

Geolokacja w Internecie

Wspomaganie geolokacji IP

Andrzej Pańkowski

a.pankowski@students.mimuw.edu.pl

Wydział MIM UW



Wprowadzenie

Czym jest geolokacja?

- To proces określania położenia geograficznego osoby lub rzeczy
 - przy użyciu systemu geolokacyjnego:
 - sonar (*SOund Navigation And Ranging*)
 - GPS
 - radar (*RAdio Detection And Ranging*)
 - ...
 - W dalszej części skupimy się na systemach geolokacyjnych które są ogólnie dostępne →

GPS (*Global Positioning System*)

- Zalety:
 - dokładne pozycjonowanie i nawigacja oparte o sygnał z wielu satelitów krążących wokół Ziemi → dostępne w każdym momencie, w każdym miejscu na ziemi i w powietrzu
 - darmowy
 - nie wymaga bazy danych, wystarczy odbiornik GPS
- Wady:
 - sygnał jest zakłócany przez przeszkody → raczej nie nadaje się do ustalania pozycji np. w budynkach

Pozycjonowanie BTS (*Base Transceiver Station*)

- Zalety:
 - to „efekt uboczny” pracy urządzeń łączności bezprzewodowej (np. GSM, WCDMA) w trybie ciągłym
- Wady:
 - mniej dokładne niż GPS
 - dokładność zależy od koncentracji BTS-ów
 - BTS-y nie udostępniają swojej pozycji, tylko IDy → wymaga utworzenia jakiejś bazy danych



Pozycjonowanie Wi-Fi (czasami zwane WPS)

- Zalety:
 - dokładność większa niż dla pozycjonowania BTS
 - może działać także tam gdzie sygnał GPS nie dociera
- Wady:
 - nawigacja teoretycznie możliwa, ale niepraktyczna
 - częściowe pokrycie zasięgiem
 - opiera się na jakiejś bazie adresów MAC punktów dostępu (*AP*)

Geolokacja IP aka Whois Lookup

- Zalety:
 - ogólnodostępne i darmowe
 - nie wymaga od użytkownika instalacji dodatkowych programów ani sprzętu → wystarczy dostęp do Internetu
- Wady:
 - możliwe określenie tylko kraju i miasta → pozycjonowanie o małej dokładności
 - opiera się na zewnętrznej bazie adresów IP (np. RIPE w Europie)
 - można oszukać → serwery proxy



