Dude, where's that IP? Circumventing measurement-based IP geolocation

Phillipa Gill
Yashar Ganjali
Bernard Wong
David Lie

Presented By:
Sanaa Taha
University of Waterloo

Outline

- Introduction
- IP-to-Location Mapping Geolocation
- Measurement-based Geolocation
- Secure Geolocation Model
- Assumptions and Adversary Model
- Delay-based Geolocation
- Topology Aware geolocation
- Conclusions and Discussion

Introduction

- IP geolocation aims to solve the problem of determining the geographic location of a given IP address, to varying degrees of granularity.
- Benefits of Geolocation:
  - Online advertisements and search engines
  - Security sensitive APP: online content provider
  - Cloud computing

contributions

- Are current geolocation Algs accurate enough?
- How can adversaries attack a geolocation system?
- How effective are such attack? can they be detected?

IP-to-Location Mapping Geolocation

- Uses DB of IP-location mapping.
- DB can be public, administrated by regional Internet, (ARIN, RIPE)
- Proprietary DBs provided by companies such as Quova and Max-Mind
- The exact method of constructing theses DBs is not public

Measurement-based Geolocation

- Uses a set of geographically distributed landmarks hosts with known locations.
- These landmarks measures various network properties, such as delay.
- These results are used as input to the geolocation alg:
  - Geolocalization: constraining the region of the target (Ping for RTT, and Traceroute for router discovery)
  - Iterarive force directed algs
  - Machine learning
  - Constrained optimization

Secure Geolocation Model

- Geolocation user:
  - hopes to accurately determine the location of the target.
  - Assuming, has access to a number of landmarks and trusts their results
- Adversary
  - Owns the target's IP address
  - Wants to mislead the user into believing the target is at a forged location.
- Internet
  - Impartial to both
  - Introducing noise : queuing delays and circuitous routes

Assumptions and Adversary Model

- Simple Adversary:
  - Can tamper only with RTT measurements taked by landmarks
  - This can be done by selectively delaying pkts from landmarks (delay-adding attack)
- Sophisticated Adversary:
  - Controls several IP addresses and use them to create fake routers & paths to target

Assumptions

- The adversary is fully aware of the geolocation algorithm, and knows both IP addresses and locations of all landmarks used.
- The Adversary can’t compromise the landmarks, but can modify the network traffic that has network measurements
- Network measurements made by landmarks reach the target.

Delay-based Geolocation

- Each landmark, L_i
  - Pings all other landmarks → end –to end net delay
  - calibrates the relationship b/w geographic distance and network delay , fn_i: d_{ij} → g_{ij}
  - Pings the target IP
  - Uses fn_i to predict distance to target
- Assumption: network delay is well correlated with geographic distance
- Network delay: queuing, processing, transmission, and propagation delay (circuitous)

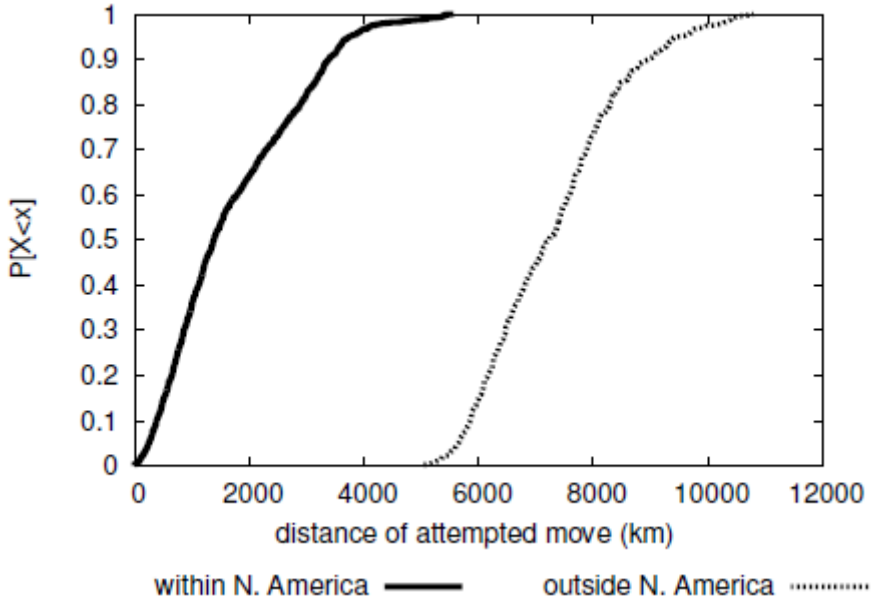
Constant-based geolocation (CBG)

