IP Geolocation Implementation for different points in an application request lifecycle

Prof. Pravin Nagare
Department of Computer Engineering
SND College of Engineering & Research Centre
Yeola, India
nagarepravin1189@gmail.com

Prof.I.R.Shaikh
Department of Computer Engineering
SND College of Engineering & Research Centre
Yeola, India
Imran.shaikh@gmail.com

Abstract —

IP Geolocation permits us to give a geographical location to an IP address allowing us to show the person behind that IP address. This can have many potential benefits for business and other types of application. The Internet Protocol of particular device is exclusive and for itself the location can be constricted down from the area to that country and even to the street address of that particular device. This technique of tracing can have very comprehensive outcomes and can sometimes only get an accurate result with some input from the user about their location. In some countries laws are in place that state a service can only track you as far as your country without your consent. If the user consents then the service can view your ISP's logs and track you as accurately as possible. The ability to determine the exact location of a person connecting over the Internet can not only lead to innovative location based services but it can also dramatically optimise the shipment of data from end to end. In this paper, we propose an efficient methodology for Reliability Analysis of Client-independent Internet Protocol Geolocation, which can be very helpful for evolution in Advertising, weather forecast, crime scene investigation, android mobile applications and many more.

Keywords: Geolocation, Geolocation Applications ; Geolocation Techniques, IP-Geolocation Implementation, Checkin-Geo

I. INTRODUCTION

Geolocation data, revealing an individual's physical location, are obtained using tracking technologies such as global positioning system (GPS) devices, Internet Protocol (IP) geolocation using databases that map IP addresses to geographic locations, and financial transaction information. Geolocation technology has been under development only since 1999, and the first patents were granted in 2004.[1] The technology is already widely used in multiple industries,[2] including e-retail, banking, media, telecommunications, education, travel,[3] hospitality, entertainment, health care, online gaming and law enforcement, for preventing online fraud, complying with regulations, managing digital rights[4] and serving targeted marketing content and pricing.[5] Additionally, the U.S. Federal Communications Commission (FCC) has proposed that geolocation software might be leveraged to support 9-1-1 location determination.[6] Uses of the information are myriad, including direct marketing and context-sensitive content delivery, monitoring of criminals, enforcing location-based access restrictions on services, cloud balancing, and fraud detection and prevention.

Geolocation technologies and their application, while offering social and economic benefit to a mobile society, raise significant privacy and risk concerns for individuals, businesses and governments. Tracking of targets from aerial platforms is an important activity in several applications, especially surveillance. Knowledge of geolocation of these targets adds additional significant and useful information to the application. Identifying the physical location of users by using devices that can passively or actively determine their location. As the correctness of geolocation technology has improved, there are more use cases for location-based networking than ever before. Conventional wisdom dictates that the Internet is a medium in which federalism is destined to fail. Virtually, the Internet naturally repels parameter by a diverse set of Government actors. Certainly, courts have reasoned that federalism on the Internet is either technically not possible or constitutionally forbidden.

The development in geolocation technology, which make it possible to rapidly, reason- ably, correctly recognize geolocation, challenges this is appreciative propose new approaches that could radically alter the way electronic commerce is governed. To illustrate this point, this

Essay explores the ways that such technologies could be used to make Internet gambling regulation more

responsive to longstanding federalism principles. As demonstrated below, geolocation technologies have the potential to make Internet gambling law both more effective and more efficient by enabling each state to enforce its own substantive regulations.
Geolocation integration can be accomplished at many different points in an application requests lifecycle. To realize the greatest value across the broadest spectrum of use cases, gathering of geolocation data is most efficiently accomplished when a request is rst made for a given resource. The Application Delivery Controller is typically deployed at a strategic point in the application and network architecture: at the perimeter of the network, acting as an inter-mediary between clients and resources. Given this strategic location, geolocation data should be incorporated into the existing context that is already associated with every request such as IP address.

