15-446 Distributed Systems
Spring 2009

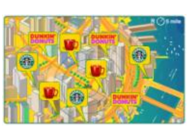
Localization

Announcements

- Assignment HW3 due 4/21
- No lecture
  – 4/16 Carnival
- Schedule change
  – 4/21 lecture moved to recitation slot on 4/22 (Wed)
  – Location TBA

Localization

- What is localization? ➔ Telling where you are
- Why? ➔ Location Base Services
  – E-911 Emergency assistance: Where are they?
  – Advertising: You are here!
  – Social networking: Where are my friends?
  – Tracking: Where are you going?
  – Virtual Tour: I am close to…
  – Service discovery: What is here?
  – Device recovery: Where's my stolen laptop?



Overview

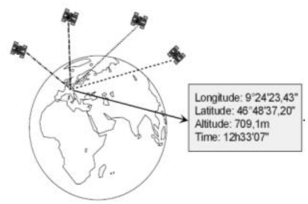
- GPS
- WiFi Positioning Systems
- IP2Geo

Global Positioning System

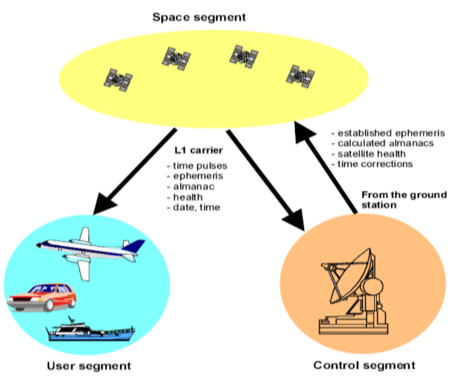
- NAVSTAR-GPS
  - NAVigation System with Timing And Ranging Global Positioning System
  - Satellite-based distributed system
- First satellite launched in 1978 by DoD
- Fully operational in April, 1995.
- Civilian use(Standard Positioning Service) and military use (Precise Positioning Service)

Basic function of GPS

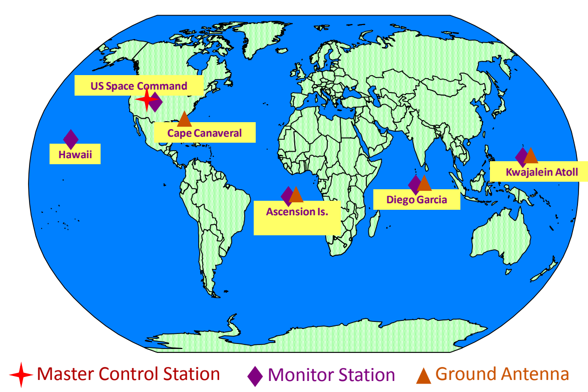
1. Exact location (longitude, latitude,height) Accuracy ~15m
2. Precisetime (UTC) Accuracy ~45ns



Three segments of GPS



Control Segment



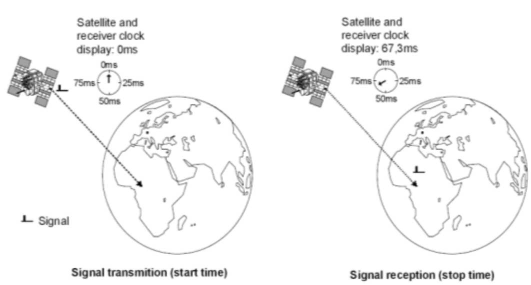
Space Segment

- 28 satellites in 6 orbits
- Each satellite has 4 atomic clocks
- Orbits the Earth at 20,180km
- 12 hours orbital time
- Atleast4satellites can be seen in any part of the planet

Positioning is based on time

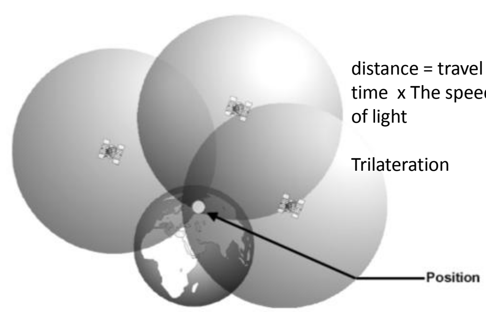
- Satellite keeps an accurate time
  - Atomic clock synchronize with each other
  - Regularly adjusted from control segments on Earth
- Each satellite broadcasts the exact location and time at 1575.42 MHz (L1).
- Receiver can calculate the travel time if the clock is synchronized.

