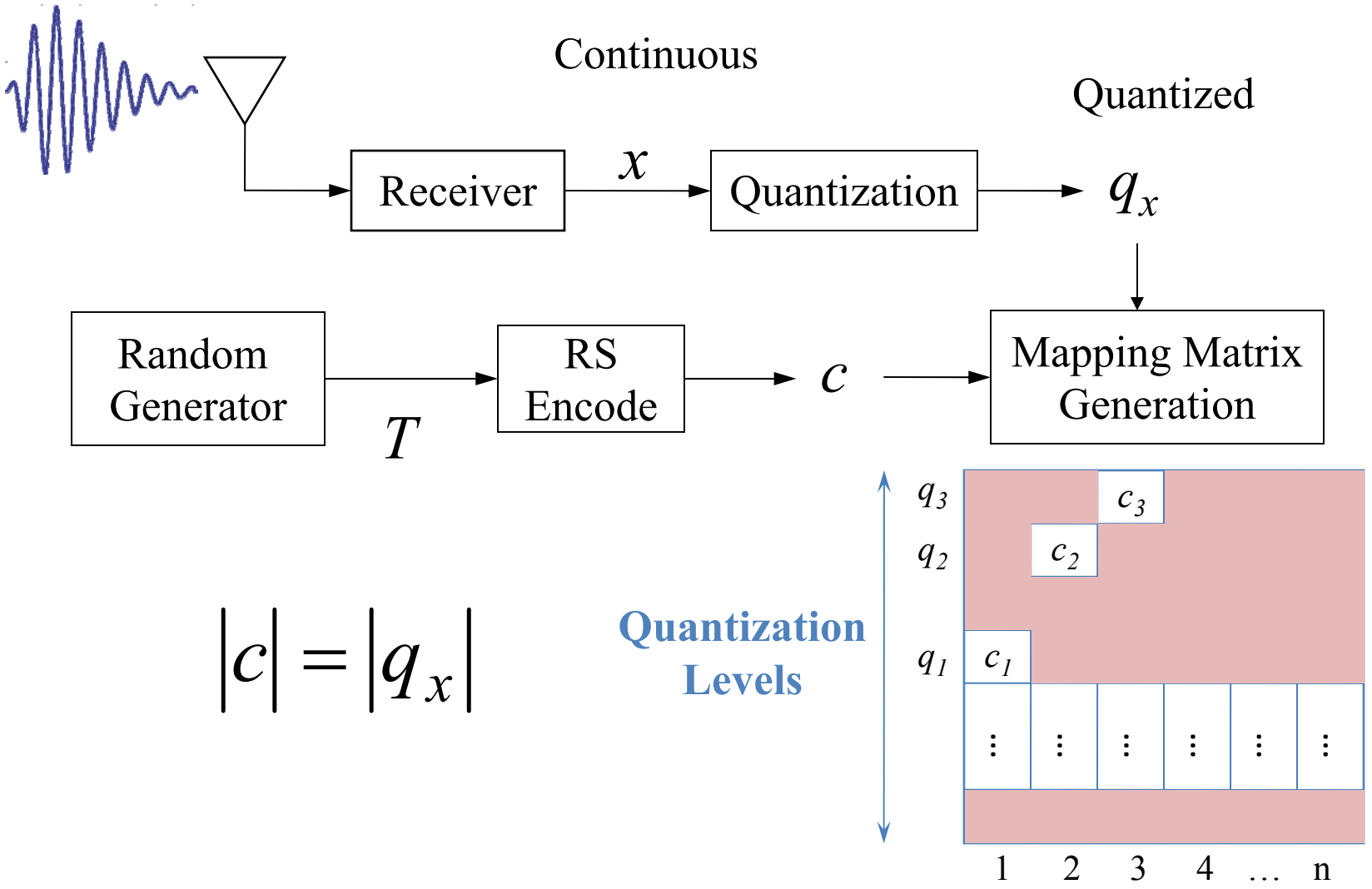
Three fuzzy extractors for location data

- Euclidean metric fuzzy extractor
  - Noise, bias, and quantization errors
  - Adjust offsets
  - Real numbers
- Hamming metric fuzzy extractors
  - Constructions
    - Reed-Solomon based fuzzy extractor
    - Secret Sharing based fuzzy extractor
  - Offline transmitter
  - Inputs are integers: quantized values of location parameters