IP geolocation plays a critical role in location-aware network services and network security applications. Commercially deployed IP geolocation databases may provide outdated or incorrect location of Internet hosts due to slow record updates and dynamic IP address assignment by the ISPs. Measurement-based IP geolocation is used to provide real time location estimation of Internet hosts based on network delays. This paper proposes a measurement-based IP geolocation framework that provides location estimation of an Internet host in real time. The proposed frame work models the relationship between measured network delays and geographic distances using segmented polynomial regression model and semidefinite programming for optimization. Weighted and non-weighted schemes are evaluated for location estimation. The proposed framework shows close to 17 and 26 miles median estimation error for nodes in North America and Europe, respectively. The proposed schemes achieve 70–80% improvement in median estimation error comparing to the first order regression approach for experimental data collected from Planet-Lab.

II. DEFINITIONS

Geolocation can be defined as a technology that uses information gain from specific PC or any type of radio or network-connection-enabled device to identify user's actual physical location. It is one of the most popular manifestations of the current development of information technologies and is recently experiencing a significant rise in popularity. When a GPS signal is unavailable, geolocation applications can use information from cell towers to triangulate the approximate position, a method that is not as accurate as GPS but has greatly improved in recent years.[7]
Technically meanings are various like GPS or satellite tracing, smartphones location, identifying locations of WiFi hot spot & identifying locations of IP addresses

III. RELATED WORK

Geolocation Techniques

1) Client-dependent Geolocation Techniques:
The Global Positioning System (GPS) has been widely used and it provides precise location information to any device with a GPS receiver. However, GPS is energy expensive for mobile devices (e.g., smartphones) and cannot be used indoors due to weak GPS signals. Skyhook [8], Place Lab and Google My Location scan the location information of cell towers and Wi-Fi access points all around the world (typically through a car), and estimate the location of a client through information broadcast from nearby cell towers or Wi-Fi access points. However, the cost of deploying cars for surveying location information is high and there is a debate on whether scanning these information is legal in some countries for privacy concerns. Moreover, all client-dependent geolocation techniques suffer from the following two limitations [9]. Firstly, these techniques require clients' support to report their locations to the server, which is not applicable for scenarios such as location based targeted advertising, context-aware security, location based access restrictions, and online service analysis. Secondly, there are many devices with only wired access [7]. and they do not have a capability of GPS, cellular, or Wi-Fi. The foundation for geolocation is the Internet protocol (IP) address, a numeric string assigned to every device attached to the Internet. When you surf the web, your computer sends out this IP address to every website you visit. IP addresses are not like mailing addresses. That is, most are not fixed to a specific geographic location. And knowing that a particular ISP (Internet Service Provider) is based in a particular city is no guarantee that you'll
know where its customers are located [10].

2) Client-independent Geolocation Techniques:

Database-driven Geolocation. Database-driven geolocation techniques try to build a database with huge number of IP/location mapping records, whose geolocation resources come from the Whois database [8], DNS [9], postal addresses from the Web [10], user contributions [11]and users' registration records [9].

Database-driven geolocations are widely used in commercial systems for their fast response time and easy deployment. However, the geolocation error is large since the geolocation resources are quite coarse-grained.

Methods		Geolocation Resources	Median Estimation Error (km)	Response Time (sec)	Deployment	
Delay Measurement Based Geolocation	GeoPing	Network Delay	382	Tens of seconds to several Minutes	Tens of Geographically Dispersed Servers	
	CBG	Network Delay	228			
	TBG	Network Delay & Topology	67			
	Octant	Network Delay & Topology	35.4			
	Wang-Geo	Network Delay & Topology & Postal Address from web	7.7			
Database-driven Geolocation	Existing Approaches	Whois,DNS,Postal Addresses from web, User Contributions	City Level	Negligible	Negligible	
	Proposed Approach	Result of our method & Login Logs	0.8			