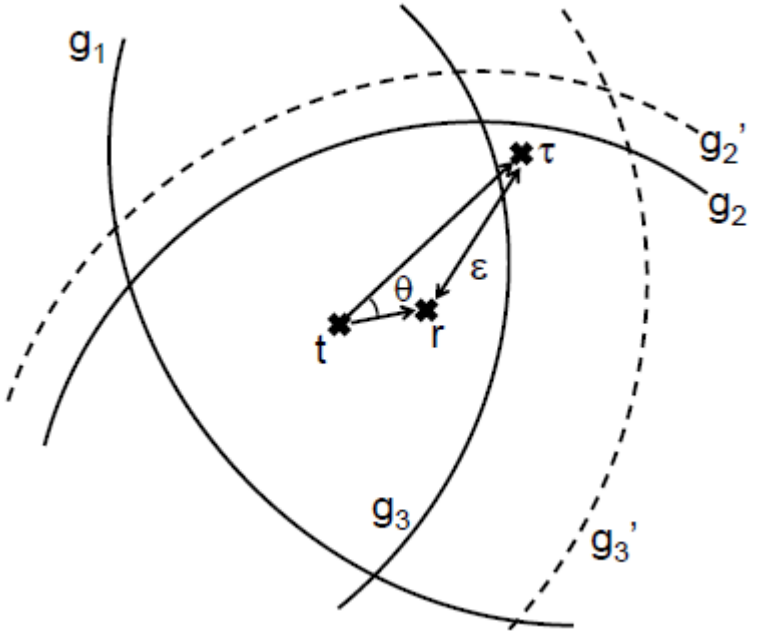
- Establishes fn by having a set of points (g_{ij}, d_{ij})
- Each landmarks computes a linear function, best line, that is closest to but below the set of points.
- Infer distance between each landmark and target IP using best line to get circle around each landmark.
- Get the region of intersection ($10^4 – 10^5$ km^2) and return the centroid of this region

Attack on CBG

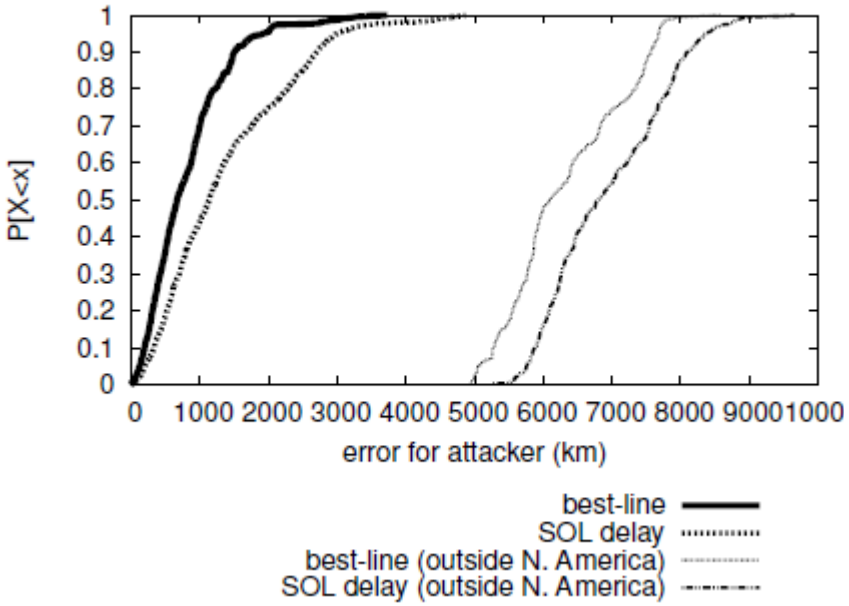
- Simple and sophisticated adversaries have equal power
- Alters the perceived delay d_{it} to $d_{i\tau}$
  - Find $d_{i\tau}$
  - Replace d_{it}
- Compute g_{it} and $g_{i\,\tau}$ by knowledge of landmarks locations $\gamma_i = g_{i\,\tau} - g_{it}$
- Using 2/3 the speed of light
- $\delta_i = \frac{2\gamma_i}{\frac{2}{3}c}$