Systemy wspomagające

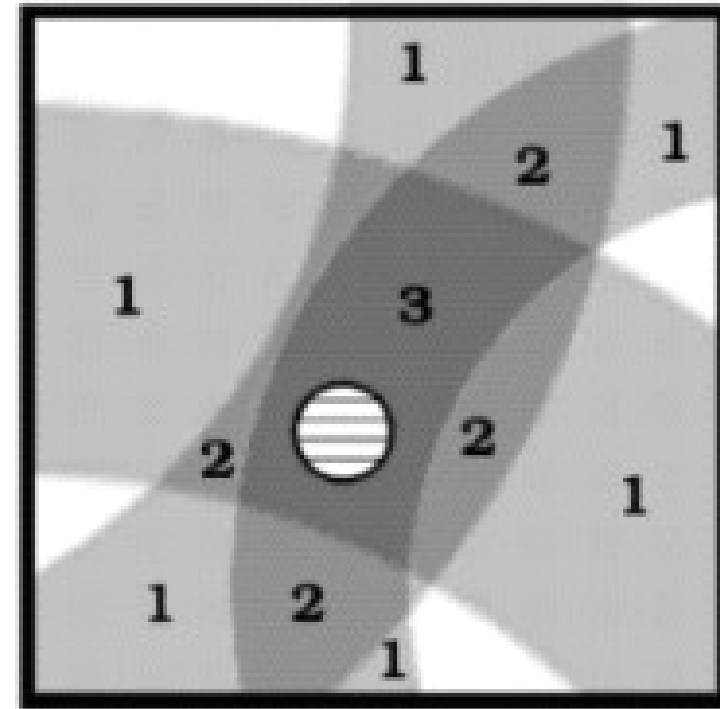
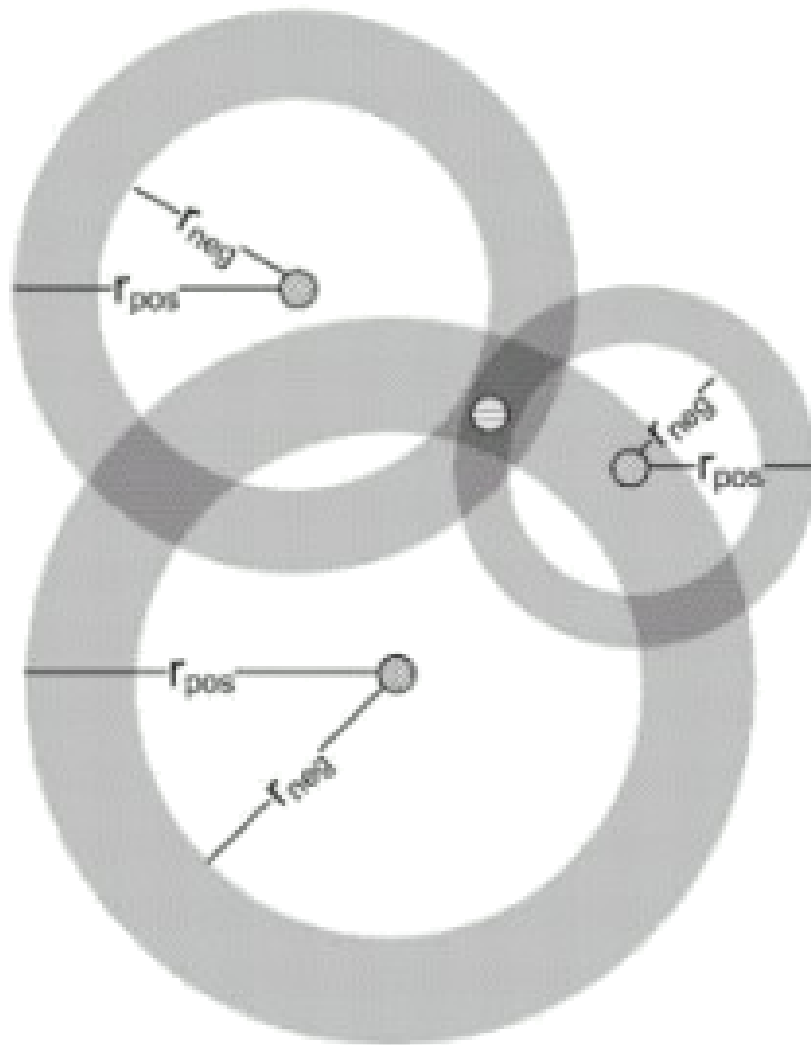
- utrzymuje i uaktualnia bazę geolokacji > 100M BTS-ów i punktów dostępu Wi-Fi → zatrudnia kierowców którzy jeżdżąc zbierają dane
- hybrydowy system XPS: dane z GPS, BTS, Wi-Fi → pozycja
- XPS jest odpłatny → firmy mogą wykupić dostęp, użytkownicy gotowe aplikacje tworzone przez SKYHOOK
- należy do czołówki firm dostarczających usługi geolokacyjne

- partnerski system geolokacji → użytkownik zbiera punkty za oznaczone GPS-em lub odwiedzone AP i BTS-y, punkty wymienia na \$\$\$
- trzeba oznaczyć ~500 BTS-ów i ~500 AP żeby dostać \$10 ;(
- ale można zapraszać przyjaciół → 20% wartości ich punktów trafia na nasze konto ;)
- aplikacje: darmowa (tylko pozycjonowanie BTS) i płatna (BTS i AP)

- na pewno utrzymuje bazę geolokacji BTS-ów i AP, być może posiada licencje dostępowe u konkurencji
- darmowe aplikacje: Maps, Latitude, Gears
- ale też AdWords i wyniki wyszukiwania
- w urządzeniach mobilnych wykorzystuje GPS
- ciężko powiedzieć czy udostępnia swoją bazę

- problem geolokacji sprowadzony do problemu ograniczeń geometrycznych
- wystarczy IP lokalizowanego hosta i – na niewielką skalę – kilkadziesiąt hostów testujących
- hosty testujące mierzą RTT i wykonują traceroute do celu badając topologię połączeń
- ograniczenia reprezentowane są przez krzywe Bezierra i w takiej postaci przetwarzane
- wynikiem jest krzywa – granica określająca przypuszczalne położenie lokalizowanego hosta

Położenie celu jako przecięcie krzywych – ograniczeń



- ⊖ Target node
- Landmark node

Octant's estimated location is: -105.057, 39.813

GeoLim could not localize target.