TABLE I: A Compassion of Geolocation result & related work in IP Geolocation

Typically, database-driven geolocations can only provide a city-level geolocation [12], [13], which cannot meet the demand of precise geolocation for many location-aware applications. Delay Measurement Based Geolocation. Delay measurement based geolocation approaches estimate the geographic location of a target IP based on the network delays from known landmarks to the target. The rationale beneath these approaches is the positive correlation between network delay and geographic distance. Thus measured network delays can be converted to distances or distance constraints. In these approaches, a number of geographically dispersed servers are deployed to measure the network delays or routes from these servers to the target IP. Depending on the geolocation resources that these approaches leverage, they can be categorized into:
(i) estimating the location simply based on network delay measurements, e.g., GeoPing [9] and CBG [14]; (ii) combining network delay and topology measurements, e.g., TBG [15] and Octant [16]; and (iii) combining network delays, network topology and postal addresses from websites, e.g., Wang-Geo [6]. With contributions from these works, the median estimation error of delay measurement based geolocation has been reduced from the original hundreds of kilometers to under 10 km [6], [16], [14], However, delay measurement based geolocation techniques have never been widely deployed in reality for the following two reasons [12]. First, they need a number of geographically dispersed servers for probing and measuring the target IP. Thus the deployment is difficult. Second, these techniques usually suffer from large measurement overhead, with a response time ranging from tens of seconds to several minutes to localize a single IP [12], [14]. For example, Wang-Geo [6] needs a response time of 25.9 seconds on average. Compared with existing client-independent geolocation techniques, this method proves the following advantages:
(i) Better geolocation accuracy with an order of magnitude smaller than the state-of-the-art;
(ii) Fast response time; &
(iii) Easy deployment. Table I shows the detailed comparison among this method and related works in IP Geolocation.

B. Studying and Mining Location Data
With the increasing popularity of location-aware devices, a considerable amount of research efforts have been attracted to study location data in recent years. Leveraging the movement trajectories sampled at high frequency from volunteers, researchers have tried to predict users' future activities [17], infer people's transportation modes [18], and identify semantic regions associated with users' activities [19]. Based on the observation that phone call records contain both the time and the associated cell tower ID of each call event, Isaacman et al. [20]propose an algorithm to identify important locations for users, and Cho et al. [21] develop a model to predict the locations and dynamics of future human movements. In particular, Isaacman et al. [20] design a logistic regression to identify important locations including home/work for a user. Different from [20] which focuses on the exact single home or work location, we are interested in all the potential home/office location candidates. Experimental results show that the candidates produced by our method based on checkins cover home/office locations for 98% of users with an accuracy of 2 km, while [20] estimates the exact home/work with 95th percentile error of 3.86 miles (home) and 21.23 miles (work).

PROBLEM STATEMENT

Statement 1. (Find records of data & finding History).
Records of data rd is set of three (x,y,z) i.e clients c shares their location x (normally using a smarthphone) at time y and transmits a message z. The finding history rd is the set of data records of client d, i.e., Cu = {cu}.

Statement 2. (Login Record & Login History).
A login record lr is set of three (w,y,uc) i.e client c logins to a website (generally using a Personal Computer) at time y with IP address o and client-agent uc. A login history lr is the set of login records of user c, i.e., lr = {lr}.

Statement 3. (Internet Protocol Set).
We define all the login Internet Protocols for a given client set C as OC = {lr.o|c C}, and a more Internet Protocol range whose geo-locations we can write as O 0 C , OC O 0 C .

PROBLEM.
Given a user set U, and the corresponding finding histories {rd|c C} and login histories {Lr|c C}, estimate the geo-location of each Internet Protocol in O 0 C.

BASED ON LOCATION-SHARING SERVICES

We know that, a client is near his house at night and in working hours near to his work place, we gather checkin histories at night and work hours respectively to find the home or office locations for each user. Using users with known home or office locations as testing data, we demonstrate that our gathering method can find users' home and office locations accurately. But, In check-in geo-location we can use tentative House or working place location to find mapping from various locations to Intenet Protocol Addresses which was used by client in his Personal Computer.

METHOD BASED ON TIMESTAMPS

For every client, we used structural approach to find geo-location in night [from 8:00PM to 7:59AM everyday] and during working hours [from 8:00AM to 6:59PM], respectively. After collecting all results, we found possible location

Statement 4. (Home-based Client & Office-based Client)
A home-based Location of client group c,h from this one group is result from structural method on the location at night hours of client c. A home-based client is xc,h = (p,n) that means p is center of these locations in group of xc and n is no. of location results in group xc,h. The home-based candidates for a client c are Xc,h = {Xc,h}. likewise, client at work place is Xc,o=(p,n) and & for office-based clients are c are Xc,o={Xc,o}.We state to Xc,h whose location reference number is in I th rank in downward order in Xc,h,I. Likewise, XC,o whose location reference number is in I th rank in downward order in Xc,o which referred as xc,o,I.
Cases like this, if the distance between two locations is more than 150 m then we abort clustering; Else the nearest location point is aggregated in new cluster. After clustering, we can get the home-based clients Xc,h & office-based clients Xc,o for each client c.