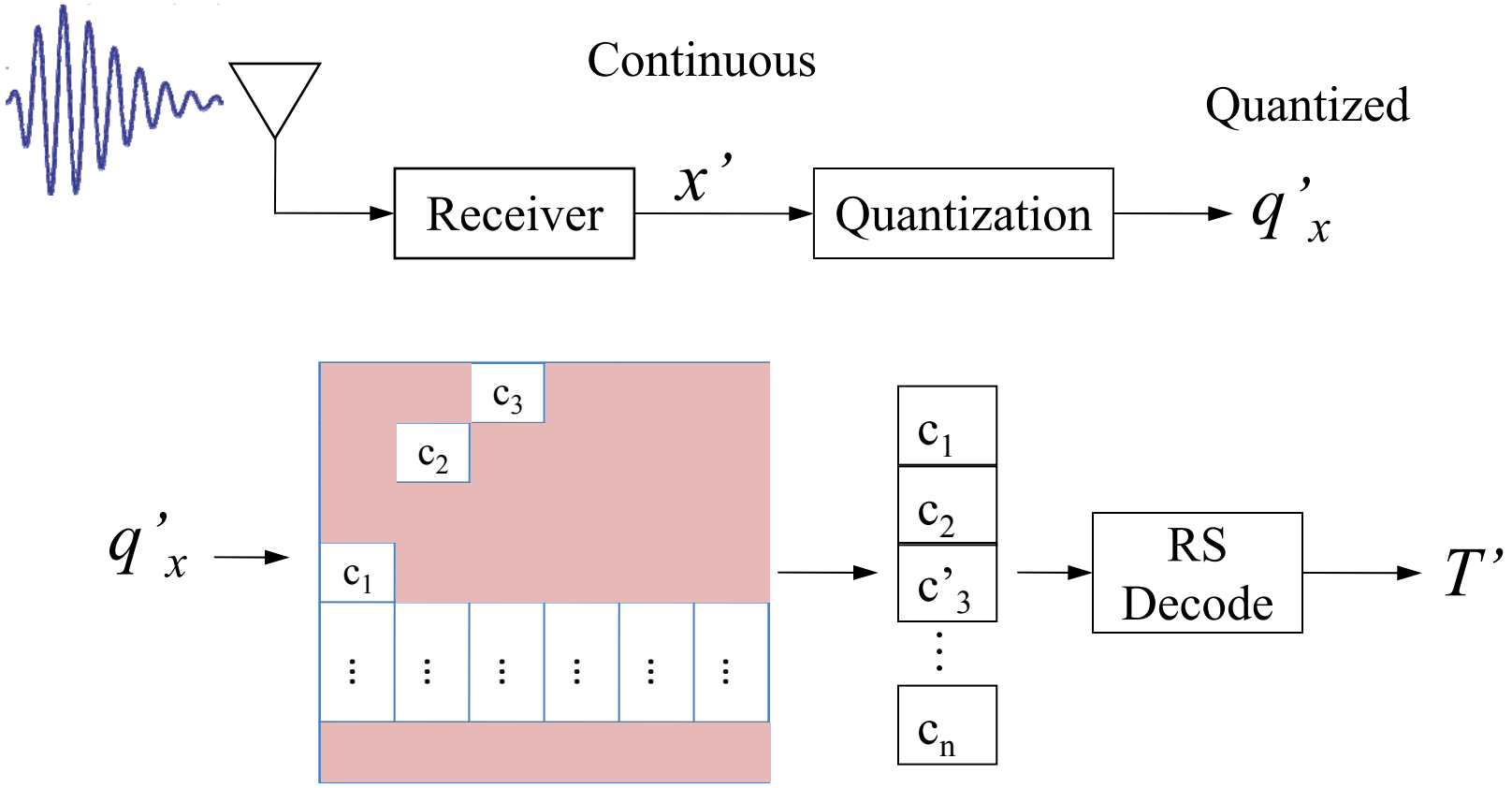
Performance of Euclidean fuzzy extractor



RS-based fuzzy extractor – "Locking"



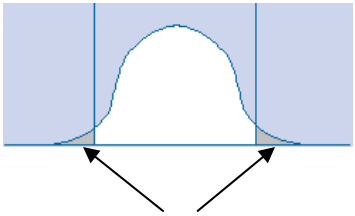
RS-based fuzzy extractor – "Unlocking"