Positioning with synchronized clock



distance = travel time x The speed of light

Positioning with synchronized clock



distance = travel time x The speed of light

Trilateration

Accounting for the clock offset

- Satellites' clocks are well synchronized
- Receiver clock is not synchronized.
- Need to estimate 4 unknowns
  - (x, y, z, Δt)
  - Δt is the clock offset of the receiver
  - R: real distance, PSR : estimated distance
  - R = PSR - Δt ·c
- $R=\sqrt{\left(X_{Sat}-X_{User}\right)^2+\left(Y_{Sat}-Y_{User}\right)^2+\left(Z_{Sat}-Z_{User}\right)^2}$
- Need 4 satellites

GPS accuracy

- ~15 meter (worsens in urban canyon)
- Source of Error (speed of light changes)

Cause of error	Error
Effects of the ionosphere	4 m
Satellite clocks	2.1 m
Receiver measurements	0.5 m
Ephemeris data	2.1
Effects of the troposphere	0.7
Multipath	1.4 m
Total RMS value (unfiltered)	5.3 m
Total RMS value (filtered)	5.1
Vertical error (1 sigma (68.3%) VDOP=2.5)	12.8m
Vertical error (2 sigma (95.5.3%) VDOP=2.5)	**25.6m**
Horizontal error (1 sigma (68.3%) HDOP=2.0)	10.2m
Horizontal error (2 sigma (95.5%) HDOP=2.0)	**20.4m**

Table 4: Cause of errors

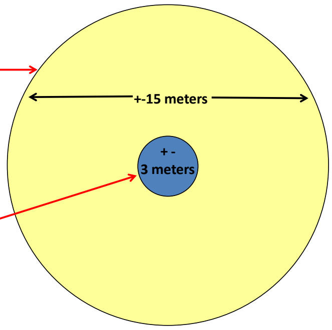
Wide Area Augmentation System

- Error correction system that uses reference ground stations
- 25 reference stations in US
- Monitor GPS and send correction values to two geo-stationary satellites
- The two geo-stationary satellites broadcast back to Earth on GPS L1 frequency (1575.42MHz)
- Only available in North America, WASS enabled GPS receiver needed

WAAS

How good is WAAS?

With Selective Availability set to zero, and under ideal conditions, a GPS receiver without WAAS can achieve fifteen meter accuracy most of the time.*



Under ideal conditions a WAAS equipped GPS receiver can achieve three meter accuracy 95% of the time.*

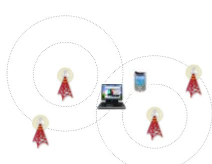
* Precision depends on good satellite geometry, open sky view, and no user induced errors.

Overview

- Need for localization
  - Location Based Service
- GPS
- **WiFi Positioning Systems**
- IP2Geo

WiFi Positioning System

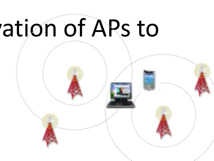
- Exploit wide-scale WiFi deployment
  - WiFi density increasing to point of overlap
- WiFi base stations broadcast unique IDs
- Position WiFi devices using a map of AP's MAC to location



Place Lab 19

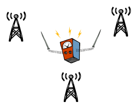
PlaceLab

- Keeps a database of WiFi measurements
  - Maps BSSID (MAC address) to locations
  - Wigle.net, public war driving database
- Mobile users download part of the DB
- Match the current observation of APs to location using the DB
- Accuracy around 20m



PlaceLab localization algorithms

- Centroid/Weighted Centroid
  - Locates the client at the average location of the Aps



- Finger printing
  - Requires signal map at each location
  - Find the nearest neighbor in signal strength space
- Particle filter
  - Requires signal map at each location
  - Takes account of the user movement