Octant's estimated region is shown with blue lines, GeoLim's estimated region is shown with red lines.

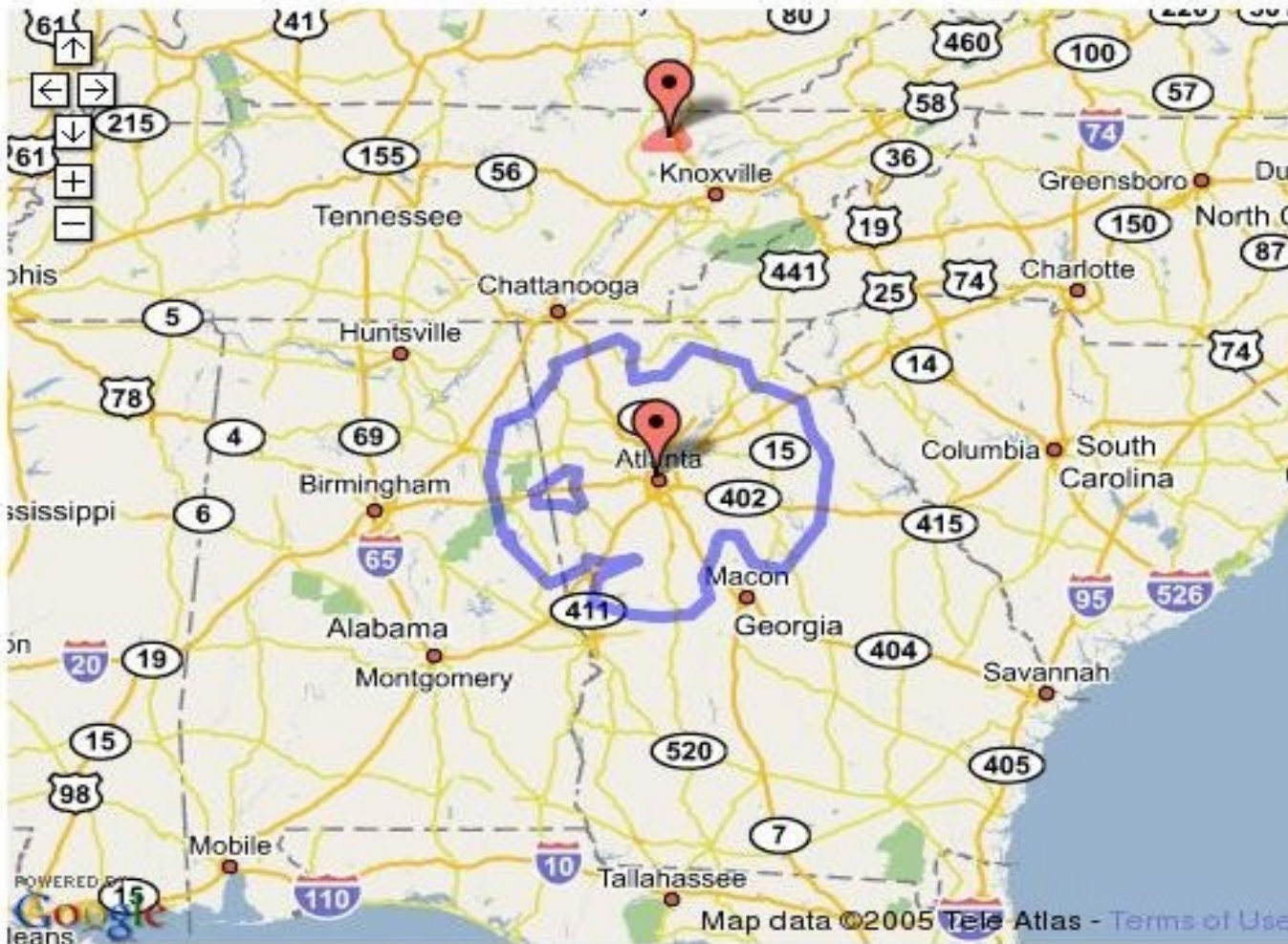


Test: www.emory.edu

Octant's estimated location is: -84.423, 33.775

