RESEARCH

Open Access

CrossMark

A landmark calibration-based IP geolocation approach

Jingning Chen[1,2], Fenlin Liu[1,2], Xiangyang Luo[1,2*], Fan Zhao[1,2] and Guang Zhu[1,2]

Abstract

Aiming at the existing IP geolocation approaches does not consider the errors of landmarks and delay; a new geolocation approach-utilized landmark calibration is proposed in this paper. At first, we find out these landmarks shared the nearest common router with a target IP by path detection; second, a deviation is assigned to each related landmark according to the corresponding organization and network connectivity; then, while the landmark's location is regarded as the points within a possible area, target IP geolocation can be converted into a constrained optimization problem; at last, we can get the location estimation of the target IP by solving the above problem, as well as the real deviation of each landmark. The algorithm analysis and experimental results show that, when a landmark is not located in its claimed position, our geolocation approach can give a location for the measured target IP, as well as the location of the nearest common router for the unmeasured target IP.

Keywords: IP geolocation, Landmark calibration, Relative delay, Nearest common router, Optimization problem

1 Introduction

IP geolocation refers to use the corresponding IP address to determine the location of a network entity in some level of granularity [1, 2], such as finding out which city is a host with public IP located at. According to the evolution of location-based service (LBS), IP geolocation-based applications are more and more popular, such as, targeted advertising according to the users' locations, adjusting the language on the site by ISP automatically according to the clients' regions, and developing the deployment strategy of the network infrastructure, and discovering fault nodes, specifying the geographic region (a city, state, time zone, or political boundary) of a cloud service, is increasingly common, and geographic region options are provided to help customers achieve a variety of objectives, including performance, continuity, and regulatory compliance [3, 4]. More importantly, IP geolocation can play a significant role in network security, such as tracing cyber fraud and attacks and extracting the system logs for computer forensics. Therefore, IP geolocation is widely applied in Internet commerce and security, as well as the cloud service.

IP geolocation was openly discussed since 2001 [1], and with a decade of development, there are a variety of geolocation methods, including GeoTrack [1], CBG (constraint-based geolocation) [5], shortest ping [6], TBG (topology-based geolocation) [6], Octant [7], SLG (street-level geolocation) [8], and LBG [9]. There are also a number of researches on the related key technologies [10–12]. Among the above methods, most of them do not consider the error of landmark (an entity with known geographical location and stable IP identifier). For instance, after getting a coarse-grained estimation area for the target IP based on CBG method, SLG gives a fine-grained location to the target IP using the relative delay between landmarks with known positions and target IP. Obviously, for the SLG method, the geolocation accuracy depends heavily on the following three conditions: (1) there are landmarks around the target IP: usually, the landmark near the target IP is likely to get accurate constraint and geolocation results; (2) the positions of related landmarks are accurate: at the fine-grained positioning stage, the landmark is used as the probe point, and when the real position of landmark is inconsistent with a claimed position, the geolocation approach cannot guarantee the real location the target IP is located in geolocation area (at the coarse-grained positioning stage, whether the location of the probe point is accurate or not will affect the geolocation result too); and

* Correspondence: luoxy_ieu@sina.com
[1]State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China
[2]Zhengzhou Science and Technology Institute, Zhengzhou, China

Springer

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 2 of 11

(3) the delay can be measured accurately: while the delay from the probe point to the target IP and landmarks is inaccurate, the relative delay between landmarks and the target IP cannot be estimated accurately, and then it is difficult to ensure the accuracy of the geolocation result obtained by relative delay.

For the above three conditions, while the first one can be met through the way of active deployment and passive acquisition [7, 13, 14] to obtain more landmarks, the second and third ones are taken as a default condition in most studies, and there is no solution focused on geolocation with inaccurate landmark. In response, a new geolocation approach based on landmark calibration is present in this paper, and by taking the error of landmarks and delay as the deviations of landmarks, our approach can be used to estimate the location of a target IP while the landmarks and delay are inaccurate. The experimental results show that the geolocation accuracy is better than the SLG approach.