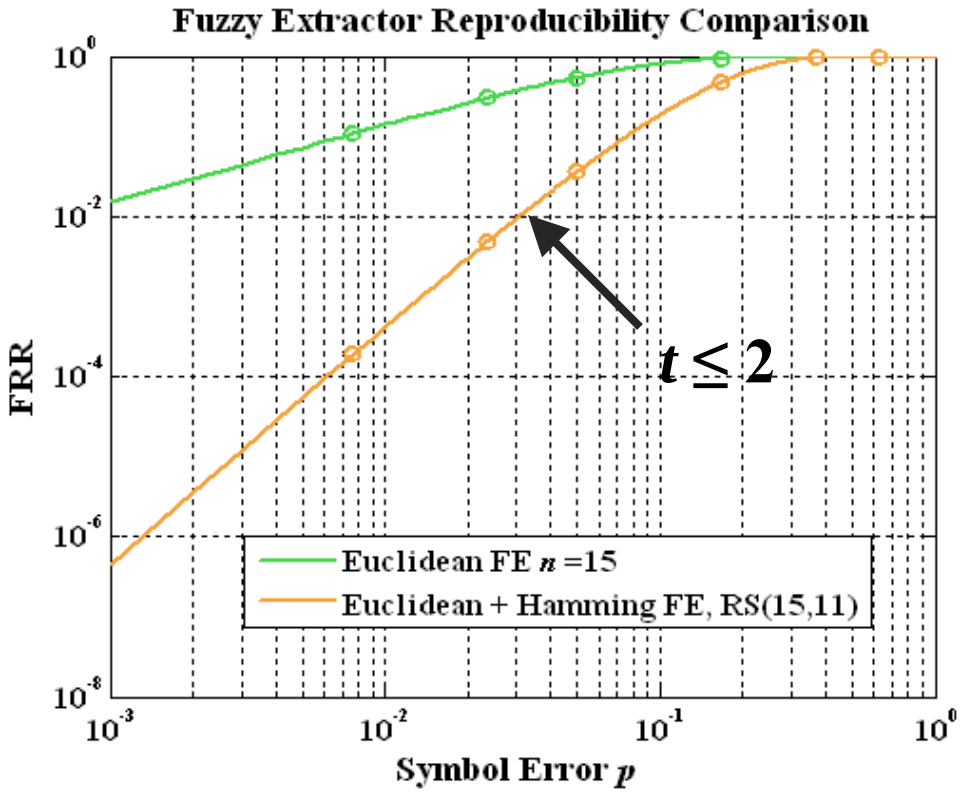


$T'=T$ if $\mathrm{dis}(q'_x, q_x) \leq t$, t is a design parameter

Performance analysis: RS-based

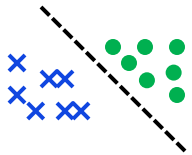
$$\Pr\{\text{error or decode failure}\} = \sum_{j=t+1}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$





Pattern classification

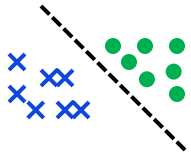
- Goal: Extract decision rules from data to assign class labels to future data samples.
  - Maximize the difference between classes
  - Minimize the within-class scatter
- The quality of a feature vector is essential to spatial discrimination/decorrelation.