PlaceLab

- Localization algorithms used to locate mobile users
  - Centroid/Weighted Centroid
  - Finger printing
  - Particle filter

Algorithm		Downtown (meters)	High density Ravenna (meters)	low density Kirkland (meters)
centroid	basic	24.4	14.8	37.0
	weighted	23.4	14.5	37.0
fingerprint (k=4)	radar	18.5	15.3	30.0
	rank	20.3	16.7	59.5
particle filter	signal strength	18.0	14.4	29.7
	response rate	21.3	12.9	28.6

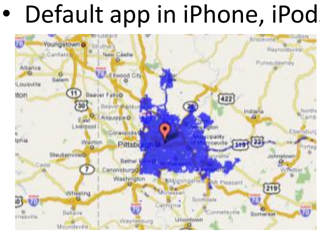
Table 2: Median error in meters for all of our algorithms across the three areas.

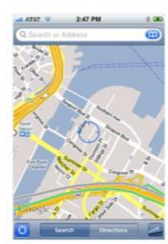
PlaceLab

- Requires a database of WiFi hotspots
- Cost
  - Infrastructure (wireless APs) are already there
  - Wireless device is cheap
- Coverage
  - Indoors as well as outdoors
  - Most areas have dense deployment of APs

SkyHook

- Commercial system extended from PlaceLab
- Uses WiFi, GPS, Cell tower (hybrid approach)
- Updates the database regularly
- Default app in iPhone, iPods





Indoor localization system

- Usually more fine grained localization needed
  - Often 3D (2.5D) : x,y and floor
  - Often want to locate users in an office
- RADAR
  - Trilateration based on signal strength from APs
  - Hard to predict distance based on signal strength because signal is blocked by walls and structures
  - Use site-surveying
- Lots of research has been done
  - MIT Cricket (RF + ultrasound)
  - AeroScout (WiFi), Ekahau (WiFi)

Overview

- Need for localization
  - Location Based Service
- GPS
- WiFi Positioning Systems
- **IP2Geo**

IP-Geography Mapping

- Goal: Infer the geographic location of an Internet host given its IP address.
- Why is this interesting?
  - enables location-aware applications
  - example applications:
    - Territorial Rights Management
    - Targeted Advertising
    - Network Diagnostics
- Why is this hard?
  - IP address does not inherently indicate location
  - proxies hide client identity, limit visibility into ISPs
- Desirable features of a solution
  - easily deployable, accuracy, confidence indicator

27

IP2Geo

- Infer geo-location of IP based on various "properties"
  - DNS names of routers often indicate location
  - Network delay correlates with geographic distance
  - Subnets are clustered
- Three techniques
  - GeoTrack
  - GeoPing
  - GeoClusters

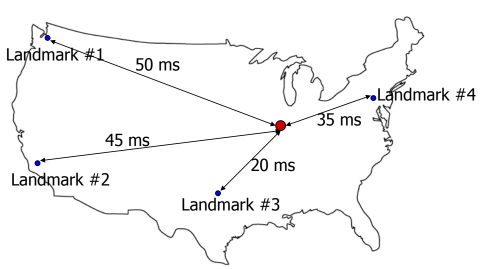
GeoTrack

- Location info often embedded in router DNS names
  - ngcore1-serial8-0-0-0.Seattle.cw.net, 184.atm6-0.xr2.ewr1.alter.net
- GeoTrack operation
  - do a traceroute to the target IP address
  - determine location of last recognizable router along the path
- Key ideas in GeoTrack
  - partitioned city code database to minimize chance of false match
  - ISP-specific parsing rules
  - delay-based correction
- Limitations
  - routers may not respond to traceroute
  - DNS name may not contain location information or lookup may fail
  - target host may be behind a proxy or a firewall

GeoPing

- Nearest Neighbor in Delay Space(NNDS)
  - delay vector: delay measurements from a host to a fixed set of landmarks
  - delay map: database of delay vectors and locations for a set of known hosts
    - (50,45,20,35) ↔ Indianapolis, IN
    - (10,20,40,60) ↔ Seattle, WA
    - •••
  - target location corresponds to best match in delay map
  - optimal dimensionality of delay vector is 7-9