2 Related works

Most geolocation approaches, such as GeoTrack, CBG, TBG, and Octant, take PlanetLab nodes [19], public ping (traceroute) servers, and deployed points as probe points and landmarks, so they do not need to consider the accuracy of landmarks. In order to steadily reduce the estimated area for the target IP, SLG finds organizations with known domain names, IP addresses, and postal addresses based on Web mining and uses shared host recognition, DNS name resolution, IP address reverse query, and other strategies to evaluate the reliability of those organizations, and reserved organizations can be used as landmarks. However, after the above verification, there are still some inaccurate organizations, for instance, if a company with no branch has a proprietary website, and the corresponding server is not located at the geographical position claimed on the website.

For the IP geolocation system, when using a landmark's location, it usually refers to the latitude and longitude of this landmark. However, the latitude and longitude is the postal address of an institution corresponding to a landmark, not the postal address of the entities corresponding to a landmark. Due to a larger coverage or rough postal address of an institution, there will be deviation between the landmark's claimed position and the real position. Taking the new campus of Zhengzhou University as an example, the covering range of this institution shown in Fig. 1 with the red line is about three million square meters, and the postal address is "100# of science Road, Zhengzhou city, Henan Province," while the corresponding latitude and longitude is (34.822975, 113.542962), as shown in Fig. 1 with the red cross. In fact, the corresponding Web server of this university may not be located in this position. Dengfeng second refractory material Co Ltd

is another example, while the postal address is "TaMiao in Dengfeng City," the identification on the map is shown in Fig. 2. Therefore, when the above network entities are used as landmarks to high-precision geolocation, we must consider the deviation between the real position and the claimed position.

It is clear that delay measurement is susceptible to network load and link intricacy and other factors. Whether the packet from a probe point to a common router and the one from a probe point to a landmark share the same path between the probe point and the common router or not will affect the accuracy of delay between the landmark and the common router. As shown in Fig. 3, denote R_i as the common router between two landmarks (L_1, L_2), and denote t_1 as the delay from a probe point P to L_1 and t_i as the delay from P to R_i. If the measurement packets for t_1 and t_i share the same path between P and R_i, the delay between L_1 and R_i can be estimated by $t_1 - t_i$, denoted as t_{i1}. If there is another path between P and R_i, and the delay from P to R_i is just measured through this path, denoted as $t_i^{'}$, then t_{i1} should be $t_1 - t_i^{'}$. However, in the process of delay measurement, we cannot know the real path of the measurement packet, and different paths from P to R_i will introduce a different error into the estimated delay between L_1 and R_i.

Denote R_1 as the claimed distance, i.e., the distance between the claimed location (L_c) and the target, and denote R_2 as the measured distance between the landmark's real location (L_r) and the target. So the geolocation area of the target deserved from the landmark is a circle centered at L_c, and the radius is R_2. In addition to the size of R_1 and R_2, SLG analyzes the landmark's influence on the geolocation result as four conditions: (1) the landmark is accurate: $R_1 = R_2$; (2) the target is farther apart the landmark's real location: $R_1 < R_2$; (3) the target is a little farther apart the landmark's claimed location: $R_1 > R_2$: and (4) the target is far away from the landmark's claimed location: $R_1 >> R_2$. In the first three cases, the geolocation region produced by the landmark can include the real location of the target, and it means that the inaccurate landmark will not affect the accuracy of the geolocation result. While only in the fourth case, the target will fall outside of the estimated area; SLG thinks that the probability of the occurrence of the fourth case is very small, and the probability that an inaccurate landmark can get an accurate geolocation result is no less than 1/2 [15].