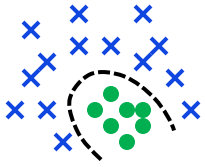
"Good" features

"Bad" features

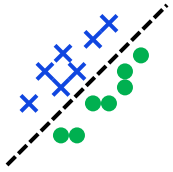
Properties



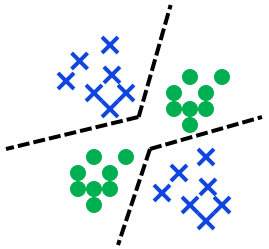
Linear separability



Non-linear separability



Highly correlated

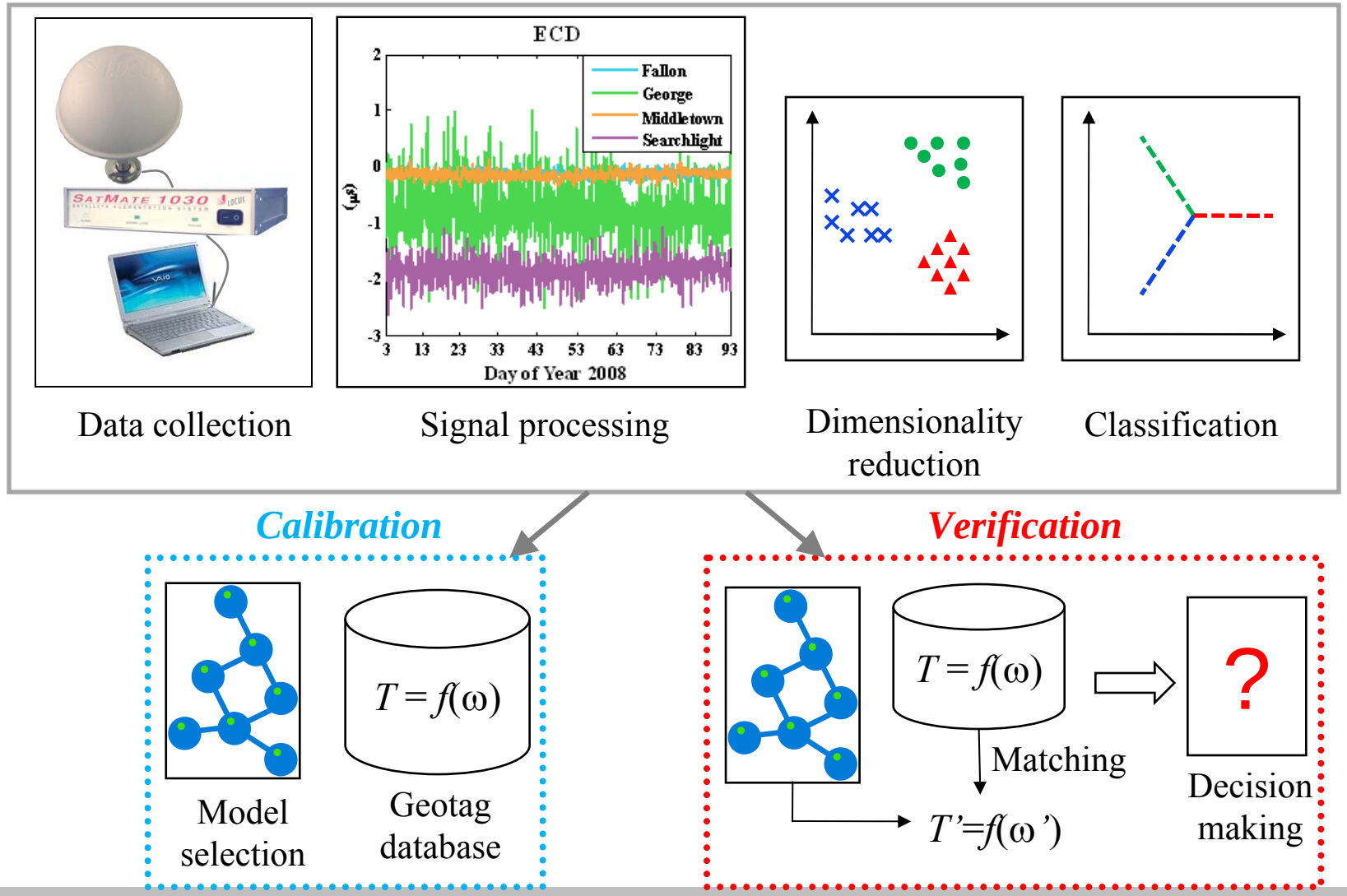


Multi-modal

Dimensionality reduction

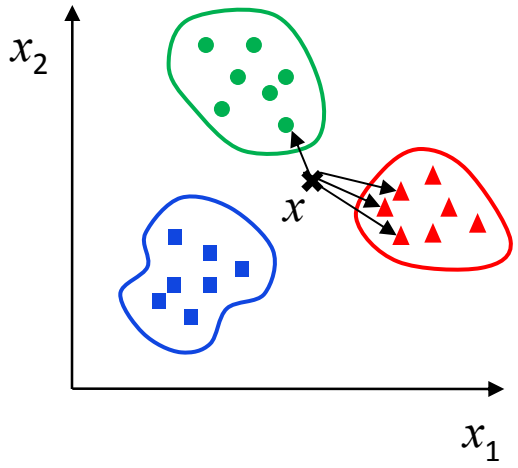