EFFICACY EVALUATION

In this method, we can easily find the geo-location by the observation that a small subsets of clients who claimed their exact location in the message which he/her send. E.g if the user post message 9-10 times that he/she at home or office. Case like this, we can assume rd.g is the home location of his/her. To find result on this method there are some requirements:

i) User has shared his/her location 9-10 times.
ii) All the home-based locations client has shared his/her location in 100 meters distance.

BASED ON LOGIN LOGS FROM PCS

In this section, we first analyze users' login logs and use a classification model to tag whether an IP address is from the corresponding user's home or office. Thus we get the mappings from each user to the corresponding home and office IPs. We then employ a /24 segmenting method to increase the IP coverage, and get the mappings from each user to the corresponding home and office IP segment. Although here we primarily identify home and office IPs for IP geolocation, the technique in this section itself can also be used to customize search results and to target advertisement based on a user's context (home, office, travel, etc) associated with an IP.

IV. IMPLEMENTATION

The activity to find out the exact geographic location of IP hosts is significant for location- receptive applications like Internet advertising, Cyber-attack detection and faults diagnosis. Even though their quick feedback, regular used database-driven geolocation methods provides incorrect locations.

The Application Delivery Controller is typically deployed at a strategic point in the application and network architecture: at the perimeter of the network, acting as an intermediary between clients and resources. Given this strategic location, geolocation data should be incorporated into the existing context that is already associated with every request such as IP address, user-agent and ability to accept specific

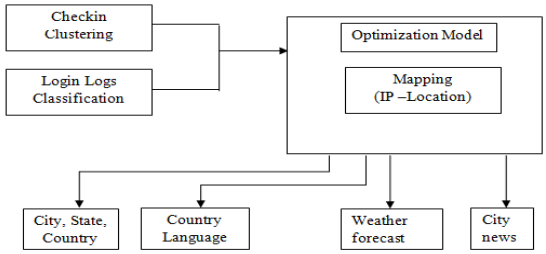
types of content. I focusing on two methods of Chekin-Geo to cover as much larger distance so that network delay is also avoided and no. of servers are less as well The methods of Checkin-Geo which we will combine in our web service/app are :-

Chekin from mobile

Log in from PC's

Summarization model of both data.

I will use Database-driven Geolocation technique and Delay Measurement Based Geolocation technique to minimize my delay and no. of servers to spot latitude longitude.Using these two techniques will lead us to optimal utilization of system resources and time delay would be less for a remote areas. Correlation of network delay geographic distance is achieved which will be of great help in de?ning degree of remoteness of a location.



1. Based on Smartphone Application
Devices like smartphones, whether they are different platform based on Operating systems, are extremely able to check-in geo location data. Smartphone applications depend on such geo data for targeting of advertisements; and many platforms will be making use of the mass of data being developed from smartphones.

2. Based on public wifi
While, identifying geolocation from public wifi is not much accurate because the range of these devices are less. As any device is connected to these hotspots they will identify.

3. Based on Internet Protocol
Laptop or Personal Computer Users are normally uses Internet through wired or wireless connection & are typically inside home. The concluding guidelines out GPS location and neither networks are through cells. In this, to find exact geolocation IP Address is used. An IP address contains no geographic info whatever and ISPs modifies IP addresses manually. Though, blocks of IP addresses are applied to areas and records kept of where addresses are applied within these areas. Also there is a high degree of consistency over who is given which IP address, right down to the individual address level.

4. Based on Application Programming Interface
There are Geolocation Application Programming Interfaces which is inbuilt in web browsers. With the help of API it is very easy to collect geographical info based on what user surfing on internet.

5. Based on Social Media Sites
Sharing location on Social media sites like Facebook, Linkedin & other site tells person where he is. With help of browser-Cookies advertisers can get information about interest, purchases, and enquiries from source computer. In summary, without users' permission – with help of these technique a large amount of info like from where this person is accessing internet, what he doing, etc.