GeoLim's estimated location is: -84.310, 36.376

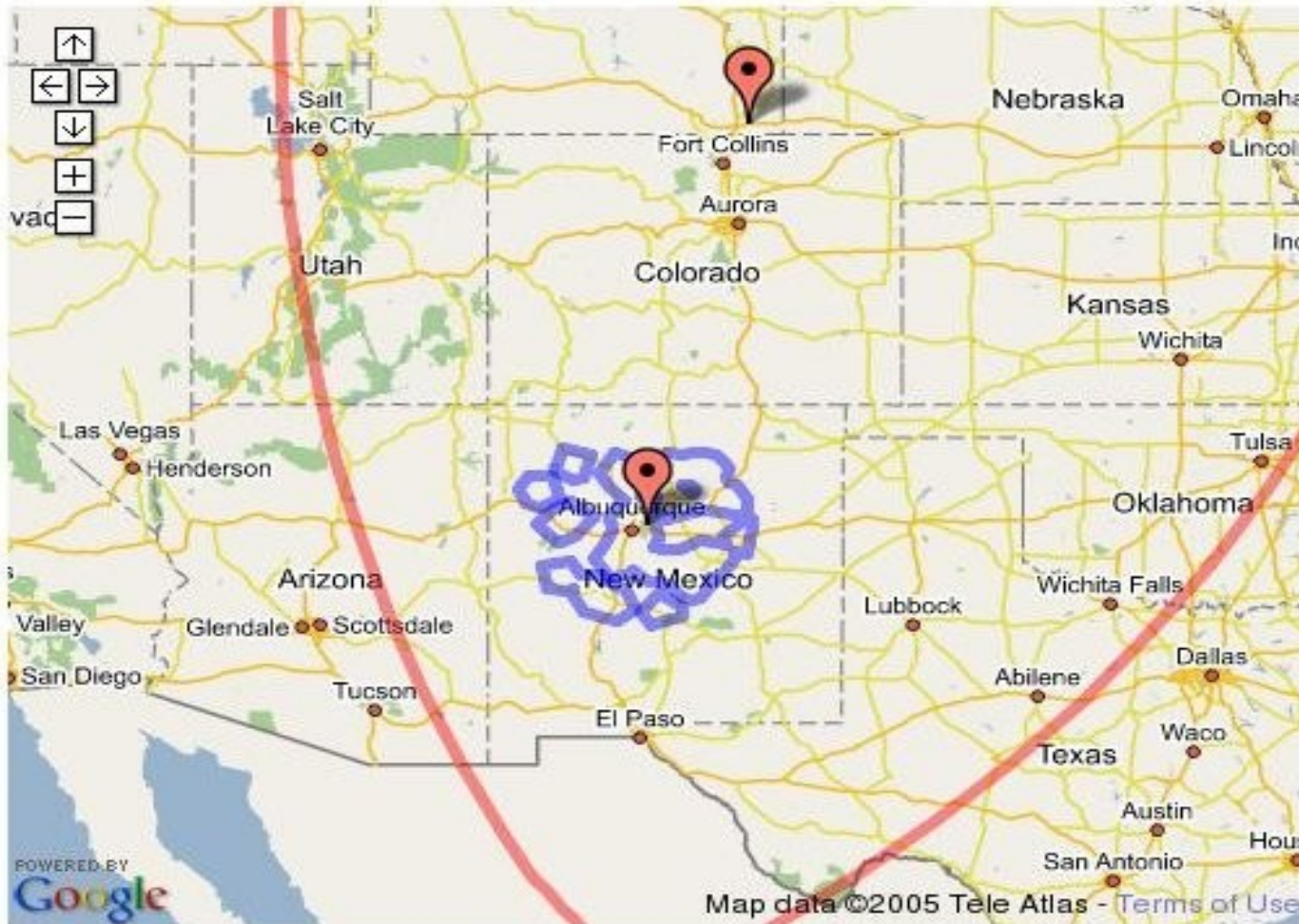
Octant's estimated region is shown with blue lines, GeoLim's estimated region is shown with red lines.



