RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Wireless Location Privacy: Depersonalization Techniques and Connected Vehicle Applications

Marco Gruteser
WINLAB, Electrical and Computer Engineering Department

The New York Times

September 1, 2009

EDITORIAL OBSERVER

A Casualty of the Technology Revolution: 'Locational Privacy'

By ADAM COHEN

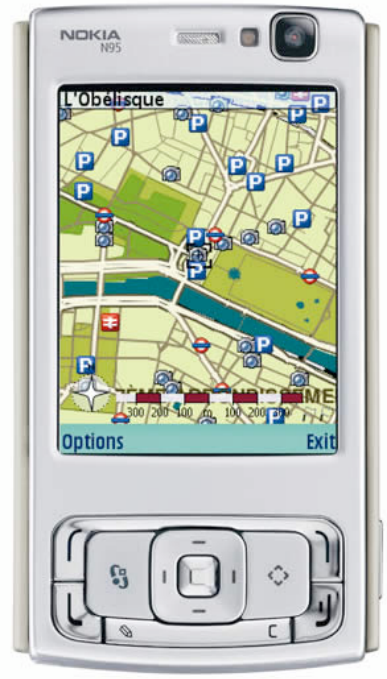
When I woke up the other day, I went straight to my computer to catch up on the news and read e-mail. About 20 minutes later, I walked half a block to the gym, where I exercised for 45 minutes. I took the C train to The New York Times building, and then at the end of the day, I was back on the C train. I had dinner on my friends Elisabeth and Dan's rooftop, then walked home seven blocks.

I'm not giving away any secrets here — nothing I did was secret to begin with. Verizon online knows when I logged on, and New York Sports Club knows when I swiped my membership card. The M.T.A. could trace (through the MetroCard I bought with a credit card) when and where I took the subway, and The Times knows when I used my ID to enter the building. AT&T could follow me along the way through my iPhone.

Controversy about Human Mobility Article

- Analyzed human mobility patterns based on cell phone handoff data
  - 100,000 users for 6 months
  - Cell tower location recorded for each call / msg
  - Avg tower covered 3 km^2
- Data was obtained outside US, country not specified
  - Phone identifiers were replaced with pseudonyms

Location Analytics Trends



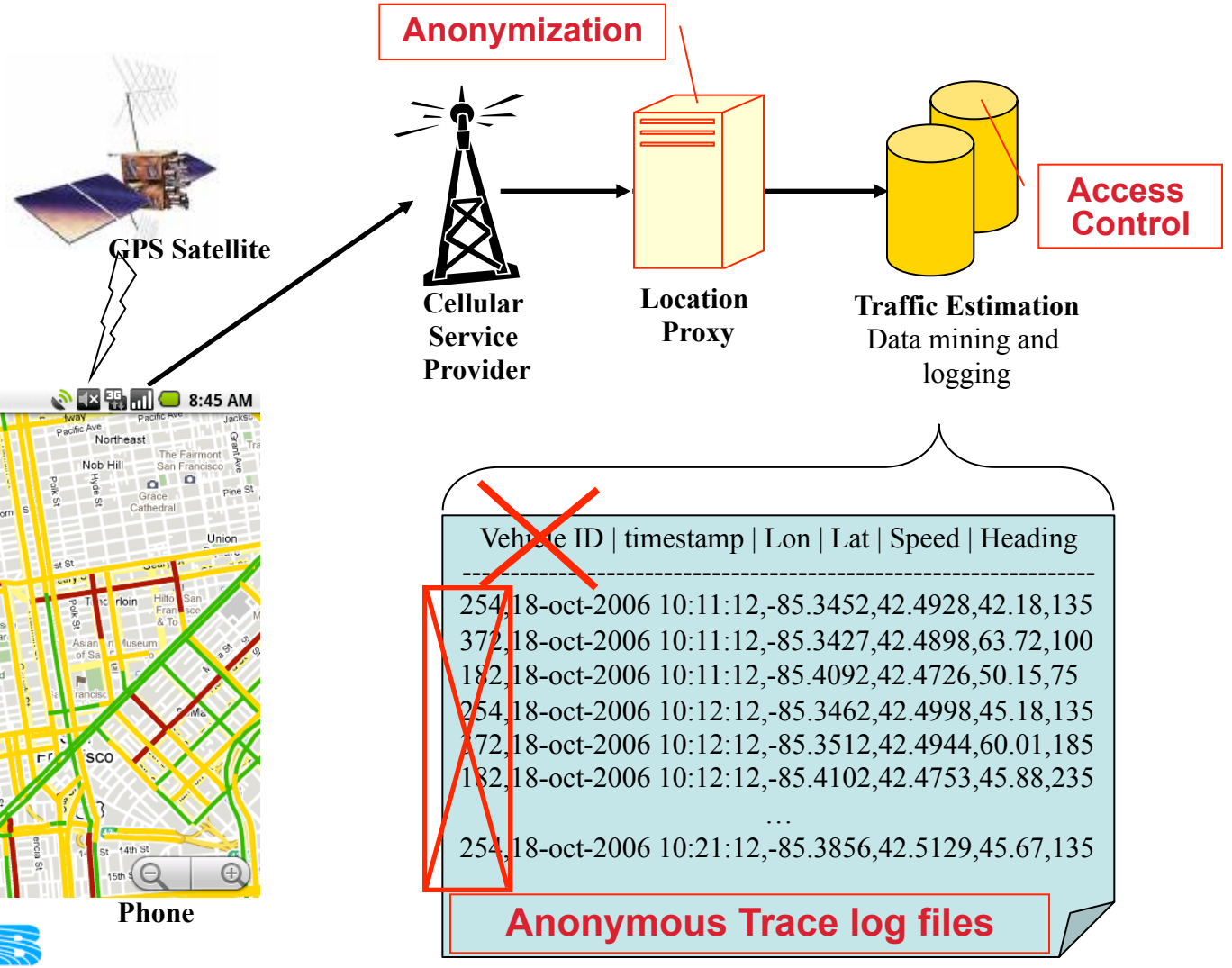
Google™

AIRSAGE

Sense Networks

INRIX

Depersonalizing Location Data



WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

Fair Information Principles

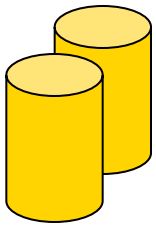
- Notice / Awareness
  - Purpose specification
- Choice & Consent
- Access/Participation
- Integrity/Security
- Enforcement/Redress

When collecting personally identifiable information

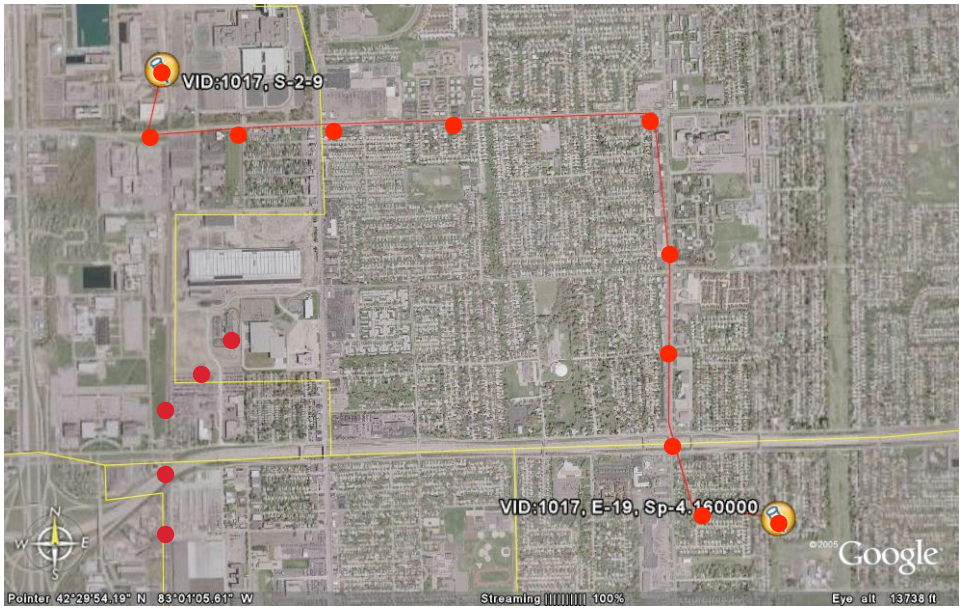
Inference/Insider Attacks Compromise Location Privacy

Still insider breaches and remote break-ins possible

Re-identification of traces through data analysis



Anonymous Trace log files



Tracking algorithms recover individual trace [Hoh05] (Median trip time only 15min)

Home Identification [Hoh06]

GPS often precise enough to identify home

WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

De-identification in the medical community

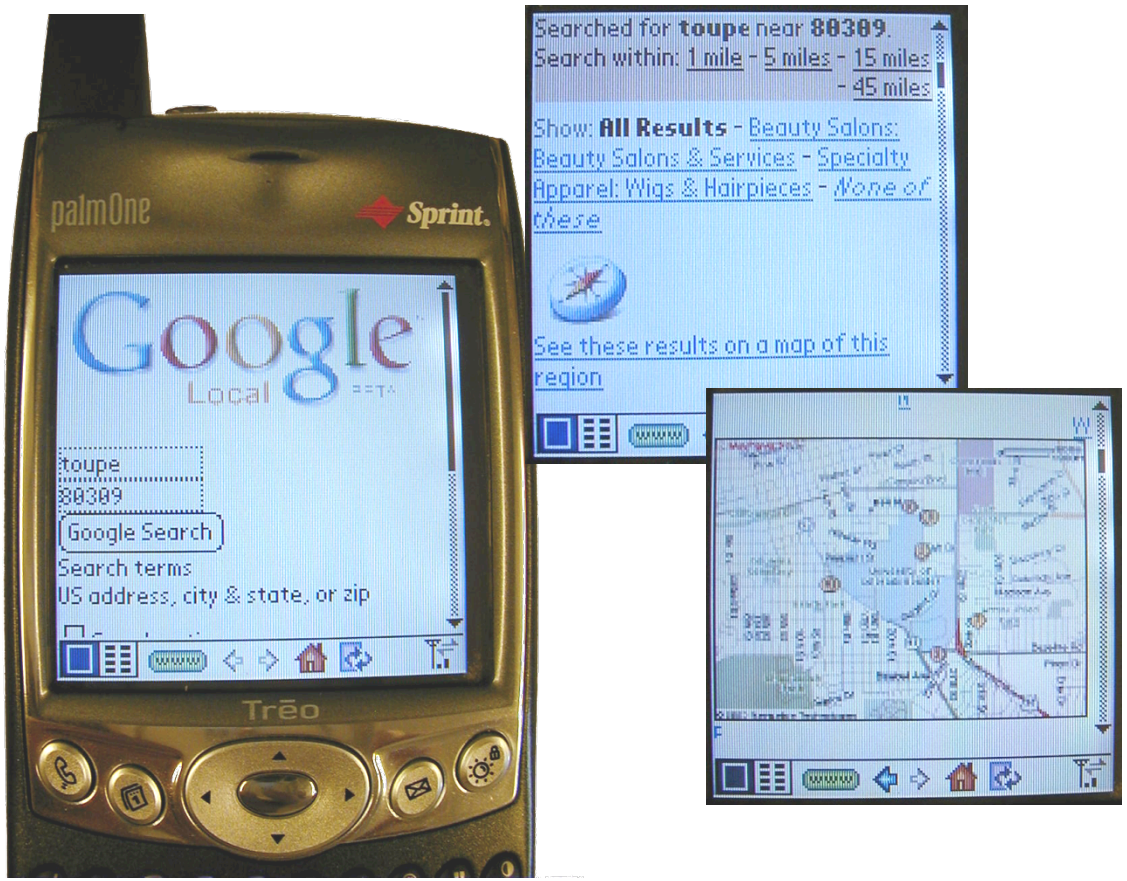
- Health Insurance Portability and Accountability Act (HIPAA) defines accepted procedures for de-personalization of health records
- Mandates removal of 18 specified (pseudo)identifiers, e.g.
  - Names
  - Geographic subdivision smaller than a state (city, street, etc.)
  - Any elements of dates related to an individual other than a year
  - Any age greater than 89 years
  - Account numbers
  - IP addresses
  - Biometric identifiers
- Custom techniques allowed if vetted by a qualified statistician

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Spatial and Temporal Cloaking

(MobiSys 2003)

Originally Developed for Sporadic Location-Based Queries



Location Identification Risks

Sensitive query, from driveway

[515110X 4300483Y 13Z]

Aerial imagery
(e.g. Goole Earth)

Geocoded Address Database (TIGER/LINE):

John Doe
1234 Main St
Anywhere, US
(515110X 4300483Y, 13Z)

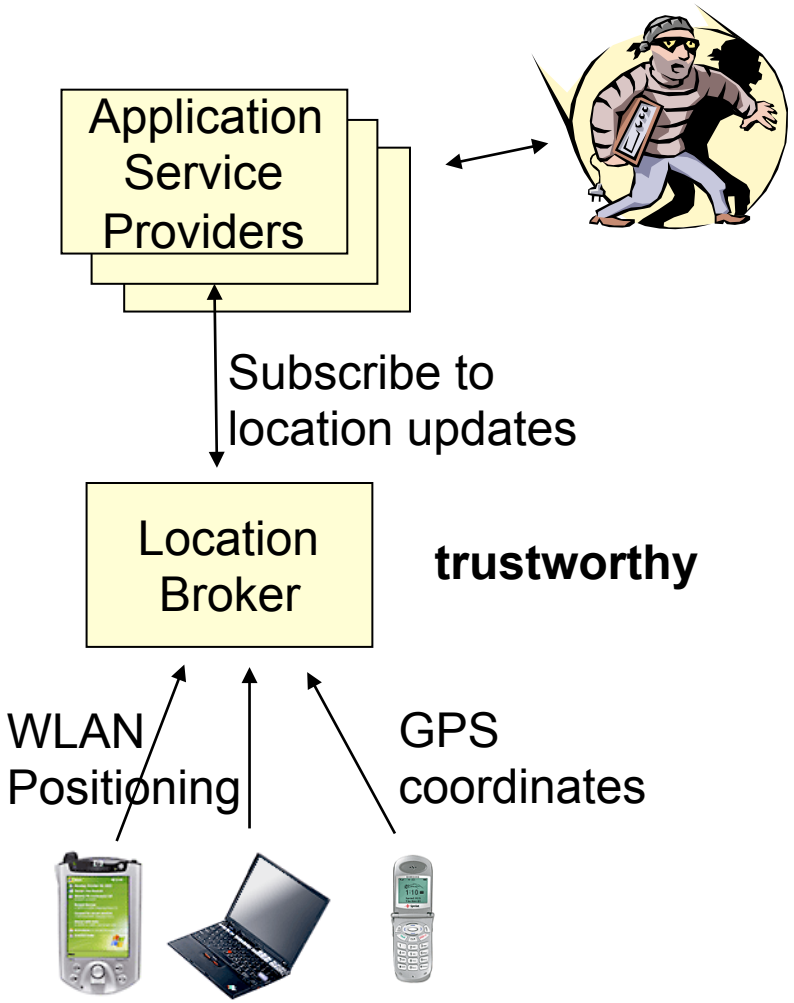
Correlation Attacks

- Automatic Toll Booth
- Customer Loyalty Cards
- Records from other location-based services …

WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

Assumptions

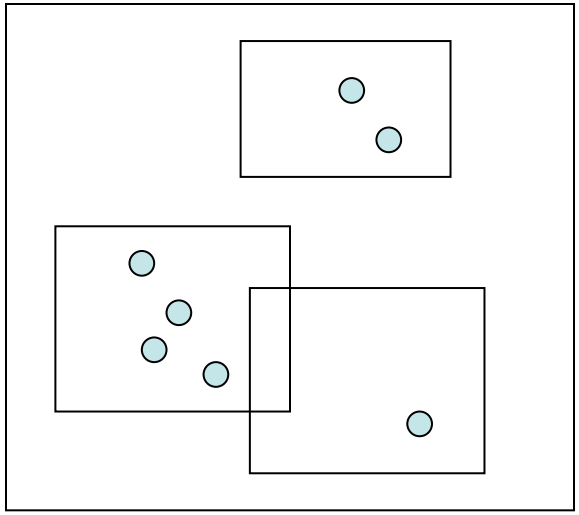
- How can we share sensitive location information with untrusted applications?
- Assumptions
  - Location broker trustworthy
  - Location broker and client secure, adversary has only access to application data
  - Location broker knows positions of all users
  - Adversary knows privacy algorithms
  - Users issue only sporadic queries



Exploiting User Density Information to Reduce Resolution

- Key idea: Reduce granularity of query location at location broker in low density areas
- Why not just use zip codes?
  - Unnecessary loss of precision
- How much should we reduce resolution?
  - Depends on population density, time of day, …