- "Curse of dimensionality" – high dimensional data are difficult to work with.
  - The added parameters or features can increase noise
  - Enough observations to get good estimates
- Dimensionality reduction
  - Efficiency – computation and storage costs
  - Classification performance
  - Ease of modeling

Geotag generation using pattern classification

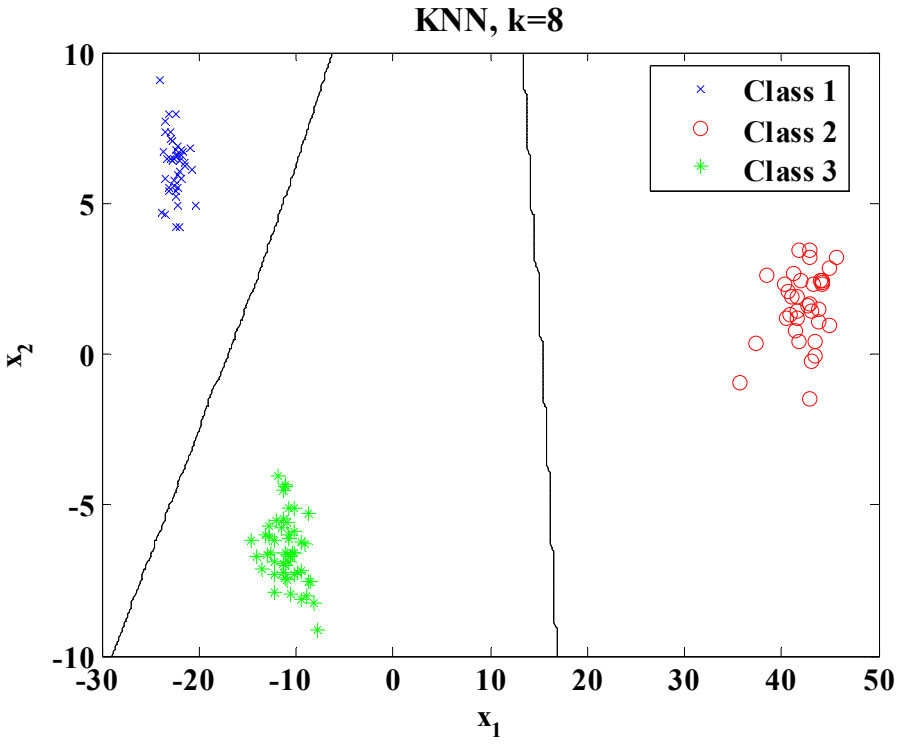


K-Nearest Neighbor (kNN)