Some unusual examples of use of Geolocation data:
1. A car rental company started using deployed GPS tracking devices to monitor driving speeds of its customers. If a customer's car exceeded 79 miles per hour for 2 continuous minutes, they were charged an additional $150 (without their consent). [2]
2. A French insurance company used both mobile phone and GPS data to track sales executive locations and cross-reference their expense accounts. This policy resulted in 21 employee dismissals and the identification of over half a million euros in false claims! [3]
You might think that the use of geo data in these examples was necessary, but the same methods could be u in a mass way and for the wrong reasons.

Proposed Applications:
1. Languages change as Country changes, but if the content is static it becomes a pain to retrieve useful information. Hence the content will dynamically be presented to the user which will not hold language as a bound to gain useful information. Content favorable for specific location be arranged such a way that it will grab attention of the user.
2. Also web service will suggest what people near his location viewed recently to interest user in viewing more articles.
3. Web service will also facilitate user with weather forecast based on their location and on anonymous location as well.

4. Public interaction will prove a vital input through various social media eg. (Facebook, Twitter) which will be of great help to government officials in crime scene investigation or to cover some major event placed in that Location.

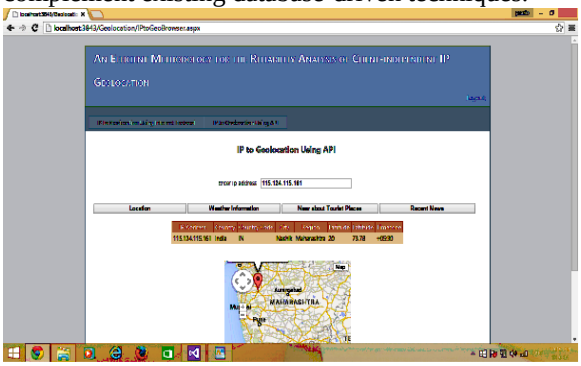
V. DISCUSSION

Limitations:
Compared with existing approaches, this method achieves both high estimation precision and fast response time, and thus it is ready to be deployed at large scale for precise location-based applications, by replacing/complementing existing widely used database-driven geolocation techniques.
However, the IP coverage of method depends on the data we use and typically cannot cover all IPs. In this case, we can combine this method and existing database-based on geolocation techniques to provide a city-level precision for IPs that proposed method cannot cover.

Data Collection:
This method needs to PC login logs from the same users. Collecting these data is not difficult given that most social network services (e.g., Twitter, Facebook, and Foursquare) provide both a mobile (app or web) and a PC version (typically through web) service. Moreover, most location-sharing services allow users to share result to other social networks. For example, result in Foursquare can be shared to Facebook and Twitter. In this case, both result and login logs can be collected from different services and associated through the same users. For example, Facebook can use its own login logs and result cross-posted from Twitter, Foursquare or Facebook itself. Note that account links among different services for a user are usually publicly posted on the user's profile page in social network services
Engineering for his en-couragement. I thank him for providing the required resources from the college. I am equally indebted to the supporting staff members of department of Computer Engineering of SND College of Engg. and Research Centre who have helped me directly or indirectly.
Alternatives for IP/Location Mapping:
We map IPs to locations through the transition of home/office locations. In fact, we have tried other mapping alternatives. However, initial experiments show that these mapping alternatives are not promising. For example, we have tried to use result with a time close to a PC login time for estimating the corresponding

CONCLUSION

In this paper, we propose method, to find an accurate IP geolocation which exploits result from location sharing services and login logs from PCs, fundamentally different from existing approaches. Both our experimental results and commercial deployment show that this method achieves fine-grained geolocations. Although the precision of this method is impacted by 1) the density of geolocation result and 2) the mechanisms Internet Service Providers use to allocate IPs, our large-scale deployment in Tencent including almost all cities in China where the two factors change in a large range shows promising estimation precision. It should also be noted that although result of geolocation initially leverages home/office IPs, it covers many more IP addresses in addition to the intuitive home/office IPs after IP segmenting (the IP range of an airport is the office IPs of employees in this airport, the IP range of an cafe can be used by nearby residents or employees as they are using the same IP /24 segment, etc.). As we can see in following snapshot, the result of geolocation demonstrates good scalability as it is only a computation intensive approach so that we can speed up the tasks by parallelization and update the IP/Location mappings at high frequency (every day in our deployment as described in, which makes Geolocation a commercial-ready technique to complement existing database-driven techniques.