Anonymity Criterion

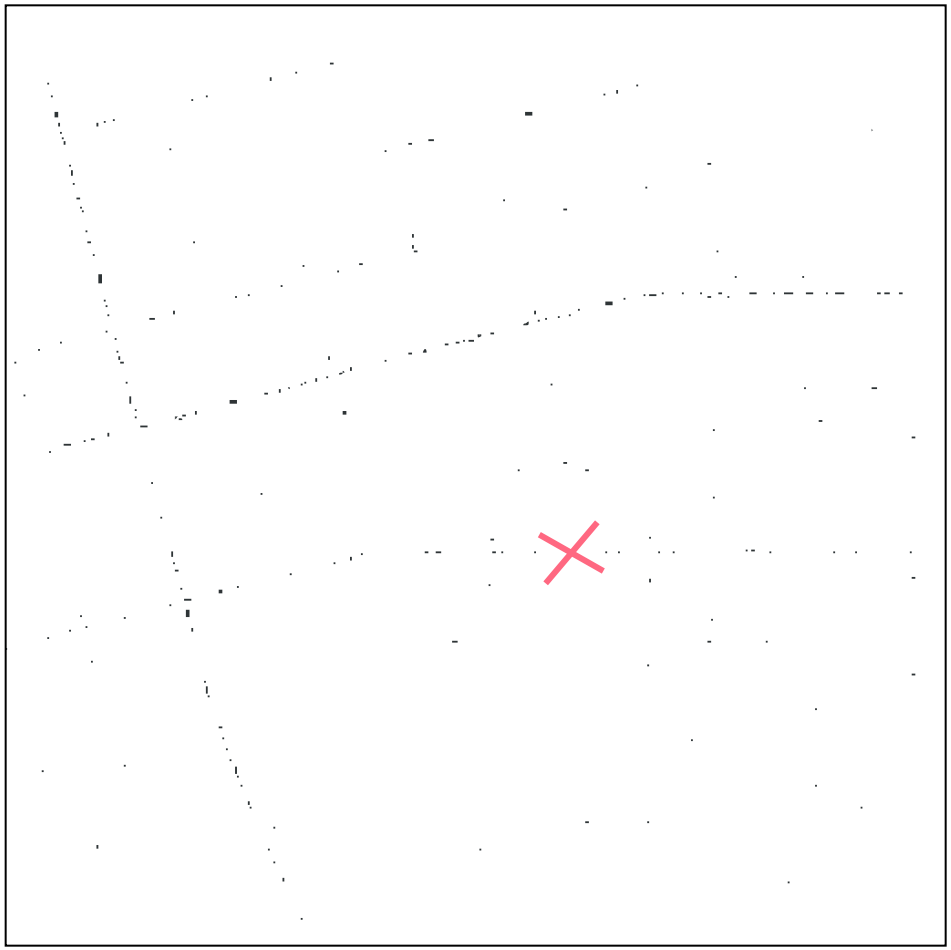


Location represented through spatial and temporal intervals (uncertainty range)
([x1,x2], [y1,y2], [t1,t2])

- k-Anonymity: Area equally likely to be chosen by k-1 other subjects
- Problem: Lower resolution yields more anonymity but reduces service quality

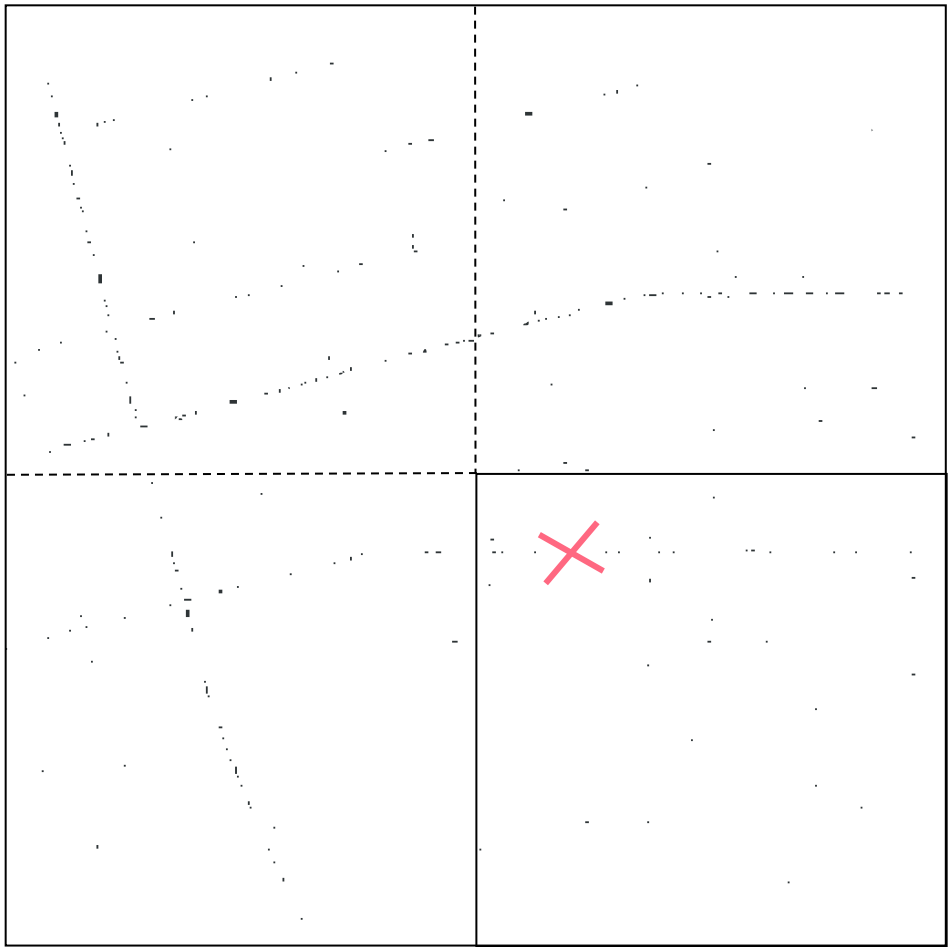
Cloaking Algorithm

Cloaking Algorithm



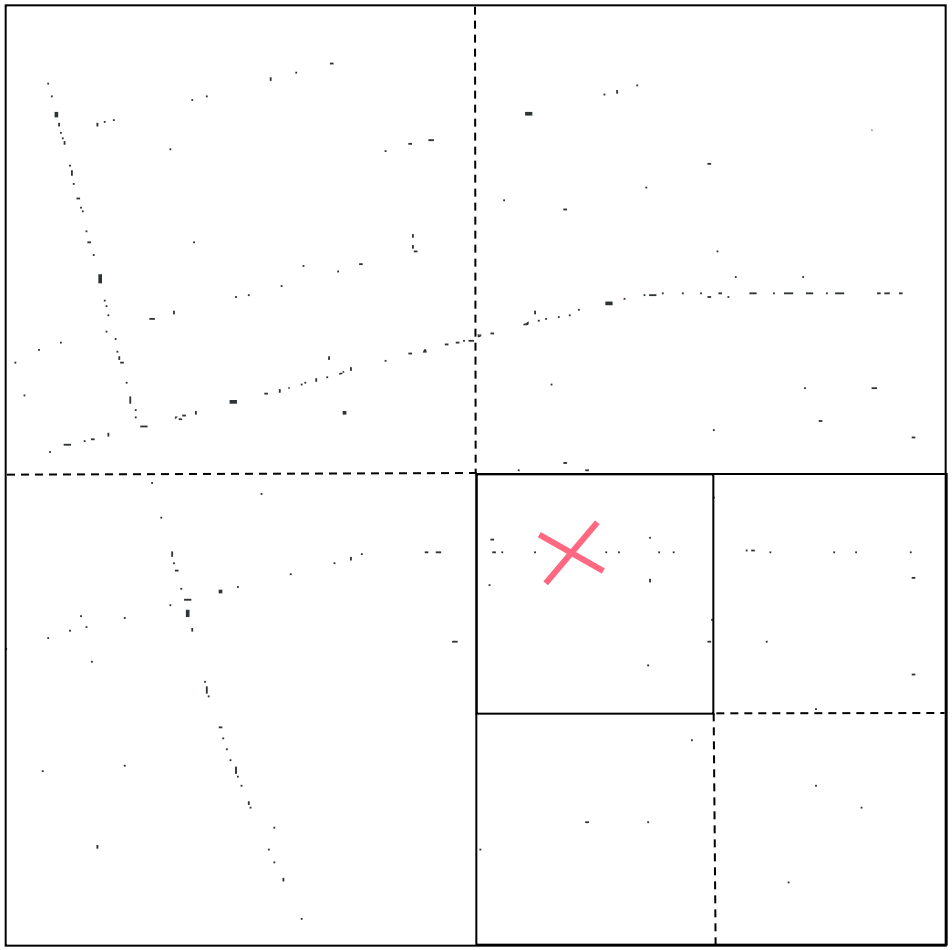
$k_{min}=5$

Cloaking Algorithm



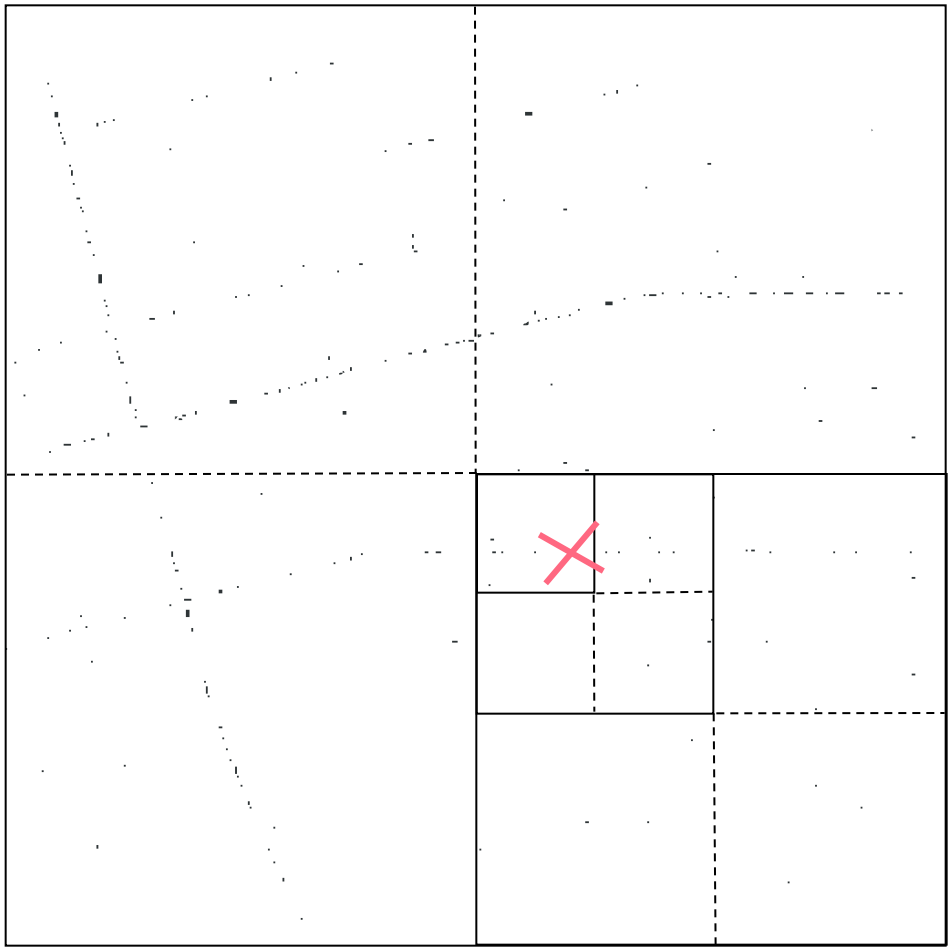
$k_{min}=5$

Cloaking Algorithm



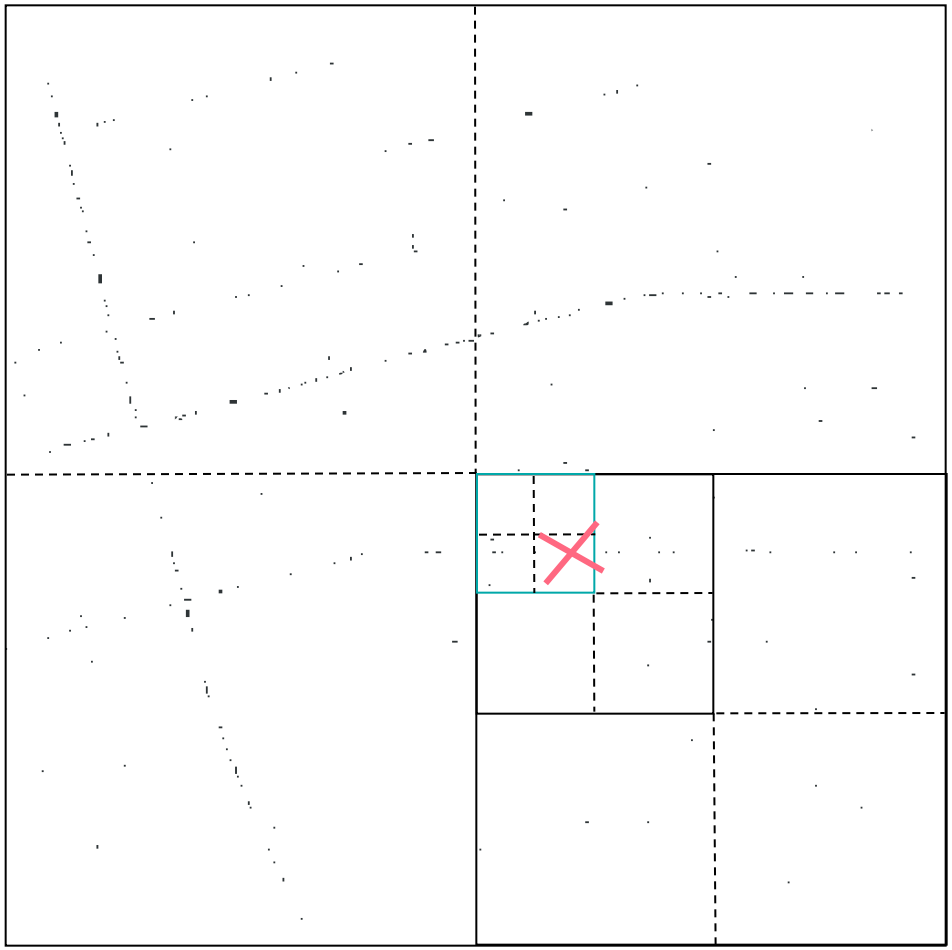
$k_{min}=5$

Cloaking Algorithm



$k_{min}=5$

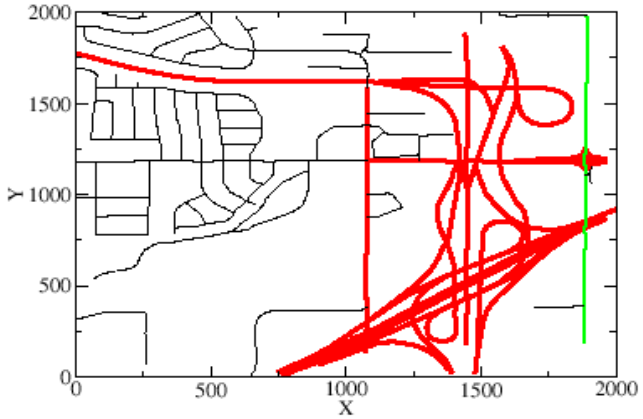
Cloaking Algorithm



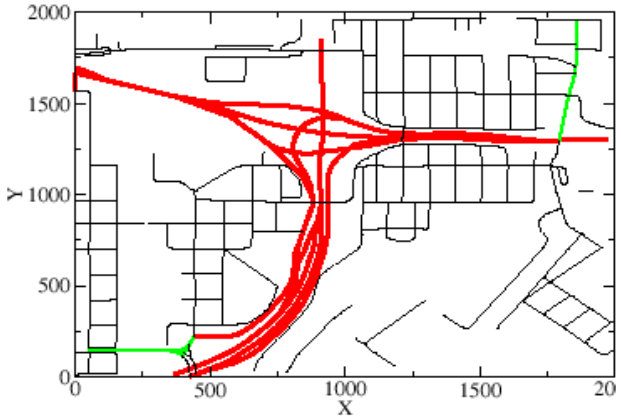
$k_{min}=5$

$k=8$

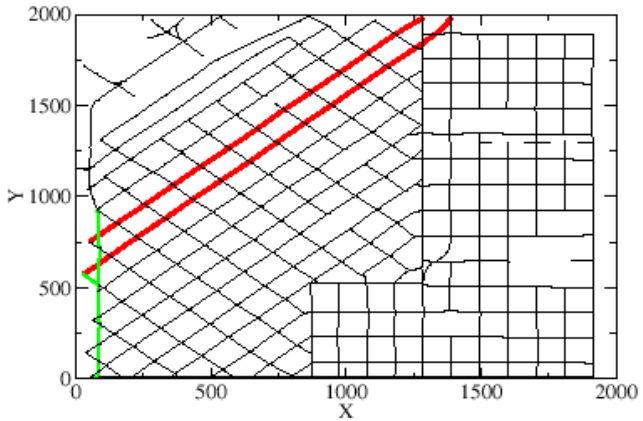
Example Accuracy Evaluation: USGS Road Maps & Traffic Models