- 'Memory' based classification
- No training phase is required: 'lazy' learning approach
- Given data point x, find the k nearest training inputs x_1, x_2,…, x_k to x using a distance metric.
- Large k produces smoother boundaries and reduces the impact of noise.
- Computational cost



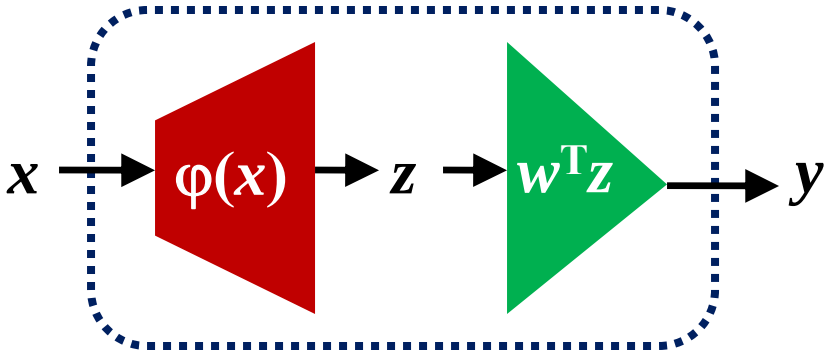
Visualization of kNN, k=8

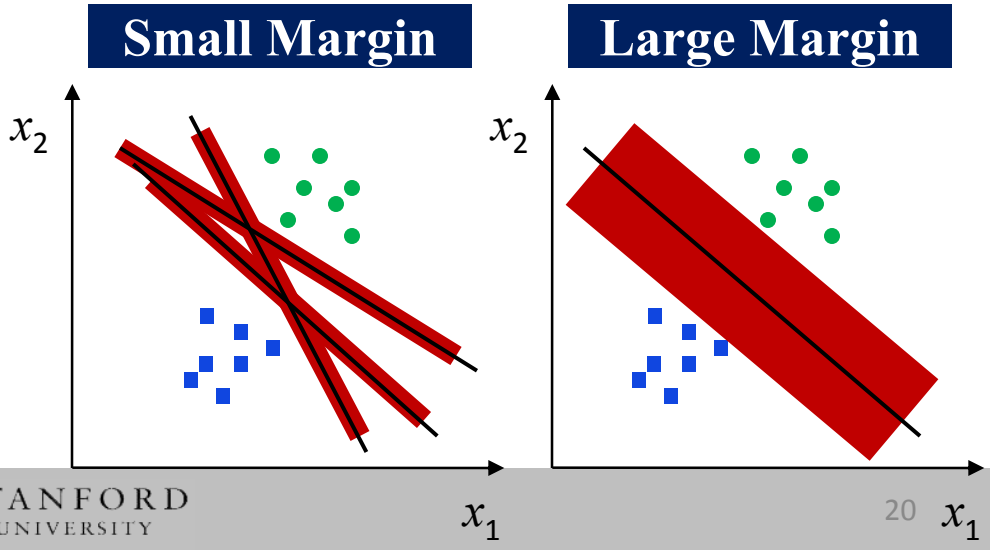


- Easy to implement but computationally intensive.
- Euclidean distance metric

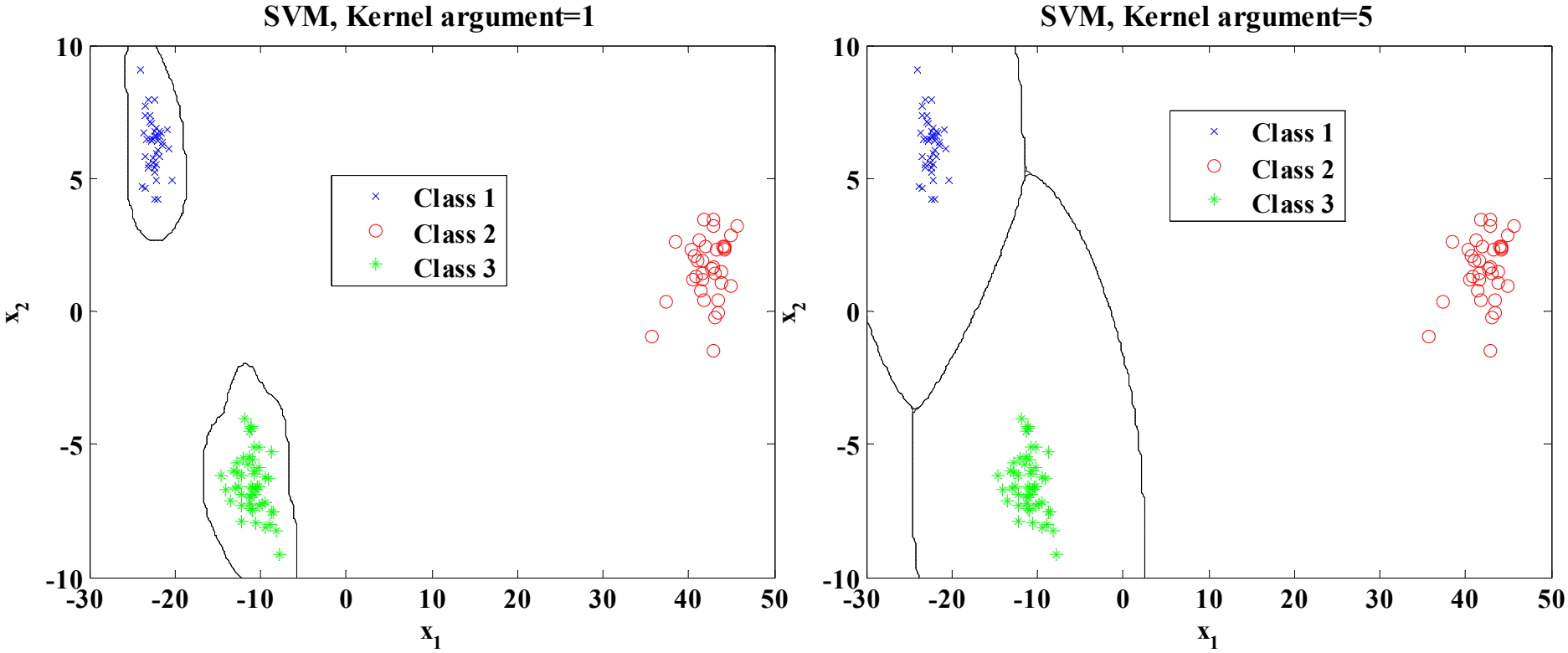
Support Vector Machines (SVM)

- Optimal separating hyperplane – find a hyperplane with minimum misclassification rate
- Non-linear SVM
  - Perform a non-linear mapping of the feature vector x onto a high-dimensional space
  - Construct an optimal separating hyperplane in the high-dimensional space
- Tradeoff – margin and capacity