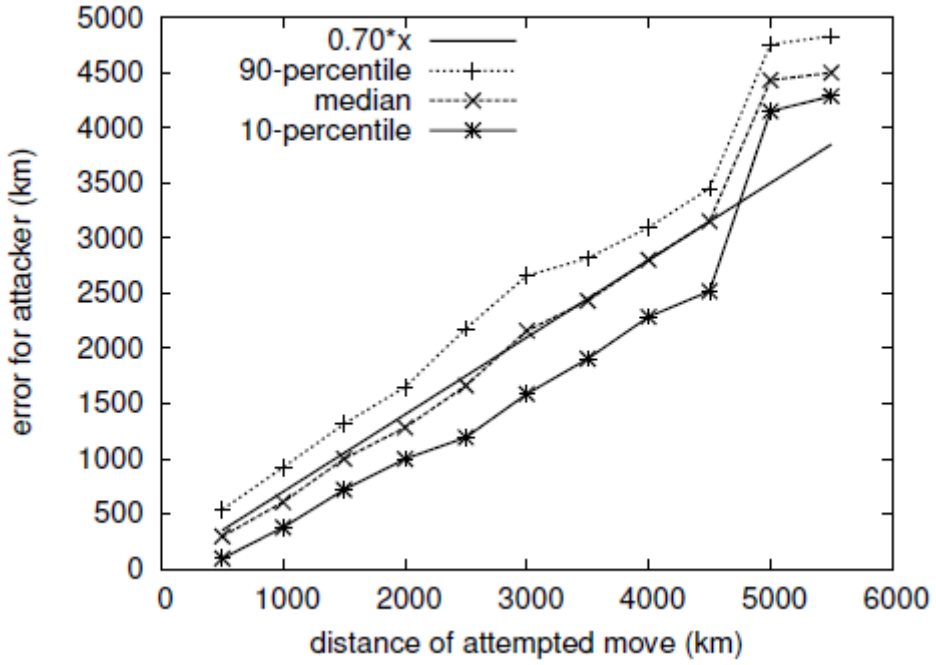
Evaluation


P[X<x]
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2000
4000
6000
8000
10000
12000
distance of attempted move (km)
within N. America
outside N. America

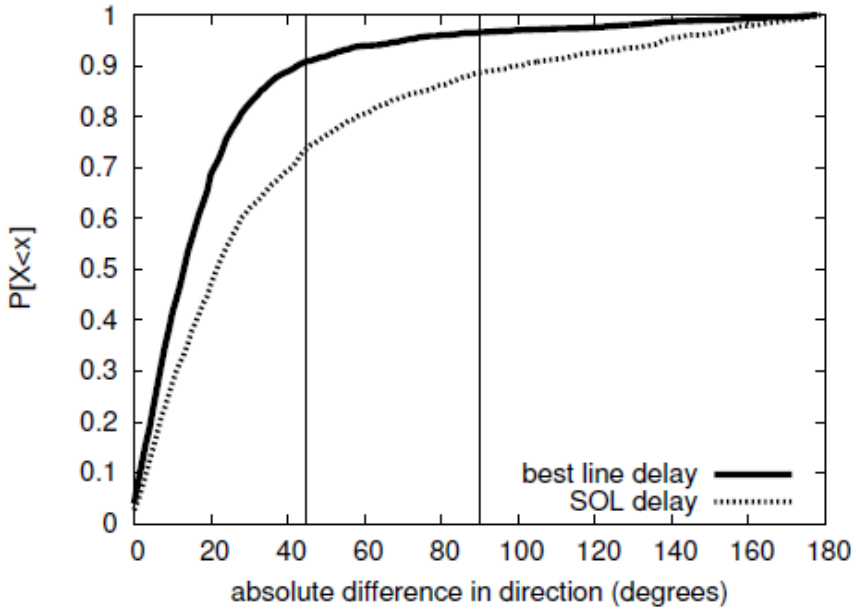

g1
g2'
g2
τ
ε
θ
t
r
g3
g3'