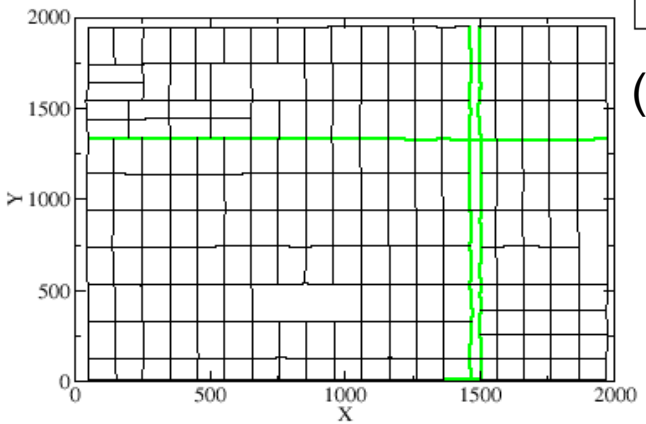
Area 1: Expressways (X:500000, Y:4407000)



Area 2: Expressways (X:500000, Y:4402000)



Area 3: Collectors (X:501000,Y:4400000)



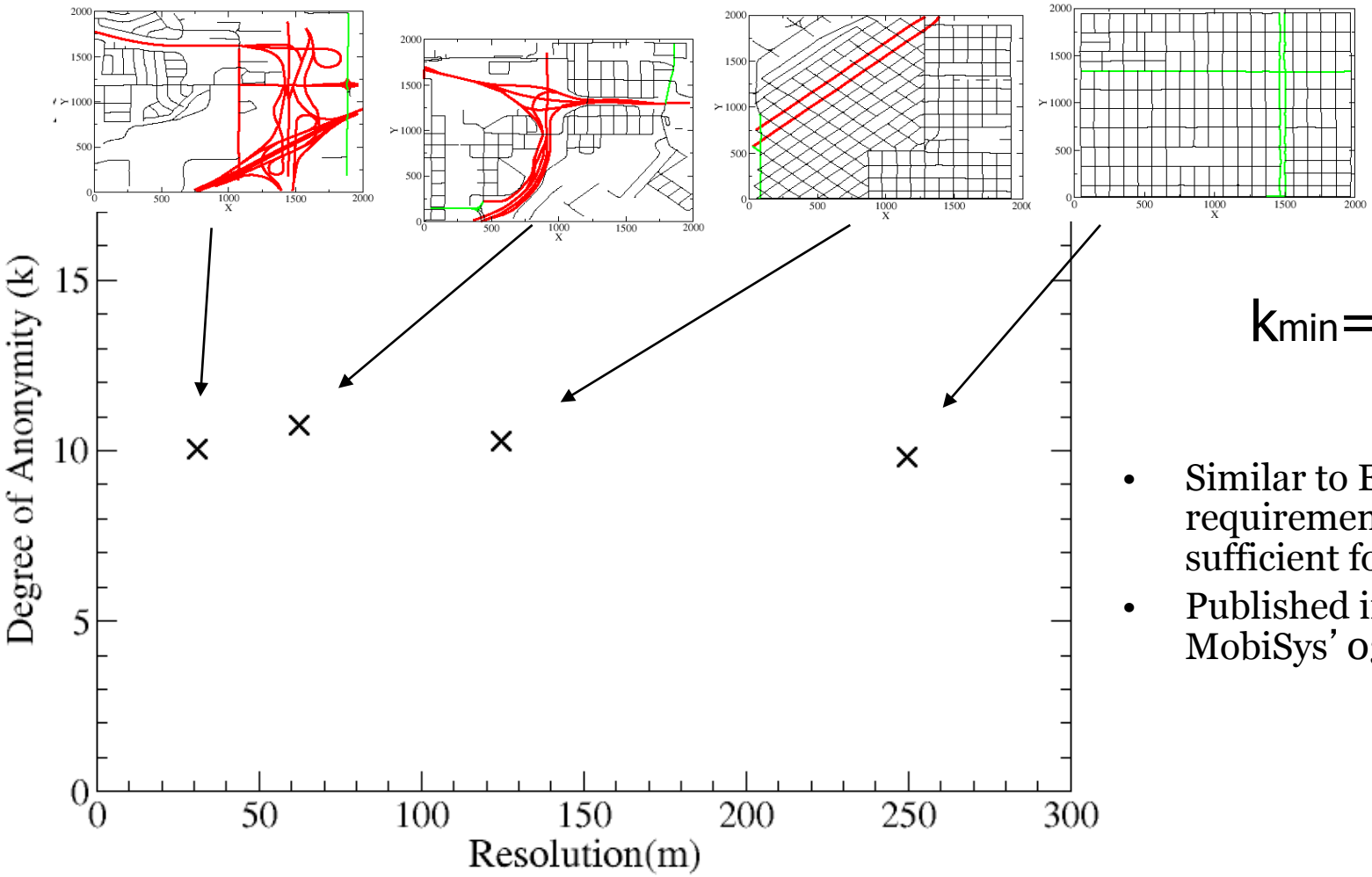
Area 4: Collectors (X:506000,Y:4400000)

USGS Class	Road Type	Traffic Volume
1	Expressway	70000
2	Arterial	22000
3	Collector	6000

(24h bi-directional counts)

WIRELESS INFORMATION NETWORK LABORATORY

Dependency on Area Characteristics



$k_{min}=5$

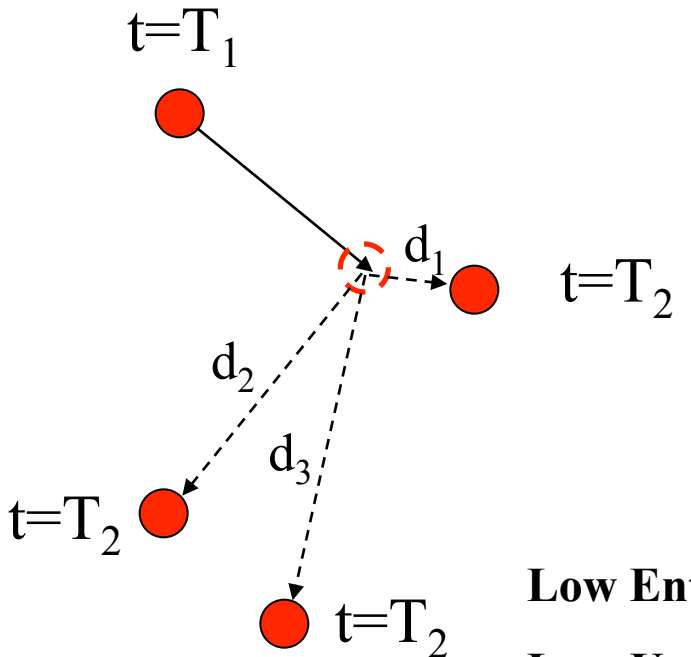
- Similar to E911 requirements, thus sufficient for some LBS
- Published in MobiSys' 03

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Uncertainty-Aware Path Cloaking – A Centralized Algorithm

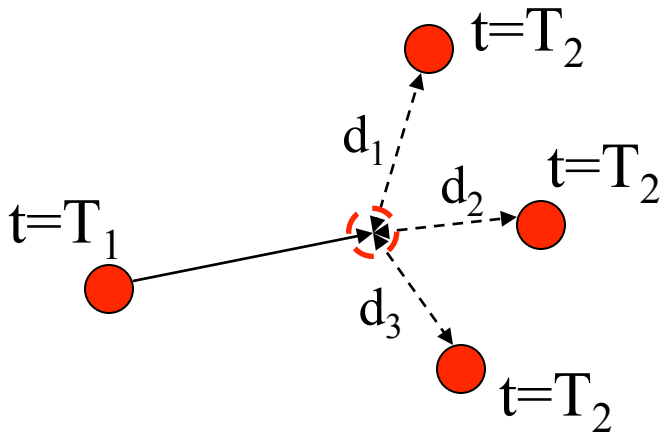
(CCS 2007)

Stronger De-identification for Location Traces: Filtering based on Tracking Model



Low Entropy

Low Uncertainty



High Entropy

High Uncertainty

$$\hat{p}_i = e^{-\frac{d_i}{\mu}}$$

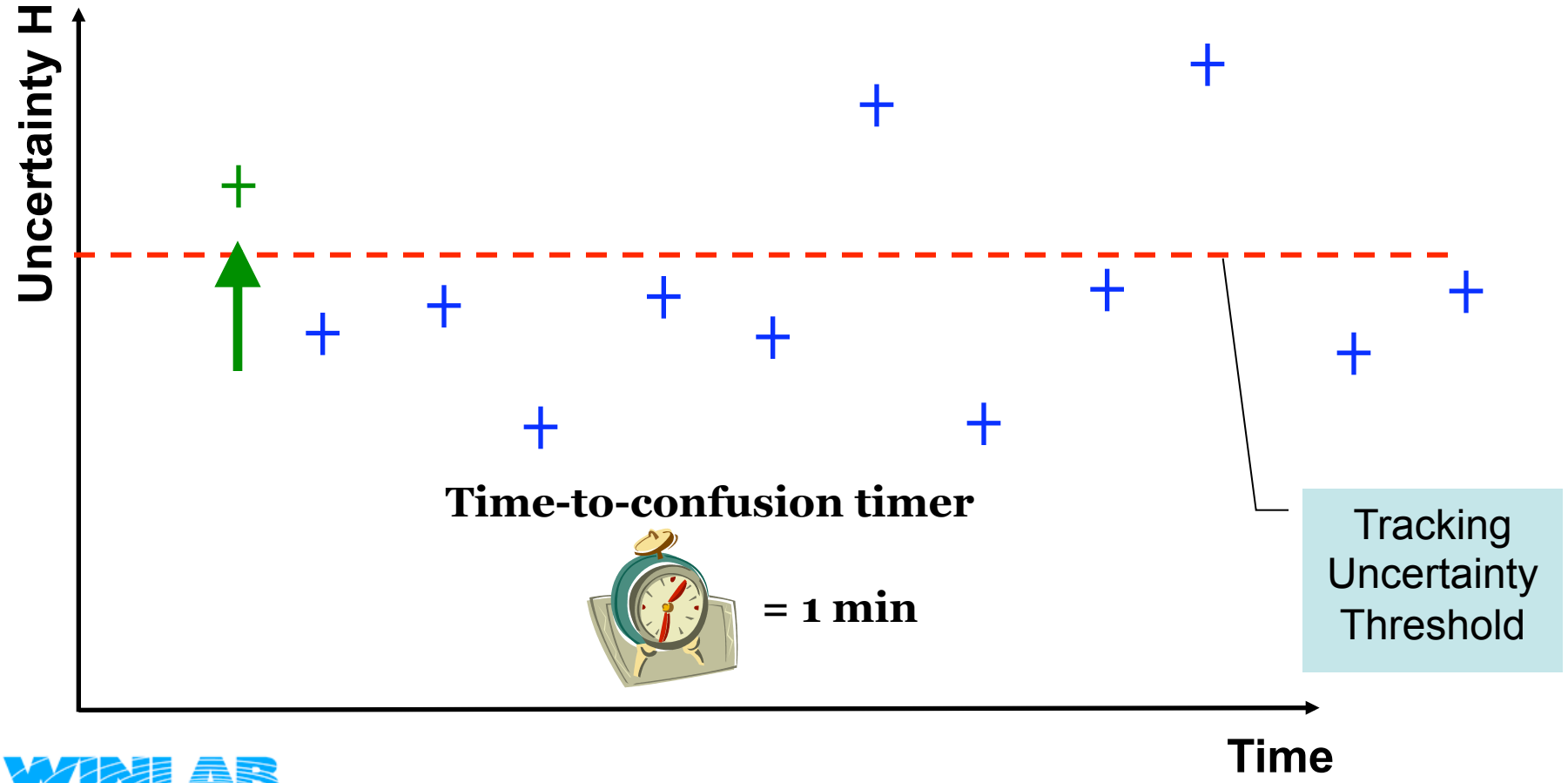
Uncertainty

$$H = -\sum p_i \log p_i$$

Normalization

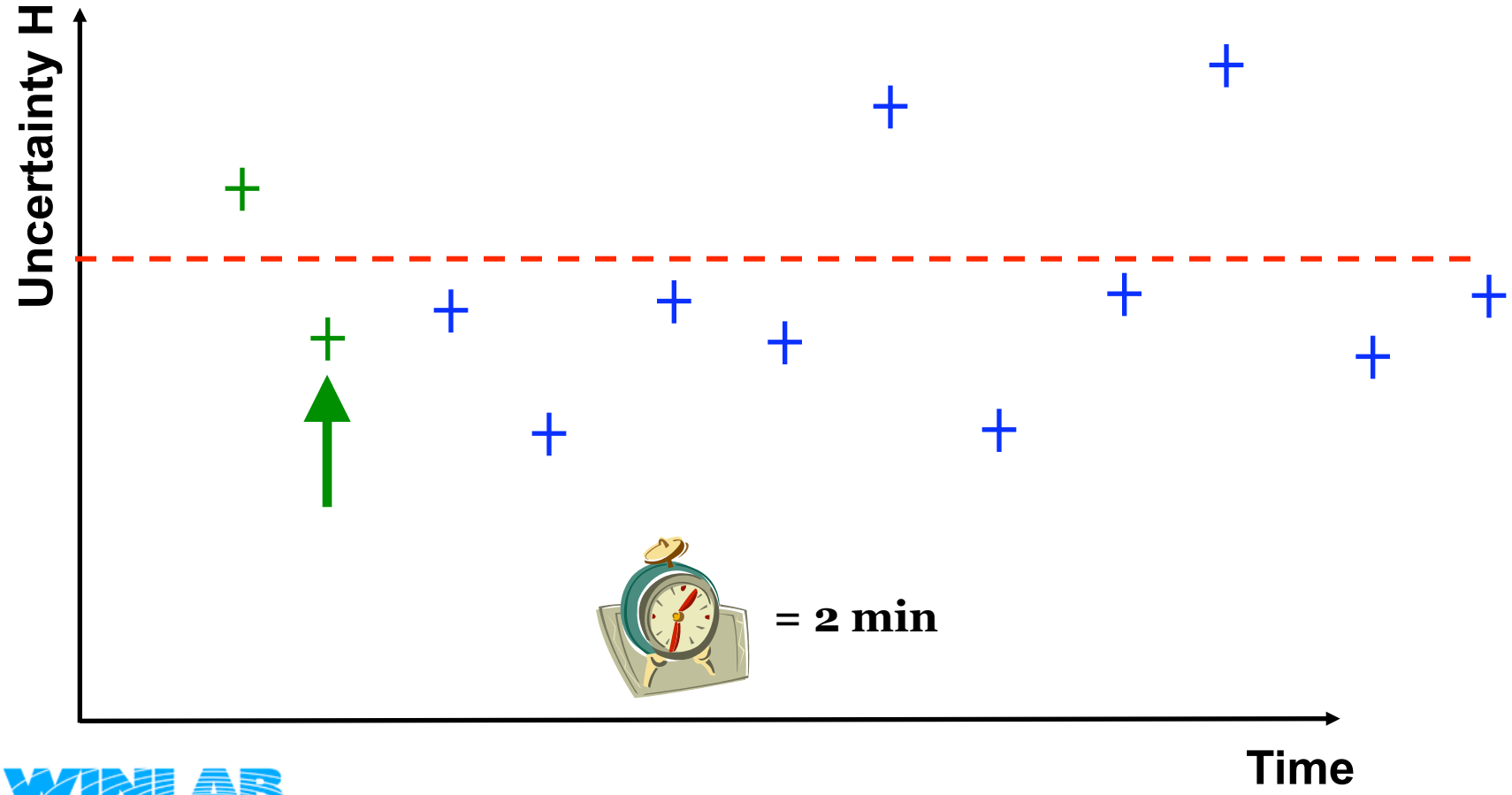
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