On the contrary, delay-measurement based approaches are difficult to be used commercially as they 1) require to deploy widely dispersed servers and 2) cannot be parallelized at large scale so that the measurements cannot be done offline as the computation in this method.

Acknowledgment
I owe my deepest gratitude and indebtedness to my highly respected and esteemed guide Prof.I.R.Shaikh, Department of Computer Engineering, SND College of Engg. and Re-search Centre for being a continuous source of inspiration, guidance, sincere criticism and regular feedback on my advancement each week I

was able to learn and improve upon my working strategy and thus, enhance my approach during this piece of work. It was because of her stimulating suggestions and comments that I could identify the right topic and always keep in the right track. I would like to take this opportunity to express my sincere thanks to our P.G.Coordinator Prof.K.P.Gaikwad and HOD Prof.S.R.Durugkar of Computer

REFERENCE

[1] "Digital Envoy wins Geotargeting Patent". 2004-06-29. Retrieved 2010-04-19.

[2] ClientSideNews, Nov/Dec 2010 issue, Page 6 "You Can Really Do That? – The Power of Geolocation Technology"

[3] Marketing Magazine, February 10, 2011 "Thetrainline brings Digital Element on board for localised ad task"

[4] Music Streaming site we7 Cranks up the Volume for Digital Rights Management and Ad Targeting

S. S. Intille, K. Larson, E. Tapia, et al. Using a Live-in Laboratory for Ubiquitous Computing Research. In Pervasive, 2006.

[5] "Geolocation and Federalism on the Internet: Cutting Internet Gambling's Gordian Knot". 2009-10-

[04] Retrieved 2010-01-02.Google Maps with My Location.http://www.google.com/mobile/gmm/mylocation/index.html .

[6]http://transition.fcc.gov/Daily_Releases/Daily_Business/2011/db0712/DOC-308377A1.pdf

[7] Ionescu, Daniel. "Geolocation 101: How It Works, the Apps, and Your Privacy". PCWorld. Retrieved March 29, 2010.

[8] APNIC - Query the APNIC Whois Database.http://wq.apnic.net/apnic-bin/whois.pl.

[9] V. N. Padmanabhan and L. Subramanian. An Investigation of Geographic Mapping Techniques for Internet Hosts. In SIGCOMM,2001

[10] C. Guo, Y. Liu, W. Shen, H. J. Wang, Q. Yu, and Y. Zhang. Mining the Web and the Internet for Accurate IP Address Geolocations. In INFOCOM, 2009.

[11] Host IP. My IP Address Lookup and Geotargeting Community Geotarget IP Project. http://www.hostip.info.

[12] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye. IP Geolocation Databases: Unreliable? SIGCOMM Comput. Commun. Rev. vol.41, no.2, pp.53-56, 2011.

[13] GeoIP City Accuracy for Selected Countries. http://www.maxmind.com/en/city accuracy.

[14] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida. Constraint- Based Geolocation of Internet Hosts. IEEE/ACM Transactions on Networking, vol.14, no.6, pp.1219-1232, 2006.

[15] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards IP Geolocation Using Delay and Topology Measurements. In IMC, 2006.

[16] B. Wong, I. Stoyanov, and E. G. Sirer. Octant: A Comprehensive Framework for the Geolocalization of Internet Hosts. In NSDI,2007.

[17] D. Choujaa and N. Dulay. Predicting Human Behaviour from Selected Mobile Phone Data Points. In UbiComp, 2010

[18] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Ma. Understanding Mobility Based on GPS Data. In UbiComp, 2008.

[19] C. Lu, P. Lei, W. Peng, and I. Su. A Framework of Mining Semantic Regions from Trajectories. In DASFAA, 2011.

[20] S. Isaacman, R. Becker, R. Caceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Identifying Important Places in People's Lives from Cellular Network Data. In Pervasive, 2011.

[21] E. Cho, S. A. Myers, and J. Leskovec. Friendship and Mobility:User Movement in Location-based Social Networks. In KDD,2011.