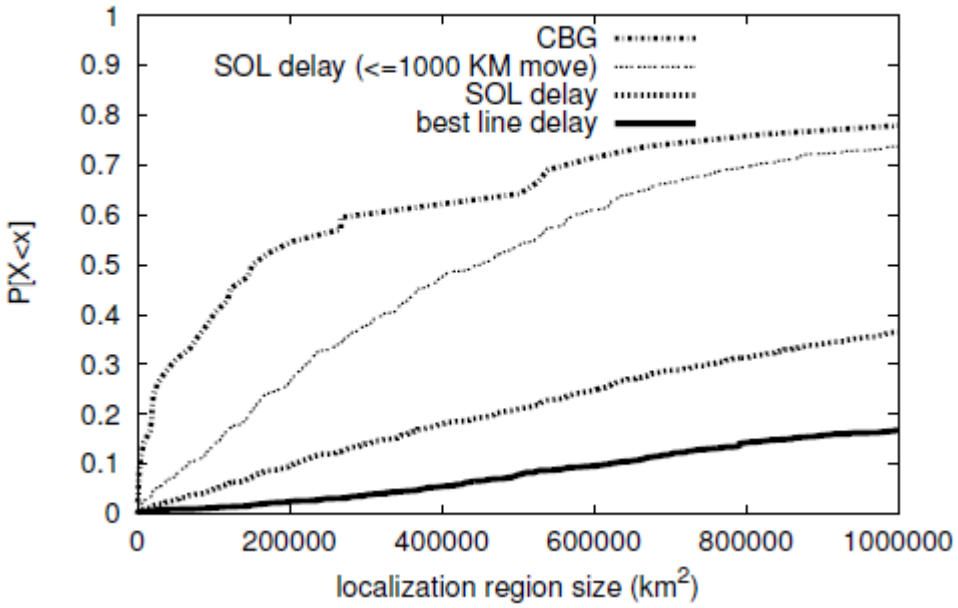
Attack Effectiveness





Attack Effectiveness and detectability





Topology Aware geolocation

- Limit the problem of circuitous end to end paths
- Localize all intermediate routers in addition to the target
- Single hop link delays are less circuitous than multi-hop end-to-end path.
- Topology based Geolocation (TBG)
  - Uses delays measured between IRs as input to a constrained optimization to know the locations of IRs and target IP
- Octant
  - Uses a geolocalization framework as CBG
  - Delays are mapped into distances using a convex hull

Delay-based Attacks

- Can not fabricate routers
- Any change made can only be reflected on the final links towards the target
- Most networks are connected to the Internet via a number of gateway routers
- Any path connecting nodes, outside the adversary's network, to the target (which is inside the adversary's network) will go through one of gateway routers

Delay-based Attacks

- Case1: If the net path from the landmarks to the target converge to a single common gateway router, then increasing the end-to-end delays between landmarks and thetarget can be detected
- $d_{it} = d_{ih} + d_{ht} + \delta_i$
- Observed latency from the gateway to target : $d_{it} – d_{ih}$

Delay-based Attacks

- Case2: If there are multiple gateway routers on the border of adversary's network, then increasing the delay between each gateway and the target can only be as effective against topology based geolocation as increasing end-to end delays against delay based geolocation.