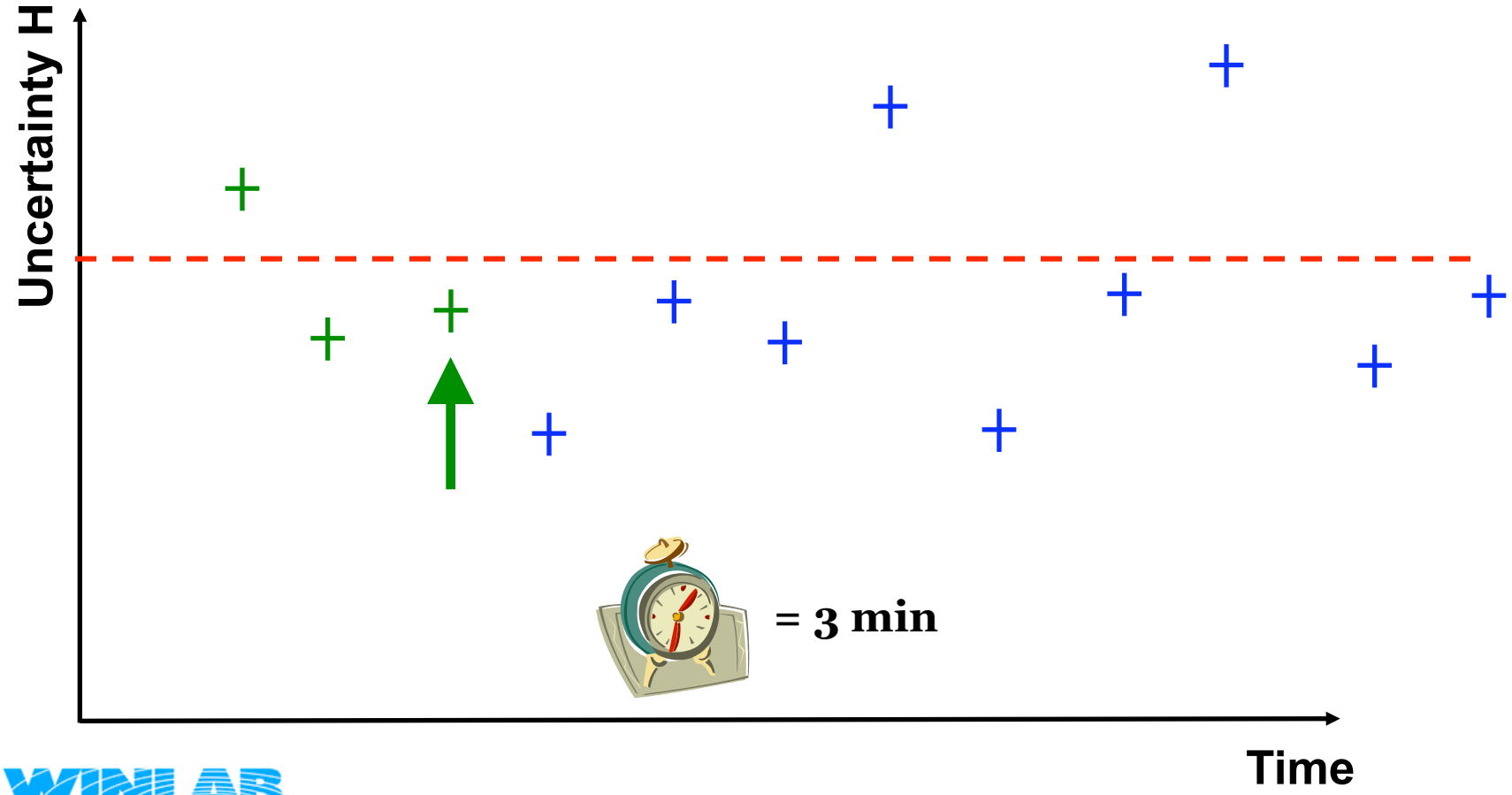
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



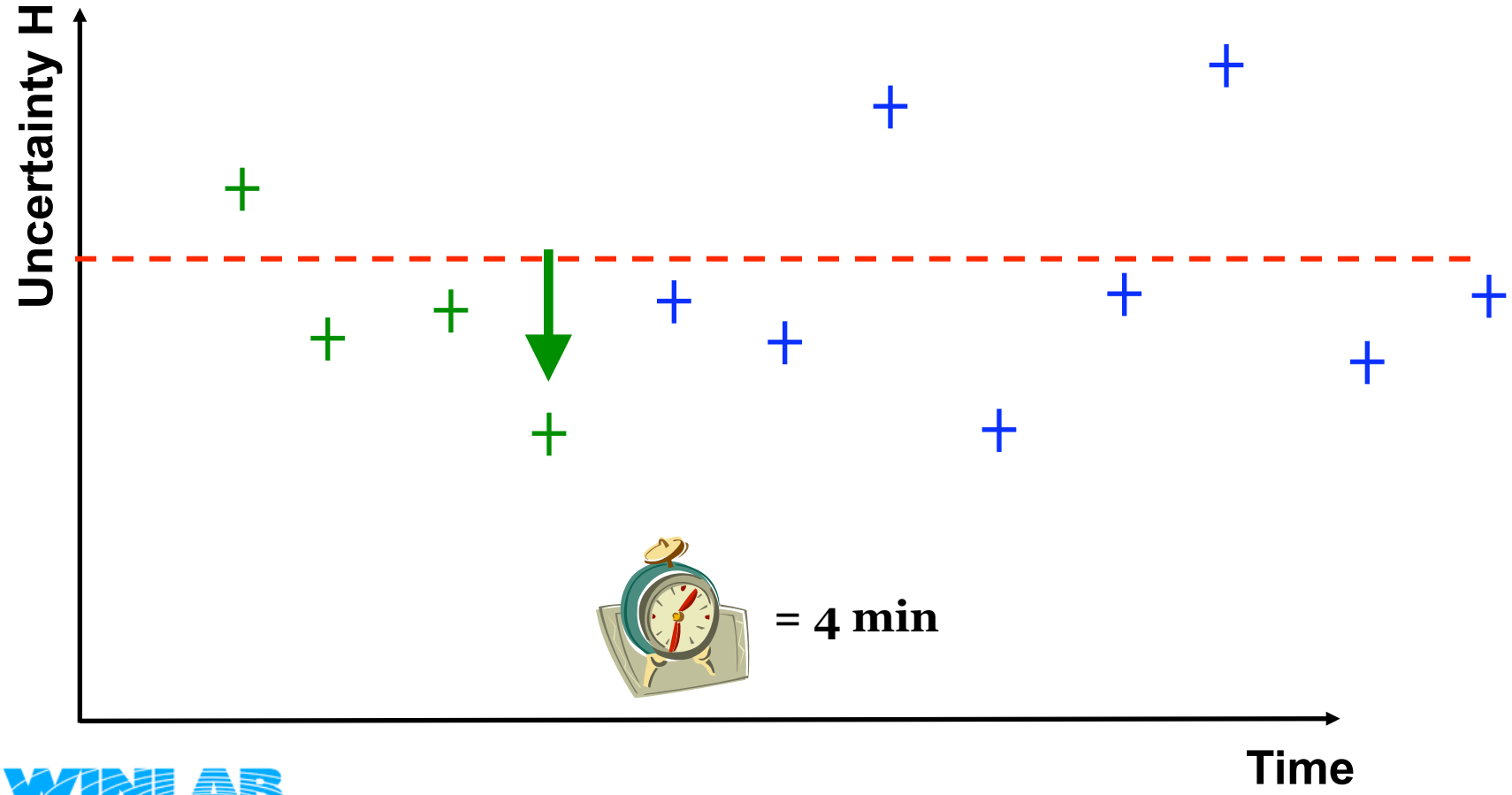
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



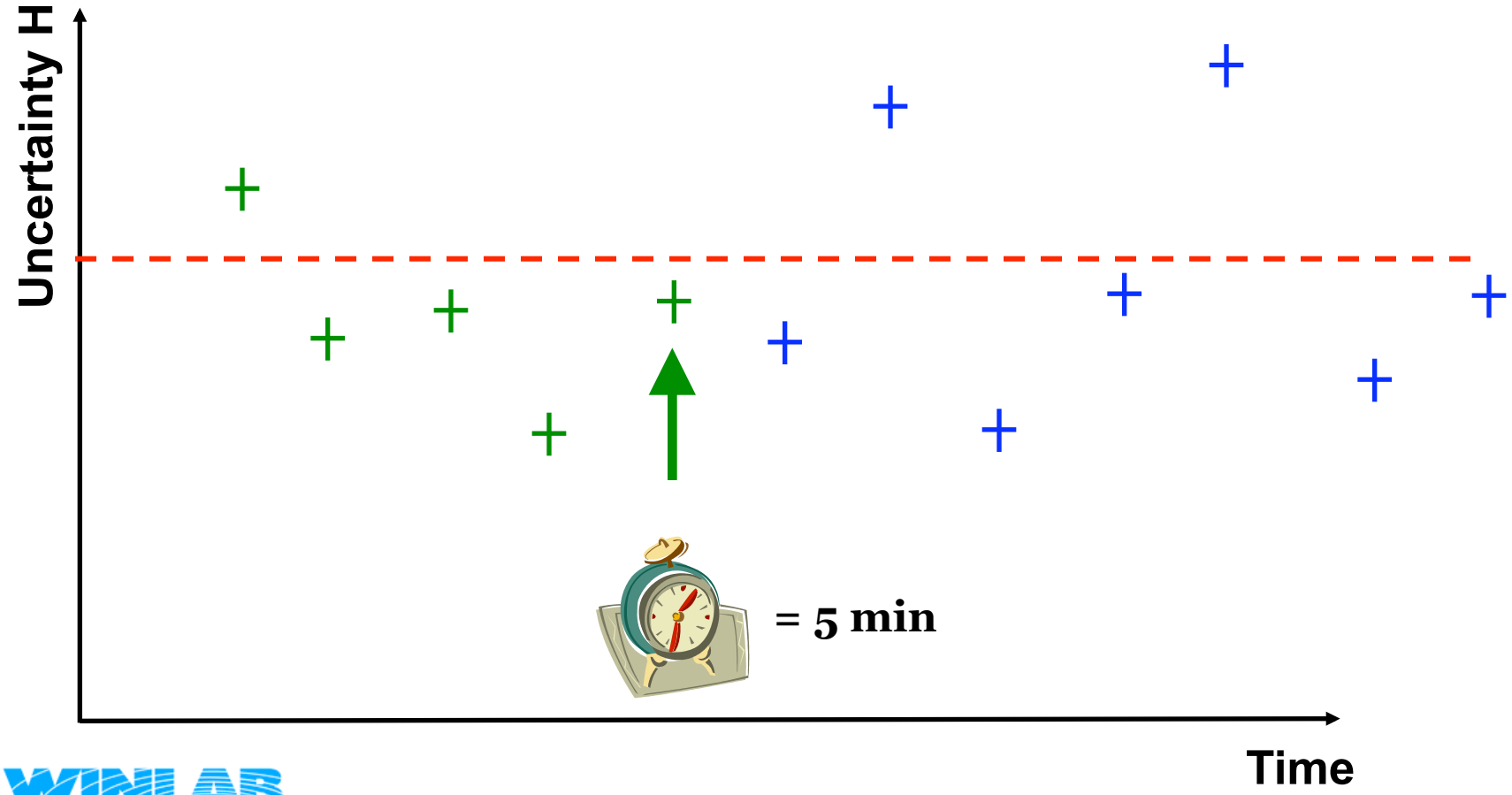
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



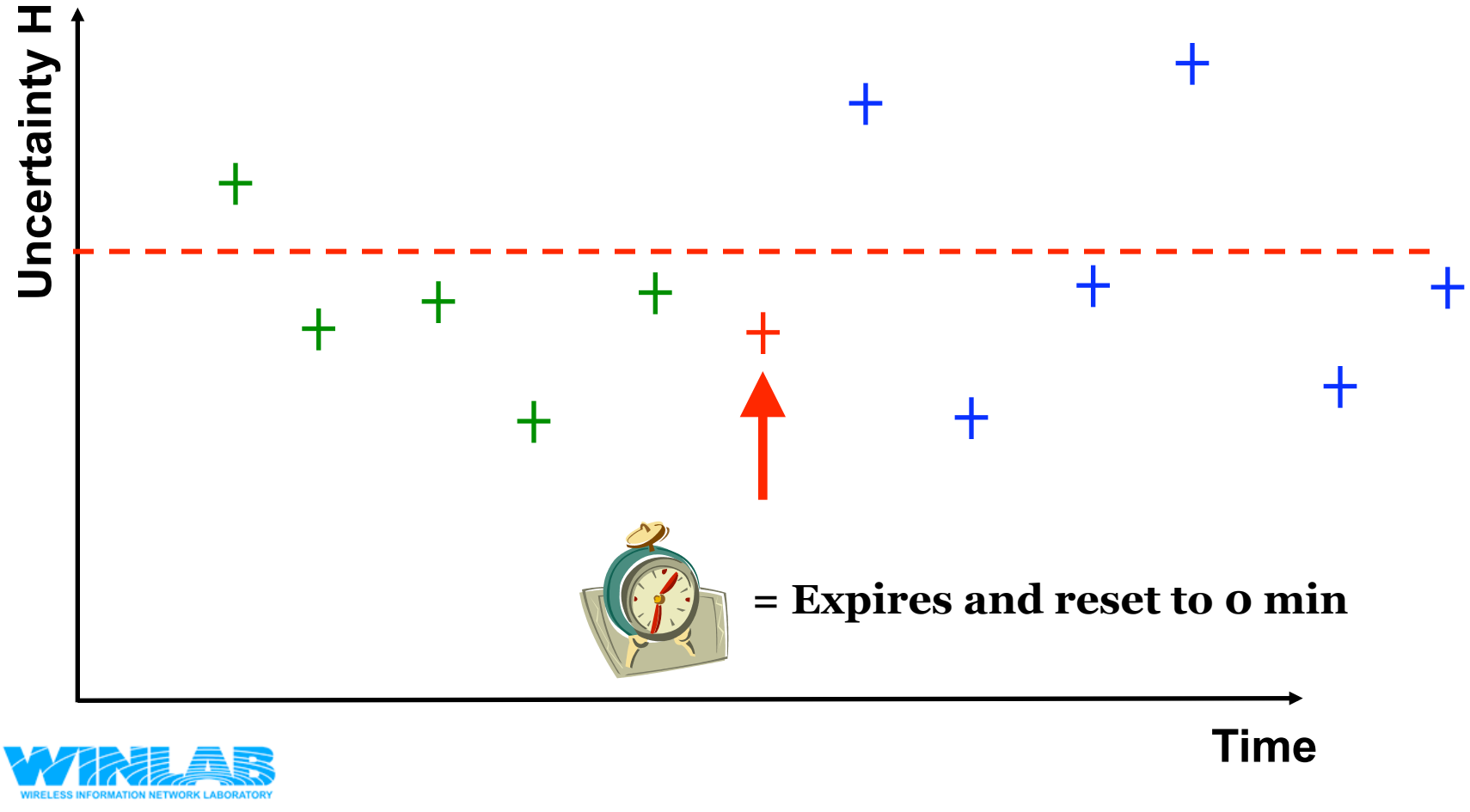
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