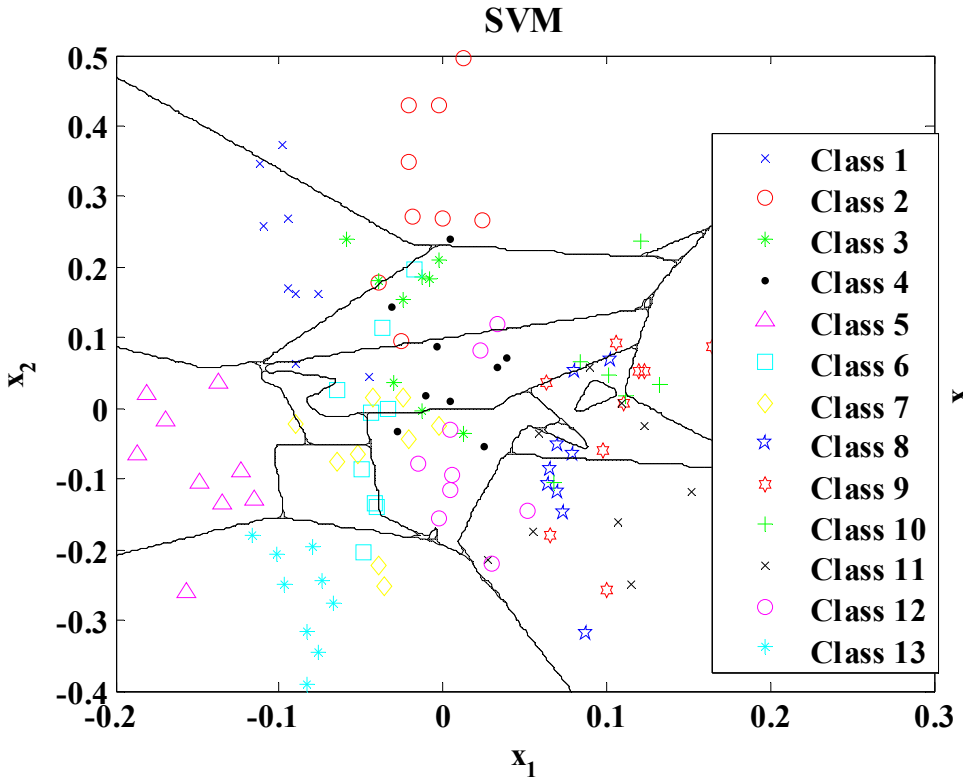
Visualization of SVM

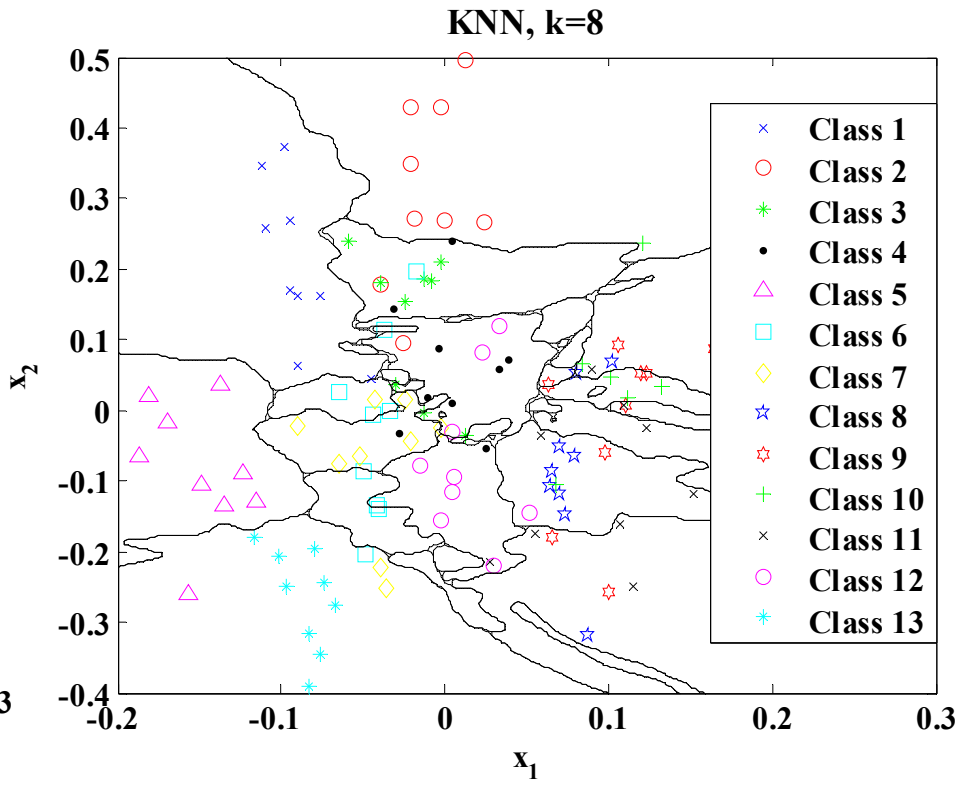


- Sequential Minimal Optimization (SMO) is applied to solve the optimization problem.
- Large kernel argument reduces misclassification errors but lowers discrimination ability.