Octant's estimated location is: -106.370, 35.124

GeoLim's estimated location is: -104.650, 41.130

Octant's estimated region is shown with blue lines, GeoLim's estimated region is shown with red lines.



- projekt który w zamyśle autorów miał być Wikipedią geolokacji punktów dostępowych Wi-Fi
- niestety mimo rozgłosu nie zyskał jeszcze popularności
- użytkownicy obawiają się swobodnych, błędnych zmian ze strony innych
- w tej chwili w bazie danych jest 15 AP

The background of the slide features a photograph of a modern building with a glass facade, partially obscured by lush green trees. The image is overlaid with a dark, semi-transparent grid pattern.

Wykorzystanie

Po co?

- Dla pieniędzy...
 - lepsze AdWords i AdSense
 - marketing – skąd odwiedzają mnie klienci
 - lokalne wyniki wyszukiwania: usługi, firmy
 - ...
- Żeby świat stał się lepszy...
 - gdzie jest mój ukradziony laptop/telefon/aparat?
 - Samsung Mobile Tracker
 - Google Maps + obrazki → Panoramio
 - wyszukiwanie osób w portalach społecznościowych
 - Taggit, grono.net



Wspomagana geolokacja IP

Motywacja

- Ciągłe brakuje dokładniejszej geolokacji na podstawie adresu IP
 - jest Octant
 - skomplikowany algorytm
 - nieprzetestowany w większej skali
 - ustalenie lokalizacji celu zajmuje stosunkowo dużo czasu ze względu na sprawdzenie RTT, traceroute ze wszystkich maszyn testowych i obliczenie wyników
 - dwa spostrzeżenia
 - *przeważnie* nosimy ze sobą telefon komórkowy
 - *przeważnie* korzystając z sieci jesteśmy fizycznie przy komputerze za pomocą którego się łączymy

Co daje zwiększenie dokładności geolokacji IP?

- Użytkownik:
 - wyświetlanie reklam bardziej dopasowanych pod względem lokalizacji usługi lub firmy → większe prawdopodobieństwo kliknięcia
 - j/w ale dla uzyskanych treści
 - wiadomości, pogoda
 - wyniki wyszukiwania, ...
- Dostawca treści/informacji:
 - profil użytkownika rozpoznany w większym stopniu:
 - \$\$\$
 - analityka i statystyki

Pomysł

- w telefonie
 - pozycjonowanie BTS → to trochę obliczeń na podstawie danych które i tam mamy
 - co 20 minut zapisujemy te dane razem z aktualną godziną
 - będziemy wysyłać je do zewnętrznego serwera w dogodnym momencie
 - serwer na podstawie danych obliczy kolejne pozycje

Pomysł, cd.

- od strony PC:
 - serwer dostarczający treść użytkownikowi rejestruje aktywność → z jakiego IP i o której godzinie przyszło żądanie?
 - po otrzymaniu danych z telefonu serwer może na podstawie czasu korelować wyliczone pozycje z zarejestrowaną aktywnością

Ważne problemy i pomysły na ich obejście

- czasy żądań i danych z telefonu nie muszą się pokrywać → interpolacja
- jak łączyć żądania i dane z telefonu, gdy z systemu korzysta wielu użytkowników? → konta i rejestracja żądań tylko gdy użytkownik jest zalogowany, np. konta Google Account i Gmail
- adresy IP mogą być przyznawane losowo z puli → trzeba nadawać wagę uzyskanej lokalizacji na podstawie liczby uzyskanych korelacji, rozproszeniu lokalizacji, daty ostatniej korelacji, itp.

Zalety

- okresowe zapisywanie danych w telefonie nie wymaga obliczeń → wykonuje je serwer
- zwiększamy dokładność geolokacji IP
- ze względów pragmatycznych dostawcy Internetu często wydzielają klasy adresów IP dla konkretnych obszarów, np. osiedla → lokalizacją jednego adresu IP można przybliżać lokalizację drugiego będącego blisko w klasie
- użytkownicy za NAT-em zwiększają liczbę danych – potencjalnych lokalizacji dla zewnętrznego adresu IP → dokładniejsza geolokacja



**Dziękuję za uwagę :)
Pytania?**