Delay Map Construction



Delay Vector = (50,45,20,35) ↔ Indianapolis, IN

GeoCluster

- Basic Idea: identify geographic clusters
  - partial IP-location database
    - construct a database of the form (IPaddr, likely location)
    - partial in coverage and potentially inaccurate
    - sources: HotMail registration/login logs, TVGuide query logs
  - cluster identification
    - use prefix info. from BGP tables to identify topological clusters
    - assign each cluster a location based on IP-location database
    - do sub-clustering when no consensus on a cluster's location
  - location of target IP address is that of best matching cluster

Constructing IP-Location Database

Registration logs

User A ↔ San Francisco, CA
User B ↔ Berkeley, CA
User C ↔ Little Rock, AK
User D ↔ San Francisco, CA
User E ↔ New York, NY
User F ↔ Clinton, AK

Login logs

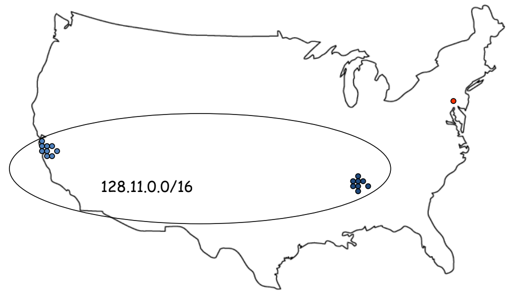
User A ↔ 128.11.20.35
User B ↔ 128.11.35.123
User C ↔ 128.11.132.40
User D ↔ 128.11.20.145
User E ↔ 128.11.100.23
User F ↔ 128.11.163.112

IP-location database

128.11.20.35 ↔ San Francisco, CA
128.11.35.123 ↔ Berkeley, CA
128.11.132.40 ↔ Little Rock, AK
128.11.20.145 ↔ San Francisco, CA
128.11.100.23 ↔ New York, NY
128.11.163.112 ↔ Clinton, AK

33

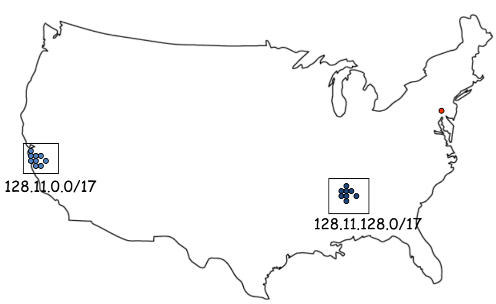
Geographic sub-clusters in a cluster



No consensus in location estimate for entire cluster

34

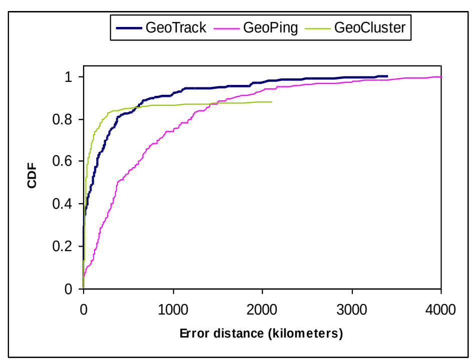
Geographic sub-clusters in a cluster



Consensus in location within sub-clusters

35

Performance



Median Error: GeoTrack :102 km, GeoPing: 382 km, GeoCluster: 28 km

IP2Geo Conclusions

- IP2Geo encompasses a diverse set of techniques
  - GeoTrack: DNS names
  - GeoPing: network delay
  - GeoCluster: geographic clusters
- Median error 20-400 km
  - GeoCluster also provides confidence indicator
- Each technique best suited for a different purpose
  - GeoTrack: locating routers, tracing geographic path
  - GeoPing: location determination for proximity-based routing (e.g., CoopNet)
  - GeoCluster: best suited for location-based services
- Publications at SIGCOMM 2001 & USENIX 2002

37

Summary

- Location based services require the location of users
- Different localization systems are built for different purposes
- Each localization system has different performance and limitations
- Localization and location based systems are interesting/active area of research