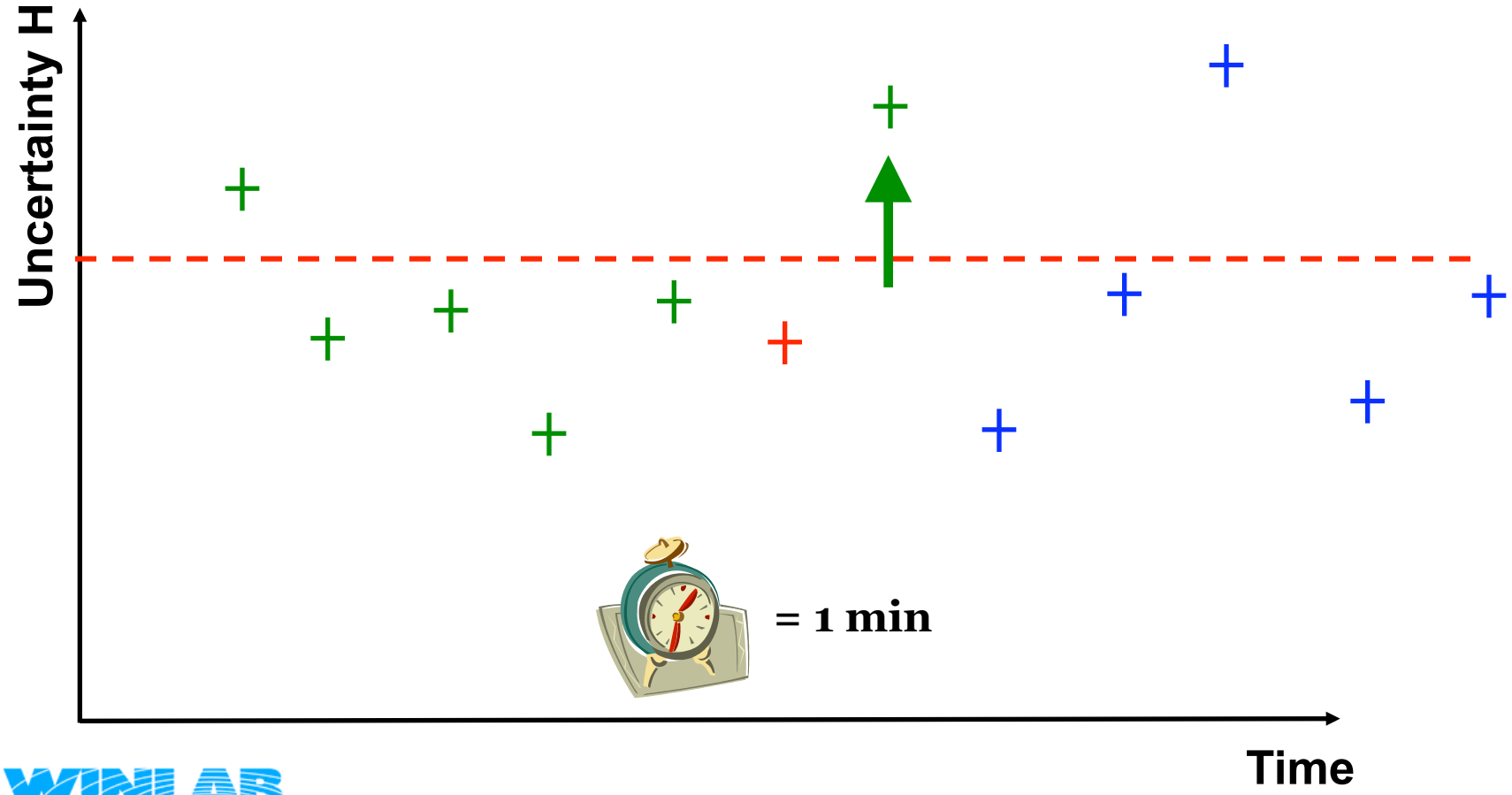
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



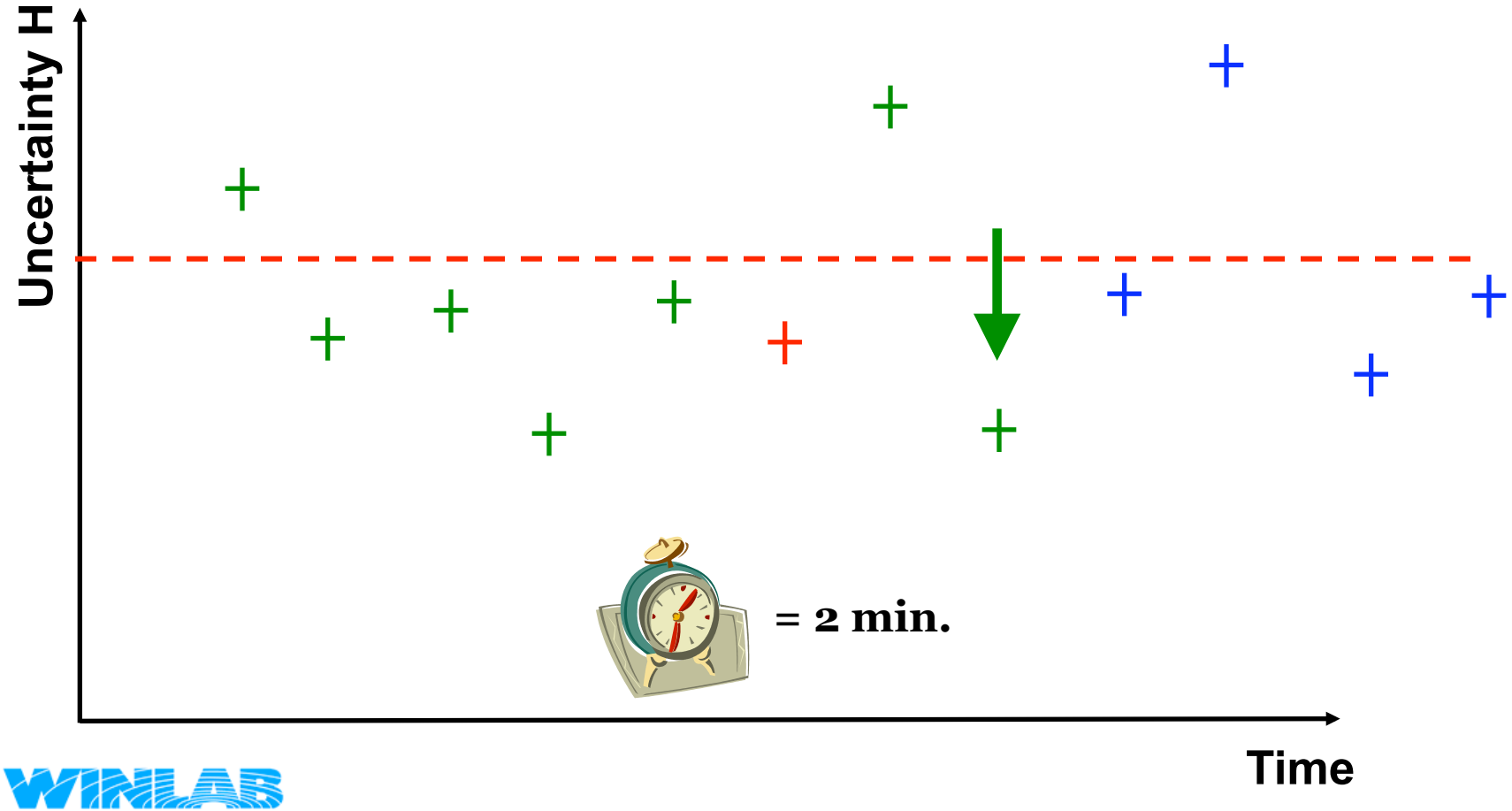
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



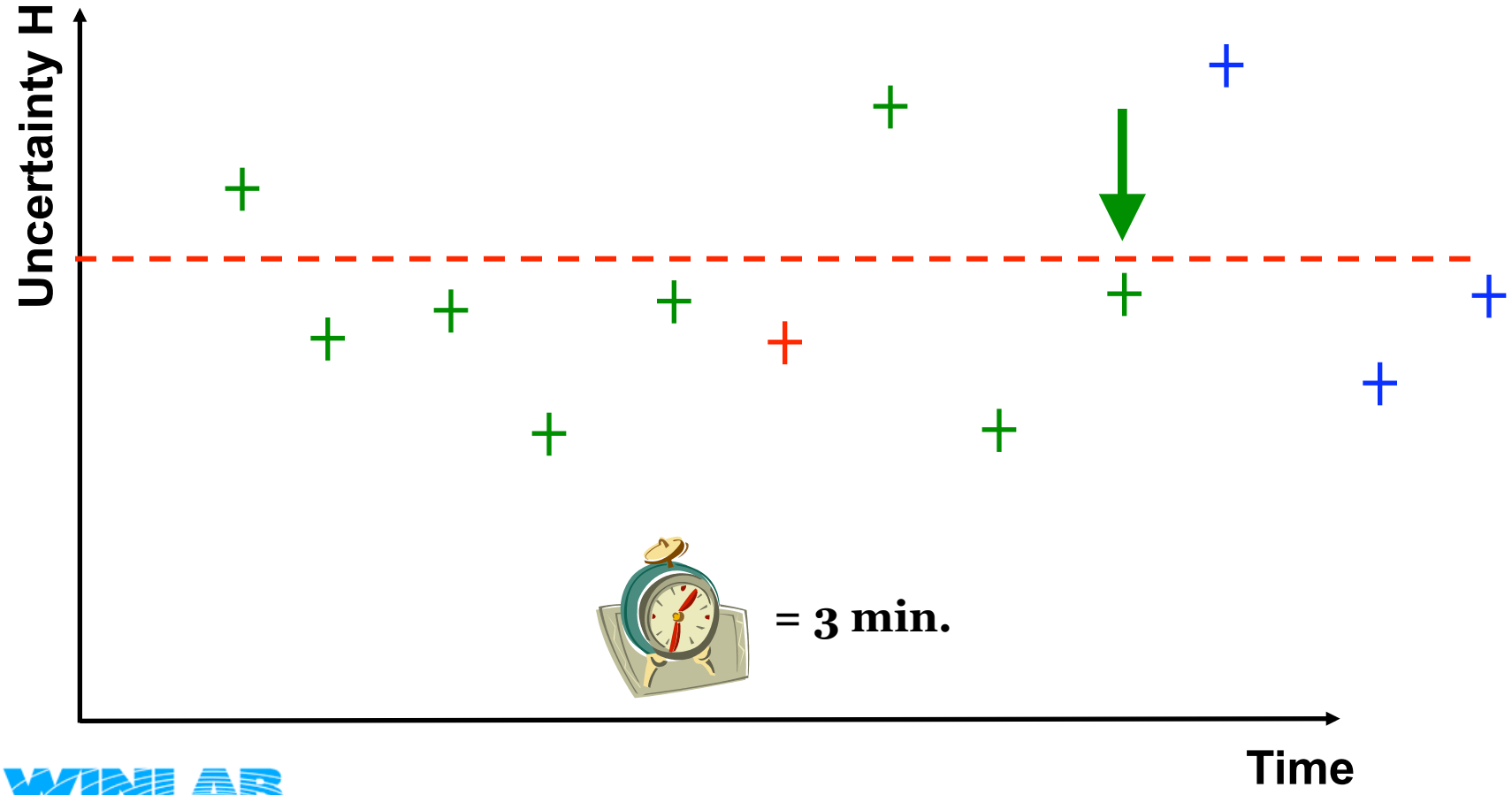
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



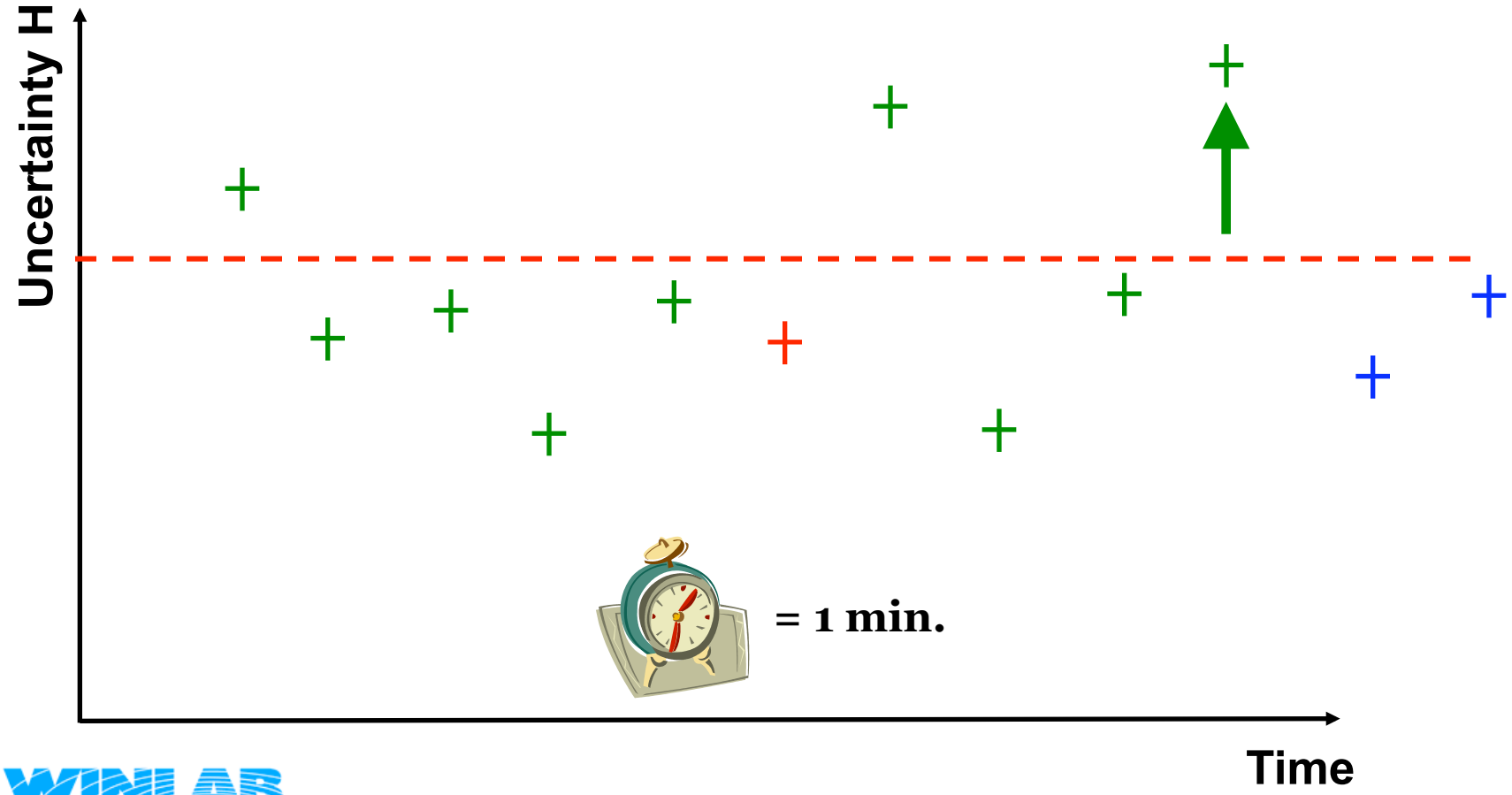
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



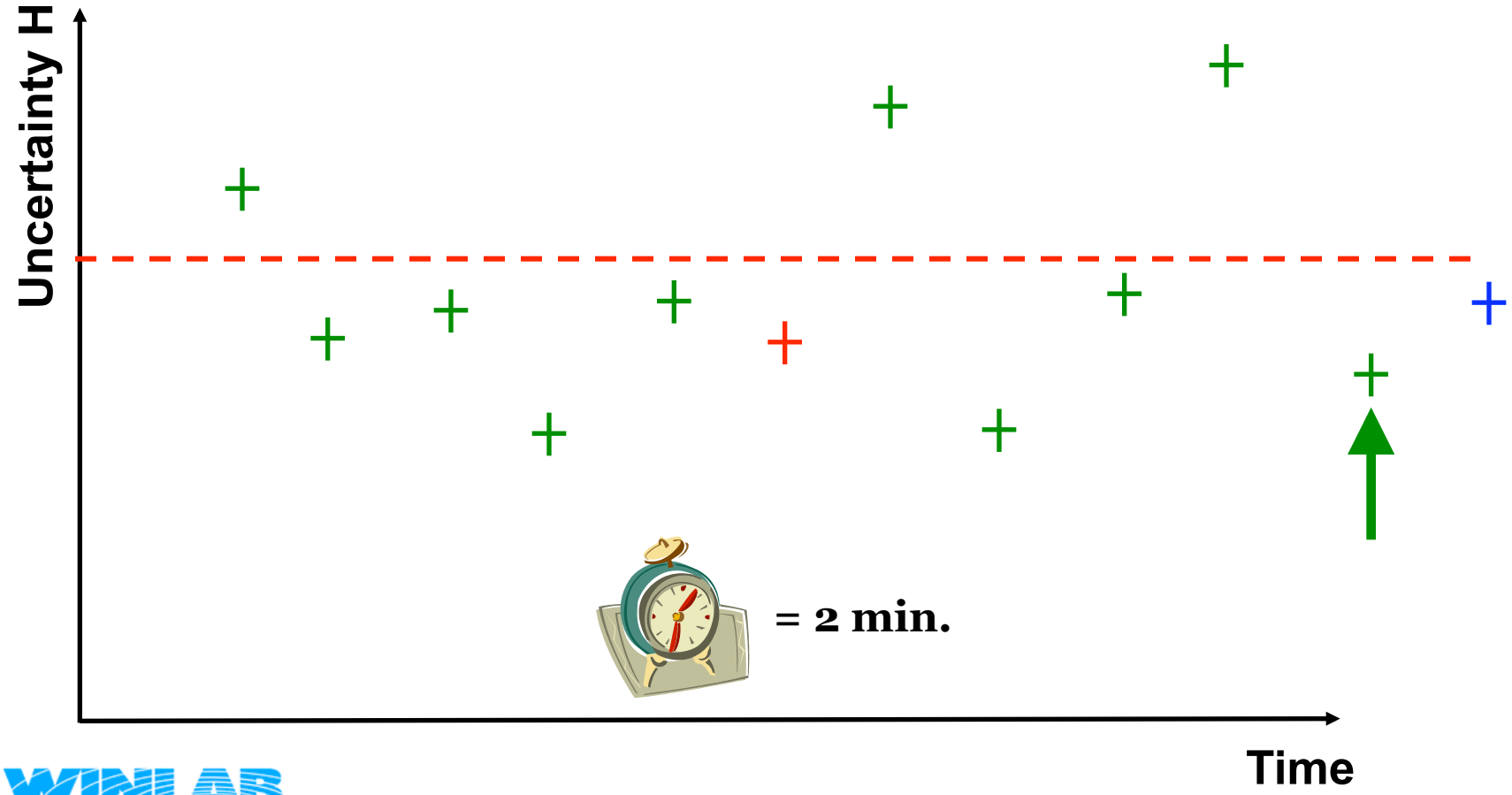
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