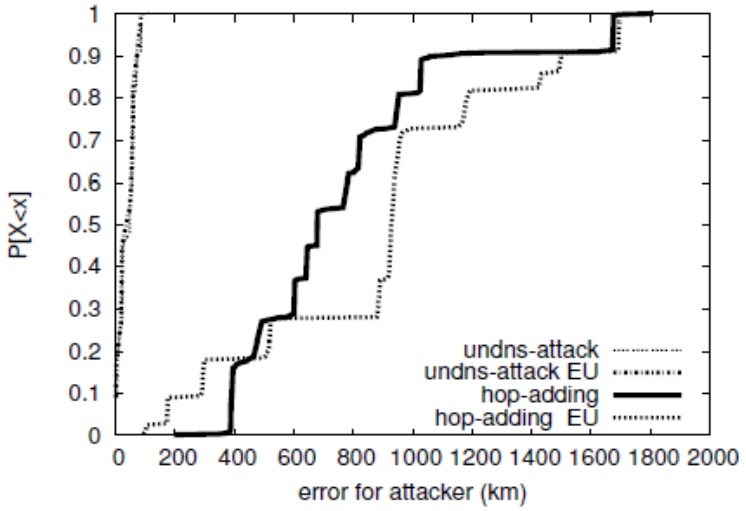
Topology-based Attacks

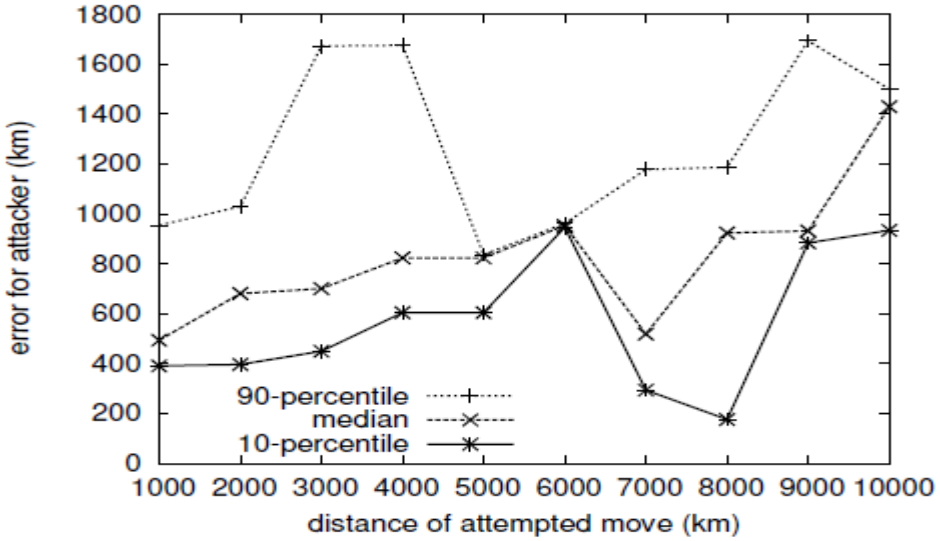
- Control a large administrative domain and can fabricate nodes
- Nodes in adversary's network:
  - Gateway routers
  - Internal routers: can be fabricated
  - End points
  - Links between internal routers can be manufactured
- Delay between fictitious nodes must respect the speed of light constraint

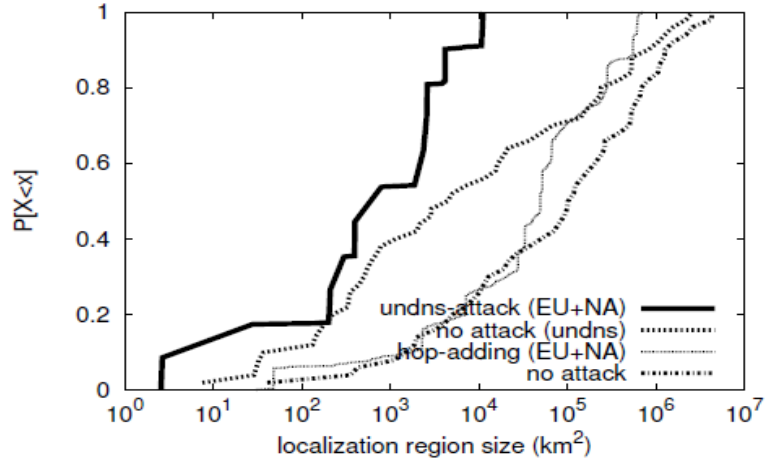
Topology-based Attacks

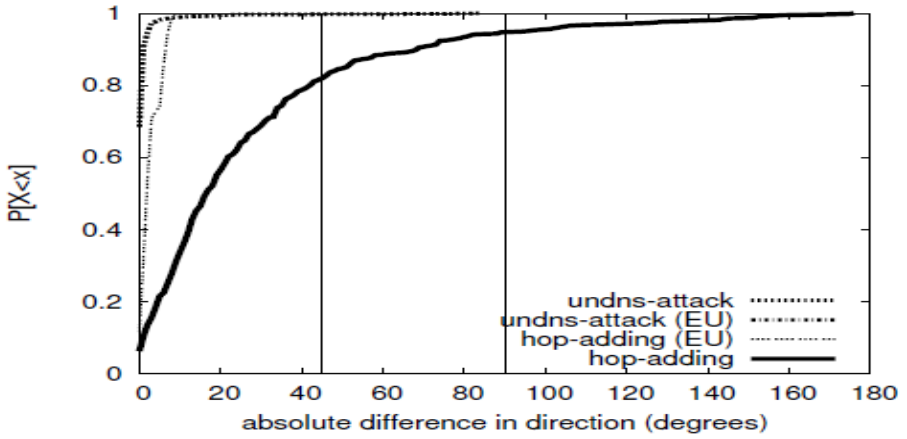
- Require the adversary to have more than one gateway router in its network
- An adversary with control over 3 or more gateway routers can move the target to any arbitrary location (geometric triangulation)
- Increase the circuitousness in the network when adding fictitious hops.
- Naming attack: when geolocation relies on undns, an adversary can effectively change the observed location of the target by crafting a domain name that can deceive undns tool.

Attack Effectiveness









Conclusion

- The current geolocation algorithms are accurate enough to locate an IP within a certain country
- Delay-adding attacks are easliy detected in delay-based geolocation and topology based geolocation schemes
- Hop-adding attacks are difficult to be detected

Discussion

- For MN with a fixed home IP address a variable foreign IP address, do you think that we can use geolocation methods to triangulate it?
- Do we need new methods for geolocation in heterogeneous networking?

Thank you

Any questions?