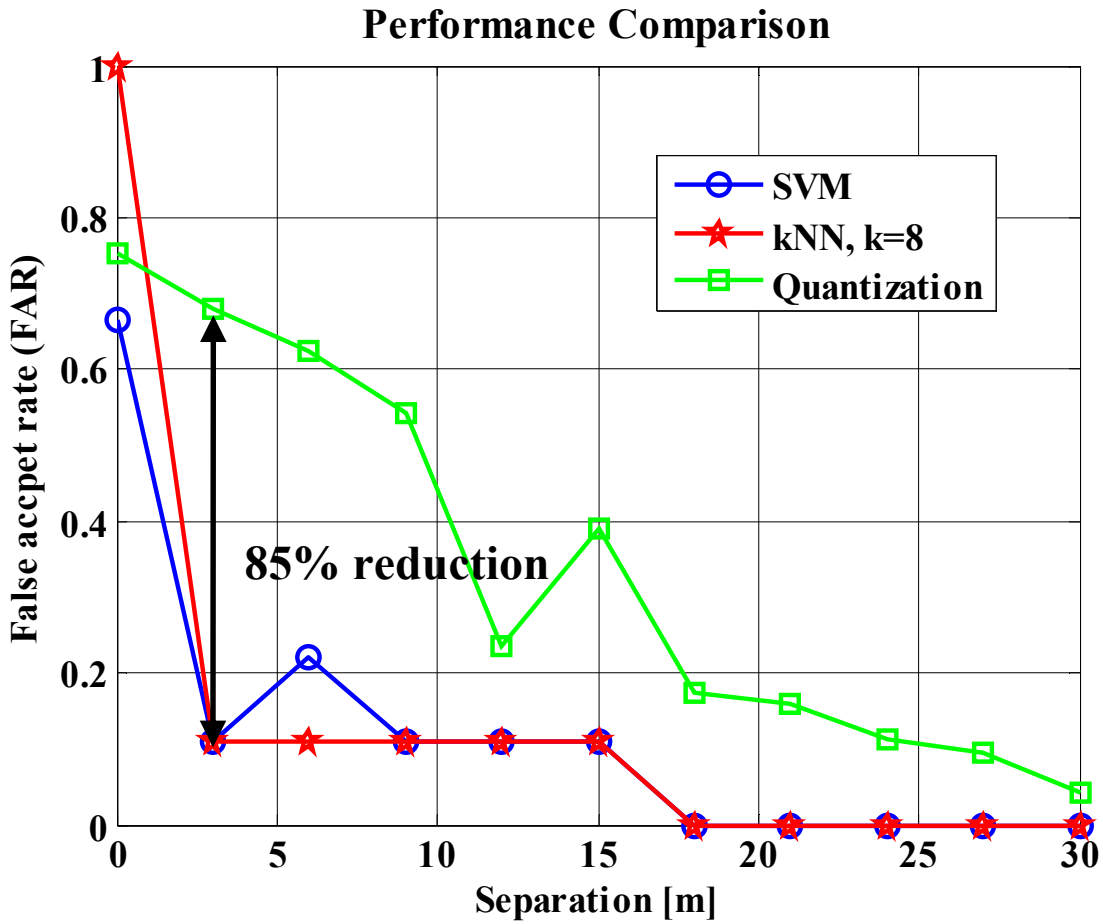
Data set to evaluate spatial discrimination

Classifier visualization





Spatial discrimination comparison



Conclusion

- Modeled geo-security and standardized the performance evaluation.
  - Geotag reproducibility and spatial discriminiation
- Developed fuzzy extractors to reduce continuity risks.
  - Euclidean metric for noise, seasonal bias, and quantization error
  - Hamming metric for offline transmitters
- Applied pattern classification for geotag generation to improve spatial discrimination.
  - kNN and SVM

Acknowledgment

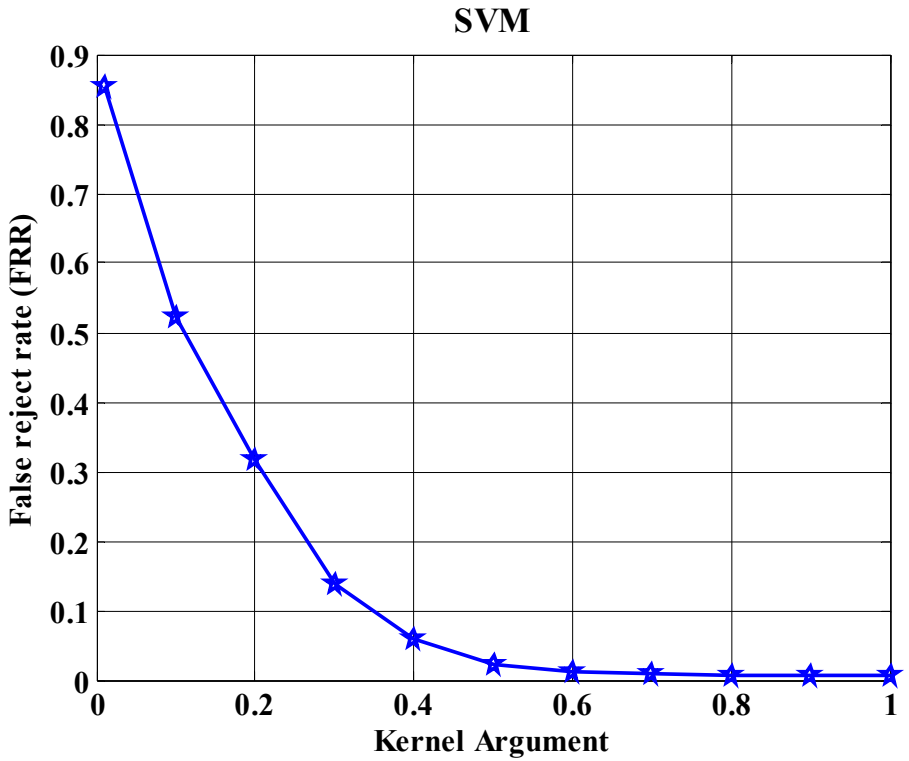
The authors would like to thank Mitch Narins of the FAA, Dr. Greg Johnson and Ruslan Shalaev at Alion Science & Technology and USCG LSU for their support and help during the research.

Thank you!

Questions?

Backup slides

Geotag reproducibility



- 90-day seasonal monitor data
- Same hypothesis problem
- Tradeoff by varying kernel argument