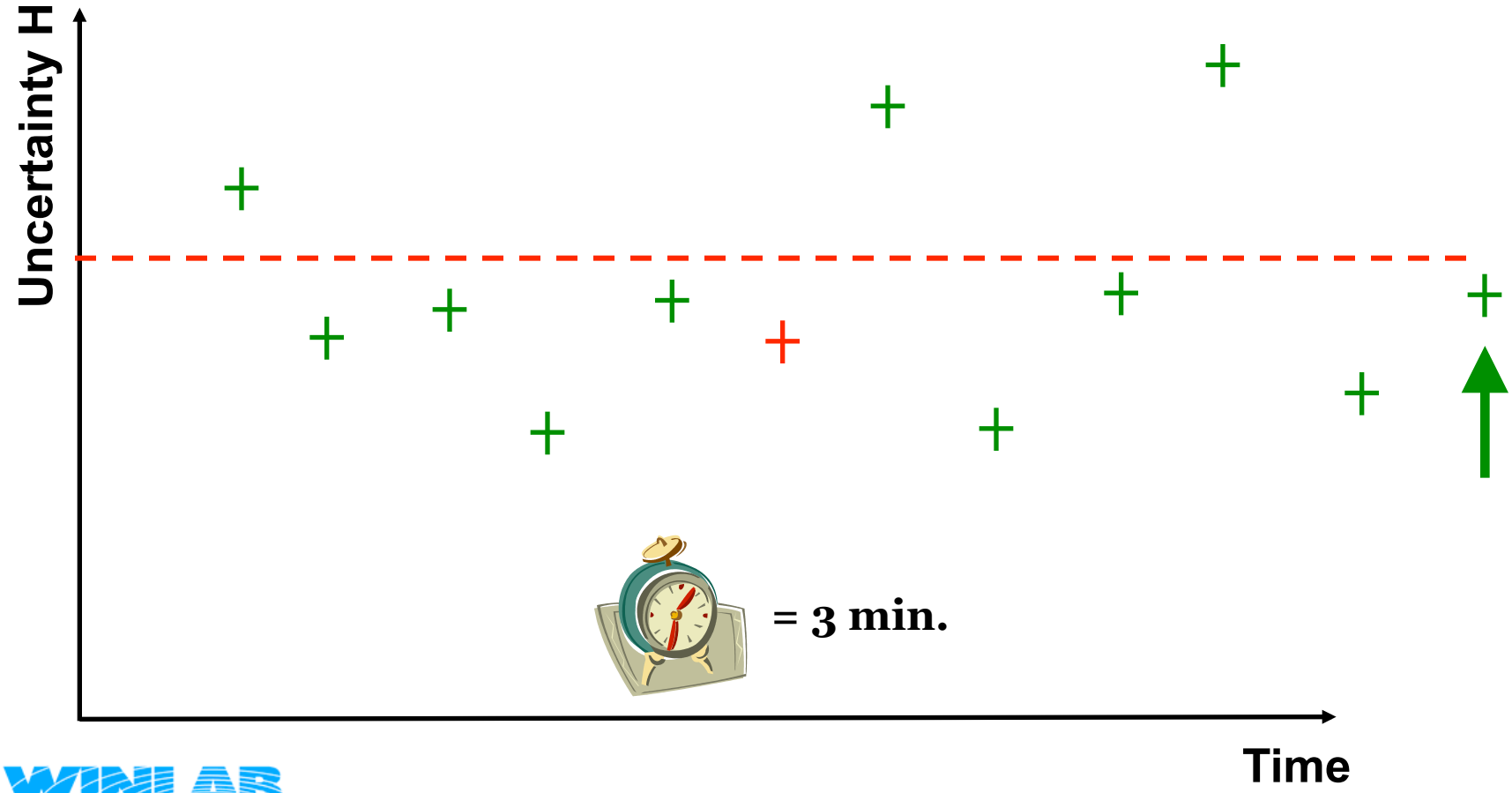
Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$

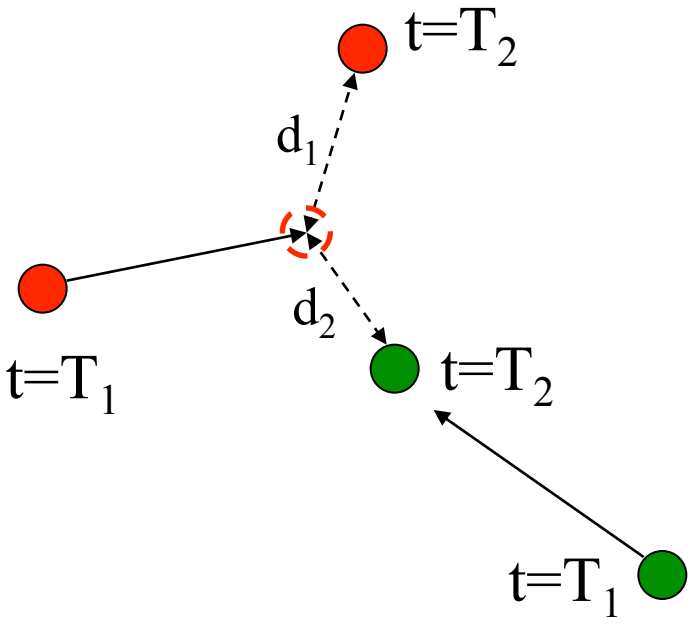


Algorithm: Uncertainty-Aware Path Cloaking

$$H = -\sum p_i \log p_i$$



Algorithm Details: User Interdependency and Pruning Process

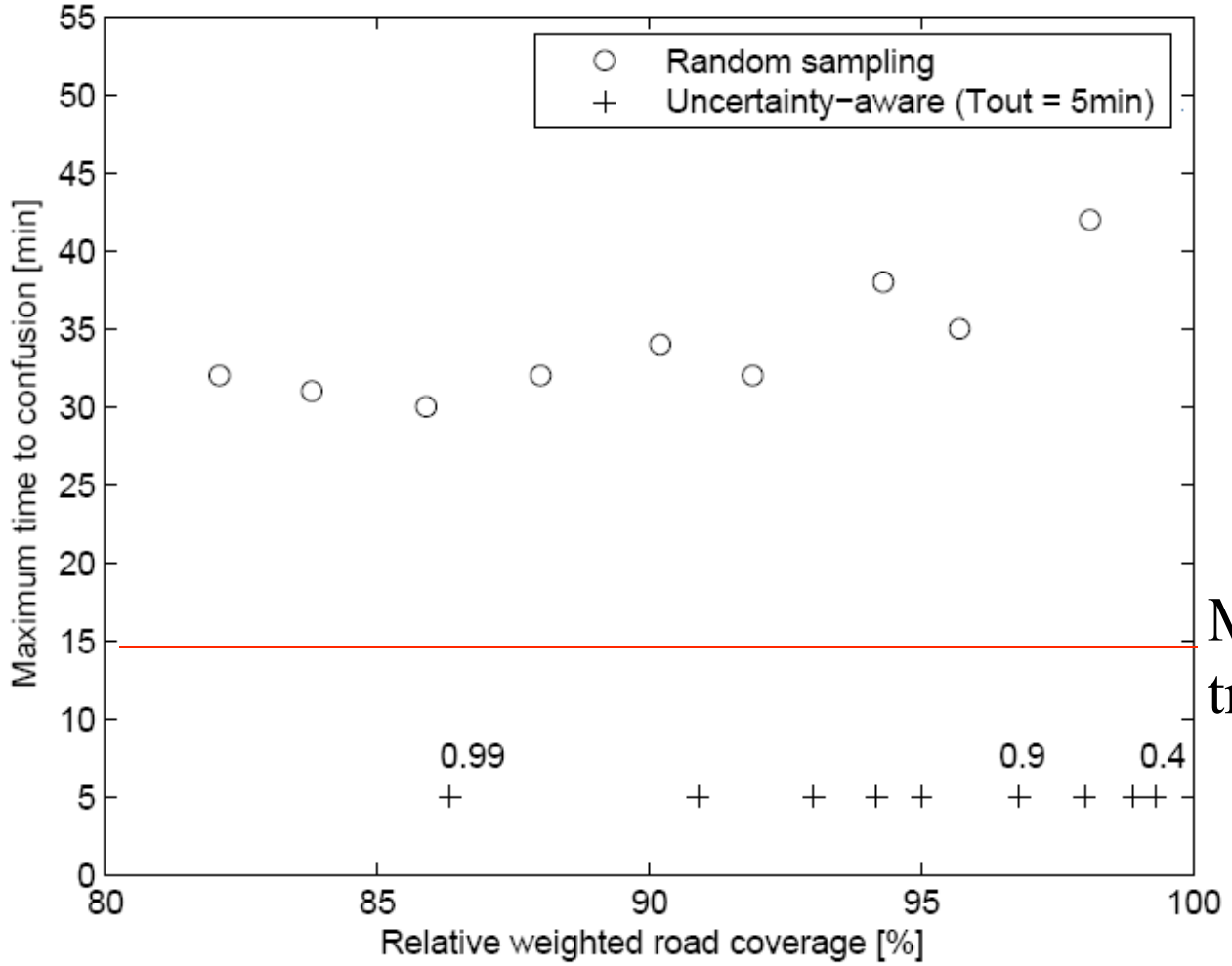


- The algorithm calculates uncertainty only with **Revealed** samples
  - Initialize the candidate set with all location samples at time t
  - The algorithm subsequently prunes the candidate set until only vehicles remain who meet the uncertainty threshold using the candidate set samples
  - The algorithm reveals all location samples in the candidate set

Evaluation

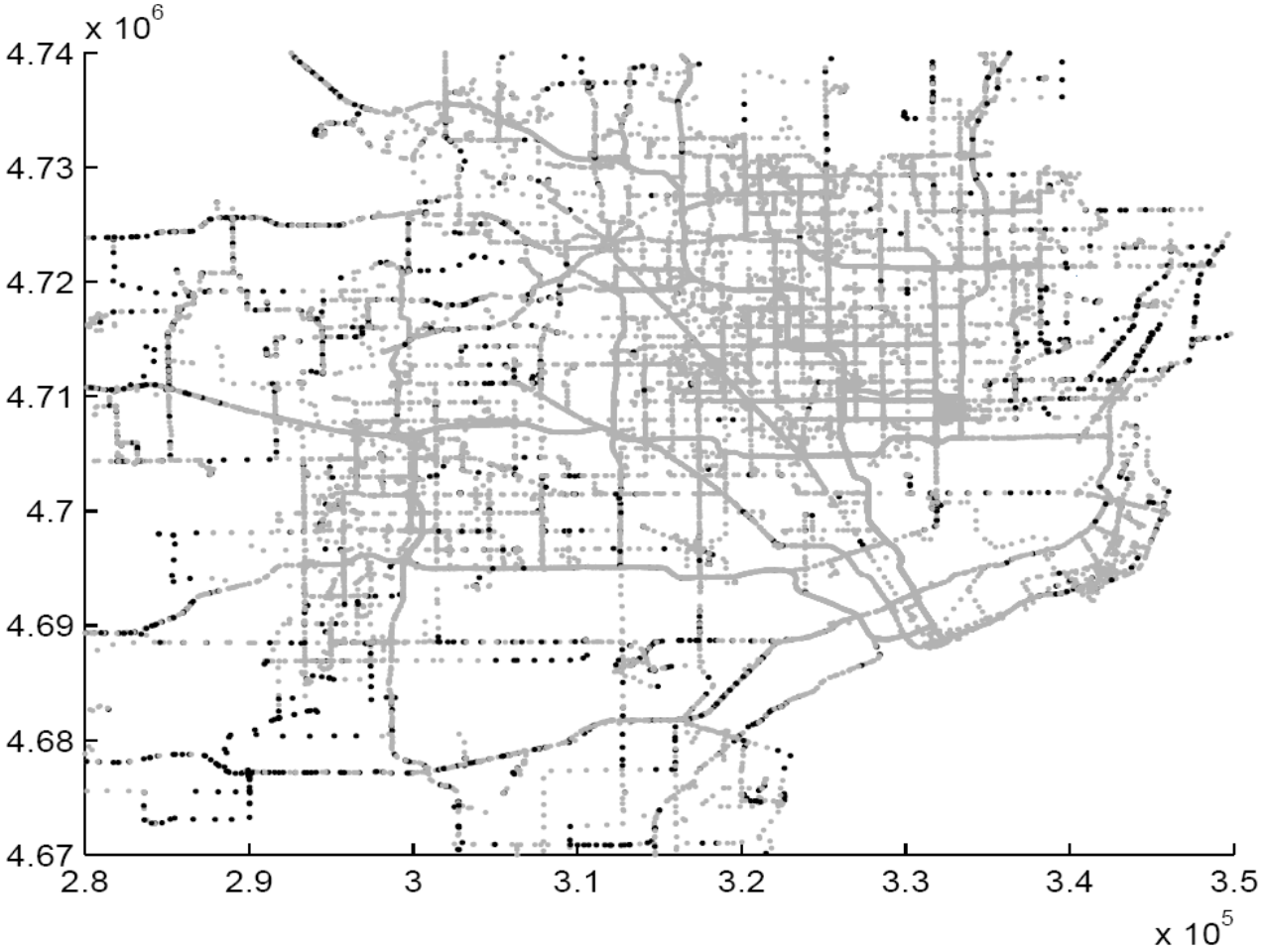
- Data set: 24-hour GPS traces of 2000 probe vehicles area (built from ~200 actual vehicles)
- 70km-by-70km
- Metrics: Tracking time and (relative) road coverage

Evaluation: Protection against Target Tracking



Median trip duration

Privacy-filtered GPS Traces



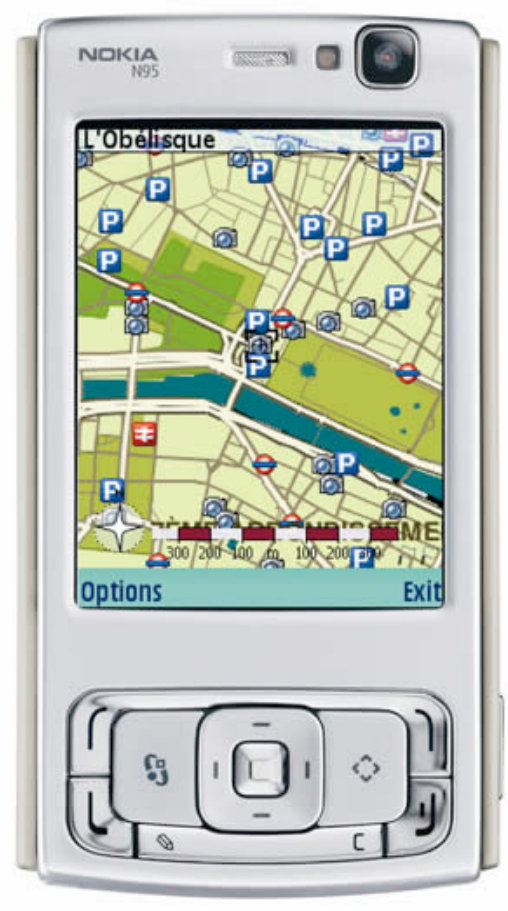
RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Virtual Trip Lines and Secret Splitting – A Distributed Architecture

with Nokia Research Center, UC Berkeley

(Mobisys 2008)

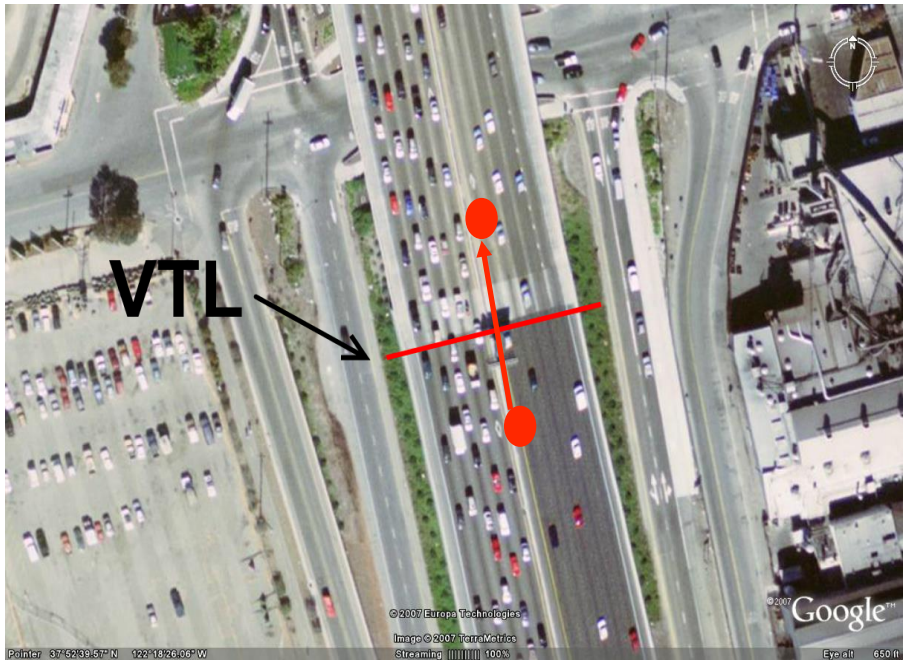
Lightweight Vehicular Networking with Smartphones

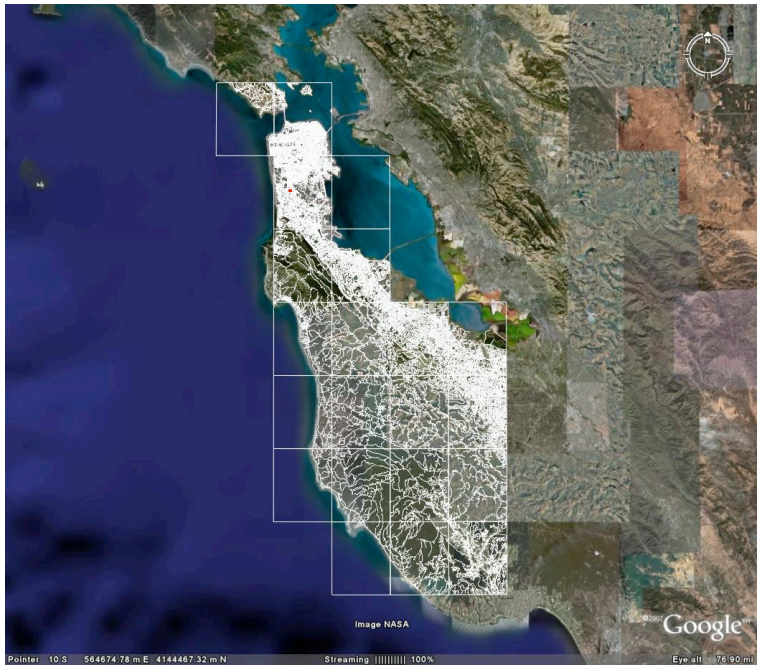


- Vehicular network applications:
  - Safety
  - Traffic Information / Management
  - Entertainment
- Phones increasingly
  - openly programmable
  - Equipped with GPS
- Phone-based solutions can quickly achieve sufficient penetration rates
- Challenge: Privacy

Virtual Trip Lines (VTLs) - Sampling in Space

- Crossing Virtual Trip Lines trigger Location Updates, rather than periodic update
- Local VTLs stored in phone



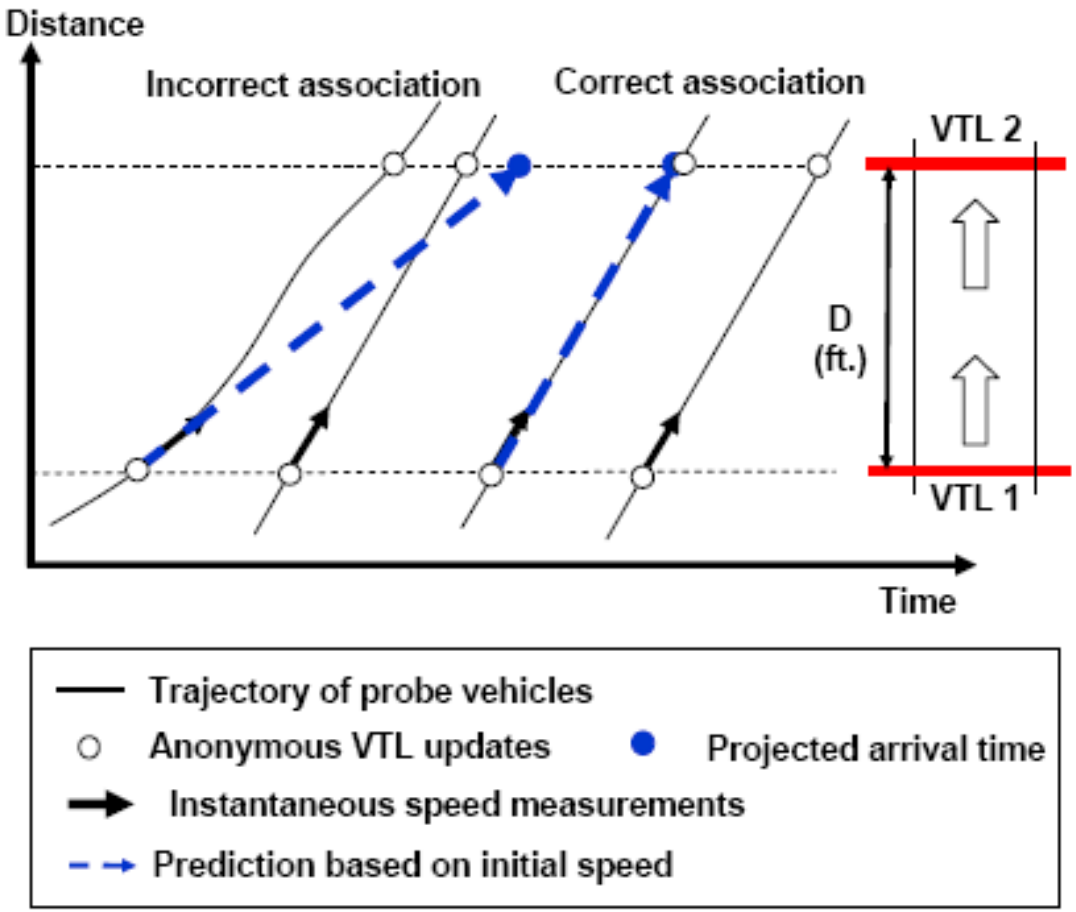


- Allows generating (anonymous) updates only from key roads, presumably less sensitive areas !

WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

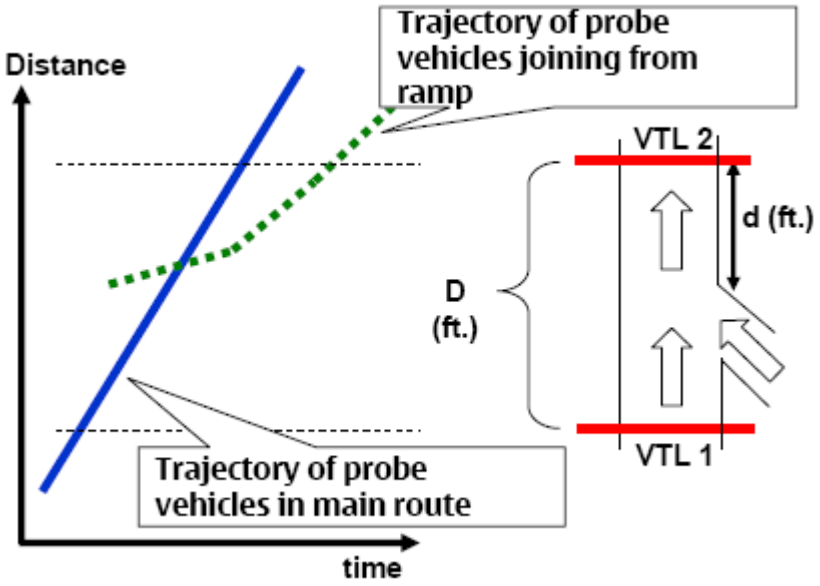
Placement Privacy Constraints: Minimum Spacing

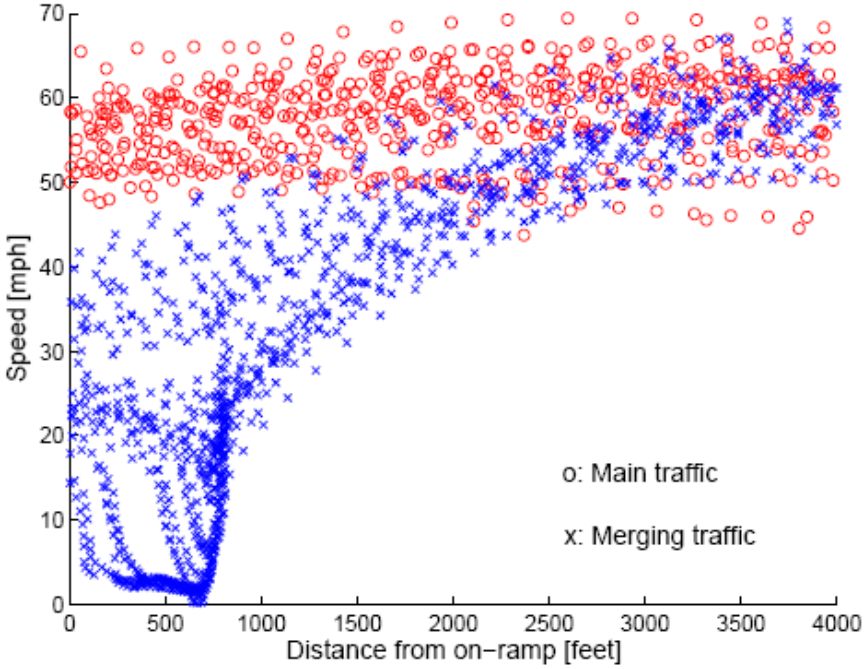
- Tracking uncertainty is dependent on the spacing between VTLs, the penetration rate, and speed variations of vehicles



Placement Privacy Constraints: Exclusion Areas

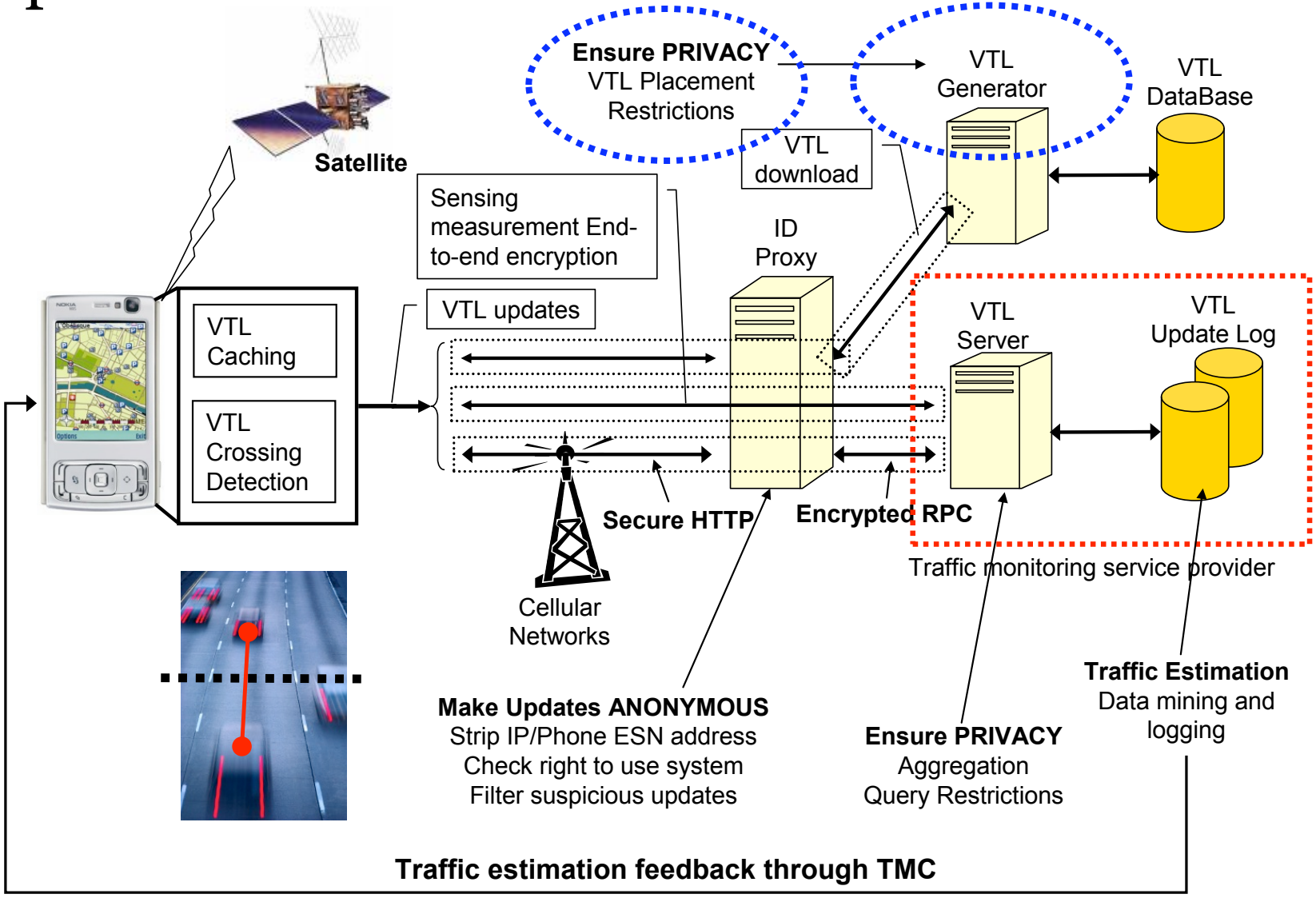
- Low speed samples are likely generated by vehicles that just entered after the ramp
- Suppress sampling on on-/off-ramps





WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

Split-Secret VTL Architecture



Cloaking Extensions and Secret Splitting

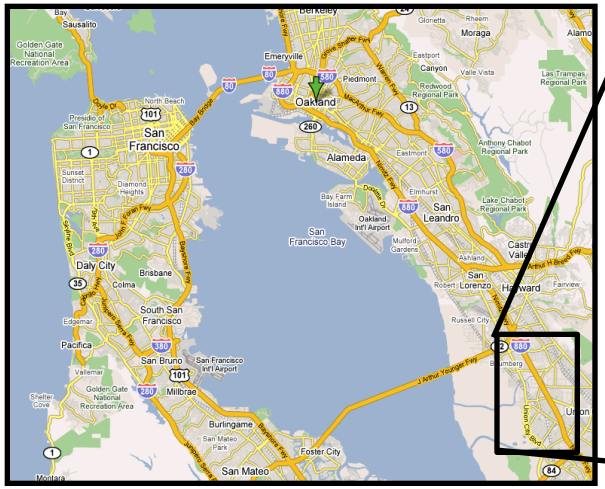
- Distributed VTL-Based Temporal Cloaking
  - Traffic estimation is relatively *immune to temporal error*
- No single entity (except handset) has access to both location and time

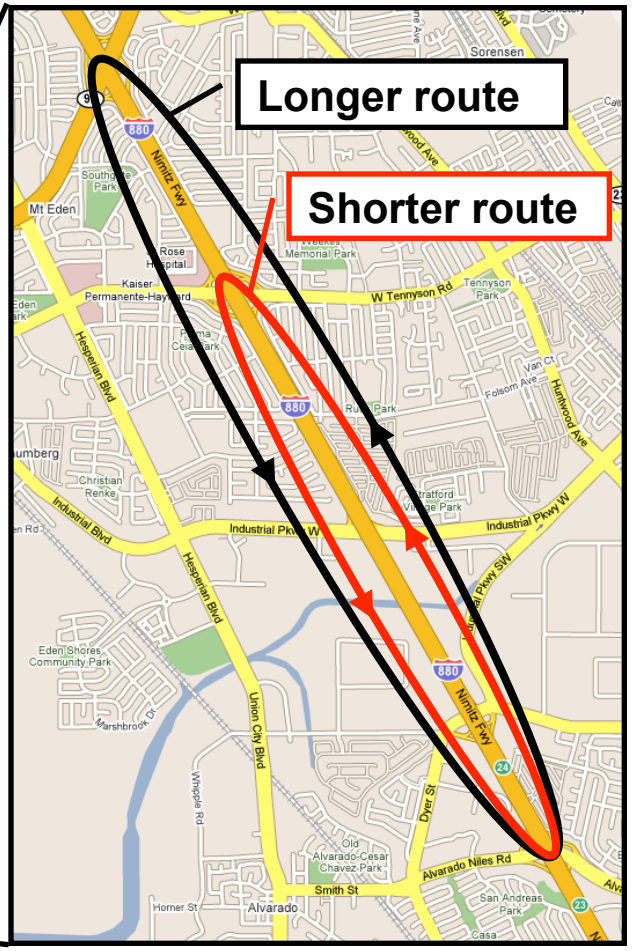
Entity	Role	Identity	Location	Time
Handset	Sensing	Yes	Accurate	Accurate
Location Verifier	Distributing VTL ID updates	Yes	Coarse	Accurate
ID proxy	Anonymizing and Cloaking	Yes	Not available	Accurate
Traffic Server	Computing Traffic Congestion	No	Accurate	Cloaked

Virtual Trip Lines

Temporal Cloaking

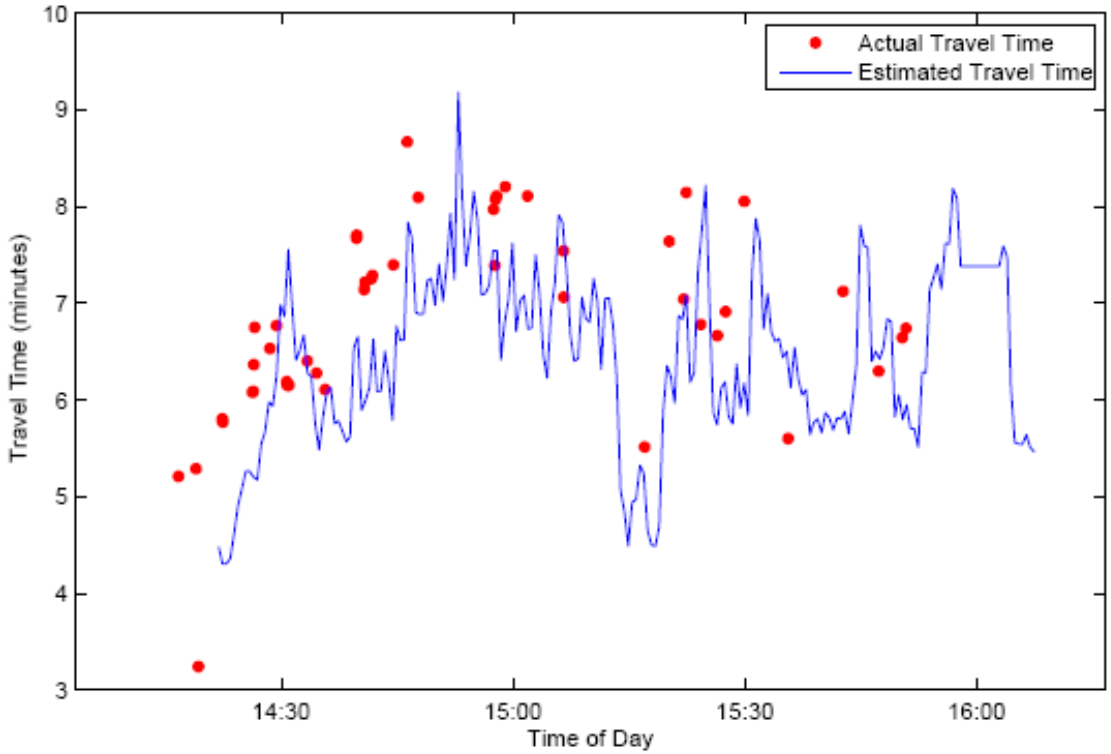
Experiment (20-cars, 1~2% penetration rate)





WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

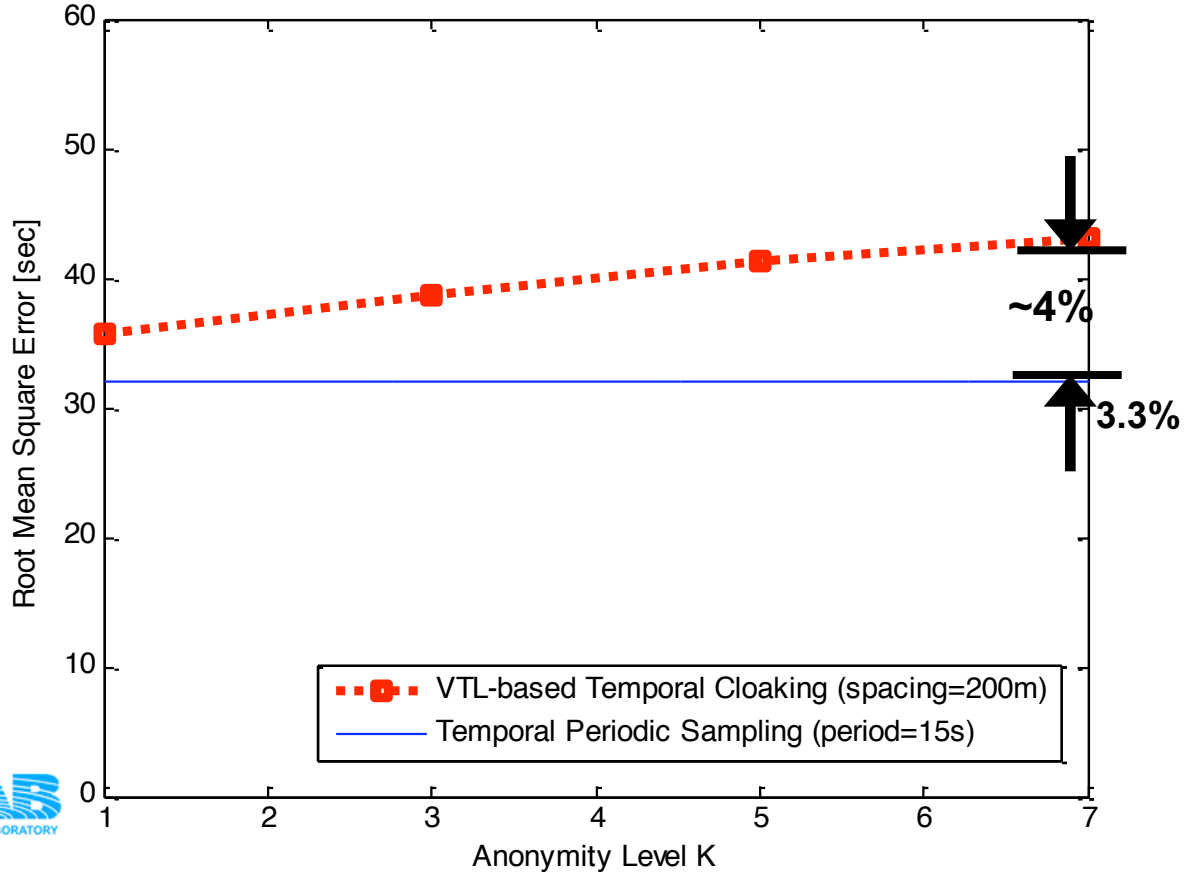
Travel Time Accuracy



- Even few probe vehicles (1~2%) achieve travel time estimates with less than 15% error

Guaranteed Privacy

- Applying the cloaking techniques reduces travel time estimation accuracy by less than 5% compared to a standard periodic sampling approach



WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

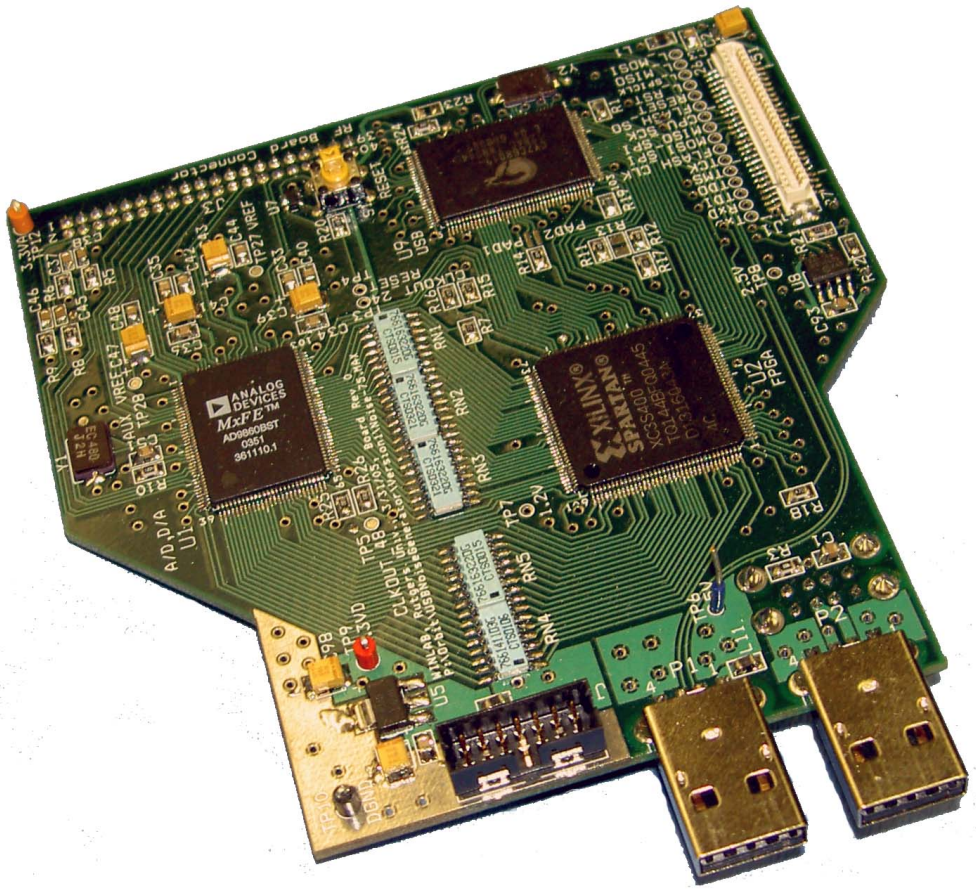
RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Identifying Transmitters via Radiometric Signatures – Can Transmitters be Tracked at the Physical Layer?

with Univ of Wisconsin (MobiCom' 08)

Trends

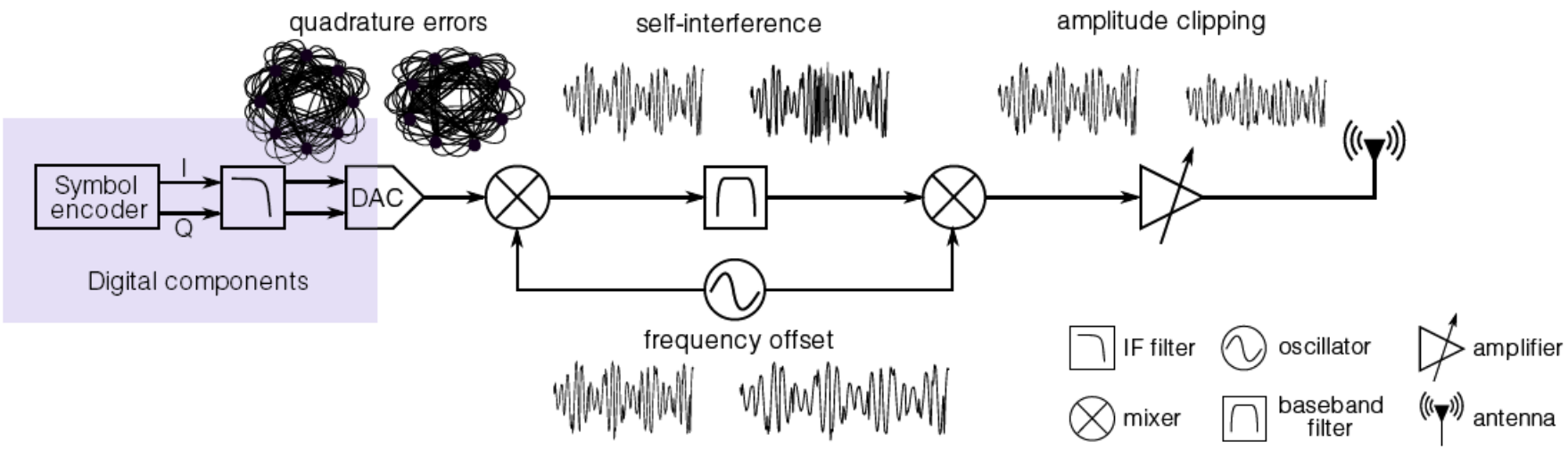
Always-on, transmitting background apps



Low-cost software (controlled) radios (e.g., GNU radio) allow eavesdropping in many bands

WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

Waveform Impairments in Analog Frontend



Transmitter Identification via Classification

- Training phase: collect fingerprint (waveform error metrics) of each transmitter

Error type	unit	reference	range	definition
frequency	Hz	2142 MHz	± 60.3	$\pm 25ppm\ f_c$
phase	$^\circ$	ideal symb	± 10	$\text{asin}(E_{max})$
magnitude	n/a	ideal symb	± 0.17	$\pm E_{max}$
EVM	n/a	ideal symb	$[0, .35]$	upto $2E_{max}$
I/Q offset	n/a	ideal origin	$[0, 0.17]$	upto E_{max}
SYNC	%	max corr.	$[0, 1]$	correlation

f_c – channel frequency E_{max} – max I/Q error

- Identification phase: measure error metrics for candidate transmitter and use classification algorithm to match with training set


ORBIT

Identification Results

- Using ORBIT testbed radios and vector signal analyzer for data collection
- K-Nearest Neighbor and Support Vector Machine classifiers

Approach	NIC population size	Bin size	Training fraction	Reported error rate[1]	Equivalent performance of PARADIS-kNN	PARADIS-SVM
Franklin et. al. [13][2]	17	8	5%	15%	0%	0%
Hall et. al. [18][3]	30	10	33%	8%	0%	0%
PARADIS	**138**	**4**	**20%**	-	**3%**	**0.34%**

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Security Analysis of In-Car Wireless Systems

with Univ of South Carolina (Security 2010)

Tire-Pressure Monitoring Systems

- First wireless sensor system built into (nearly) every car

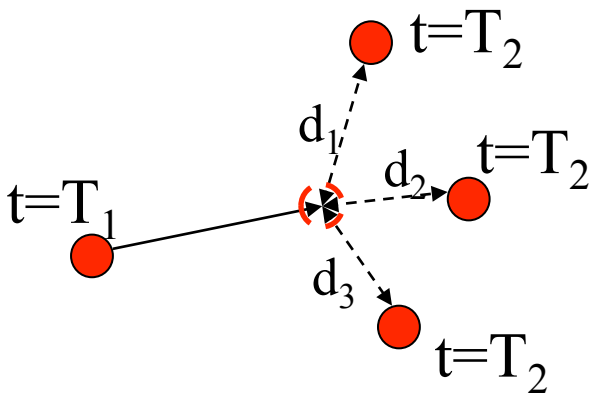
Reverse Engineering

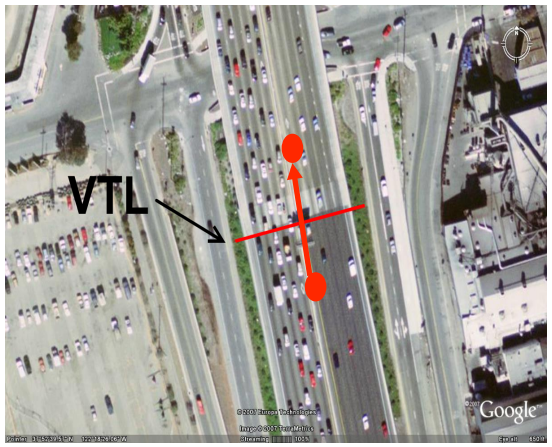
- Determine
  - Frequencies
  - Modulation
  - Packet format

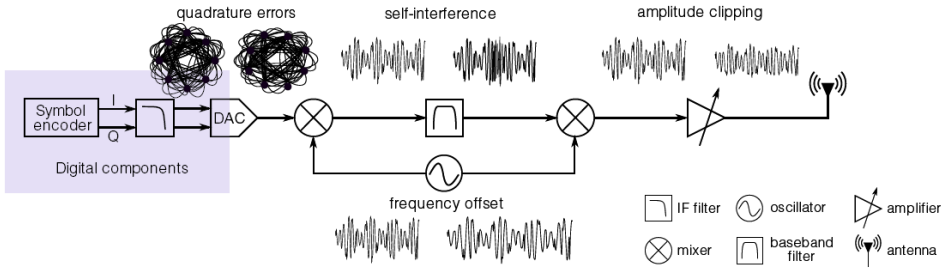
Security and Privacy Concerns

- Privacy concerns
  - 32 bit static identifiers
  - Activation signal
  - Achieved up to 40m range
- Security concerns
  - Ineffective input validation allows spoofing
  - Bricked the TPMS ECU

Summary: Mechanisms for Depersonalizing Location Traces







WINLAB
WIRELESS INFORMATION NETWORK LABORATORY

RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Thank you