It can be seen that the existing geolocation approaches often give algorithms and results with accurate landmarks and delay. Eexcept that SLG points out that the inaccurate landmark will affect the accuracy of geolocation results, there is almost no study related to geolocation approaches which can give geolocation result with inaccurate landmarks and delay. In fact, except taking the probe points

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 3 of 11

Fig. 1 Coverage and claimed position of the new campus of Zhengzhou University

Fig. 2 Rough claimed position of Dengfeng second refractory material

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4
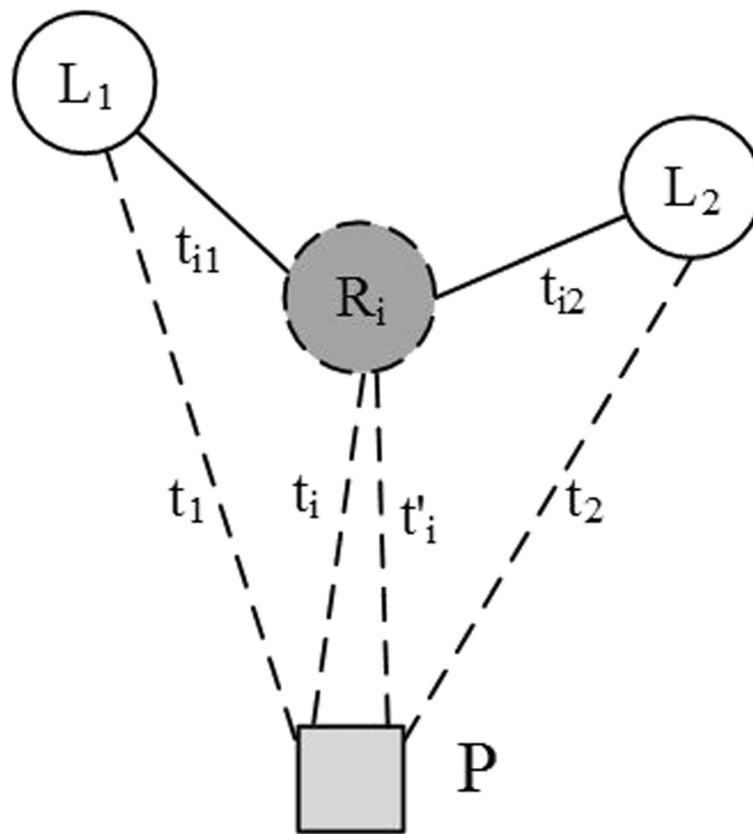
Page 4 of 11

**Fig. 3** Error of estimated delay between landmark and common router

which are deployed actively as landmarks, other landmarks derived from Web mining and geolocation databases may be not accurate. At the same time, it is difficult to eliminate the measurement error of estimated delay between the landmark and common router. In addition, link intricacy can affect the geolocation result too. While the estimated delay between the landmark and the common router is inaccurate, it means that the estimated delay is larger or smaller than the real delay. Combined with the known conversion coefficient of delay and distance, the estimated delay between the landmark and the common router, and the landmark's location, the distance constraint of a common router can be obtained, and this con constraint will larger or smaller than the distance between the landmark and the common router.

This paper argues that, because it is difficulty to verify the accuracy of the landmarks and delay, taking the landmark as a point located at its claimed position is inconsistent with the actual situation. Taking all possible error of landmarks and delay as the deviation of a landmark, so, instead of a point, we use a possible area to indicate the landmark's position, and then a new geolocation approach to estimate the target's location based on landmark calibration is proposed.

3 The principle of landmark calibration

The principal principle of the proposed approach is shown as the following: a landmark is no longer a point corresponding to the claimed location, and by introducing a deviation into the landmark, we can get a circle centered at a claimed location. And the radius is the deviation, and then the landmark is supposed located in the possible region covered by the circle. So it means that the real location of the landmark is the point in the circle. When geolocating the common router, taking a point from the possible region as the landmark's location according to a certain strategy, a possible location of the common router can be obtained combined with an estimated delay between the landmark and router. A set of landmarks correspond to a set of sampling points and a possible location of the router. While different sampling points of landmarks result in another possible location of the router, the geolocation result is the union of all possible locations. The schematic diagram of the proposed approach is shown in Fig. 4.

The proposed approach also needs paths from a probe point to landmarks and target IP, as well as the nearest common router and the related landmarks. Those steps are the same as the existing approaches (such as SLG),

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 5 of 11

Fig. 4 The proposed approach
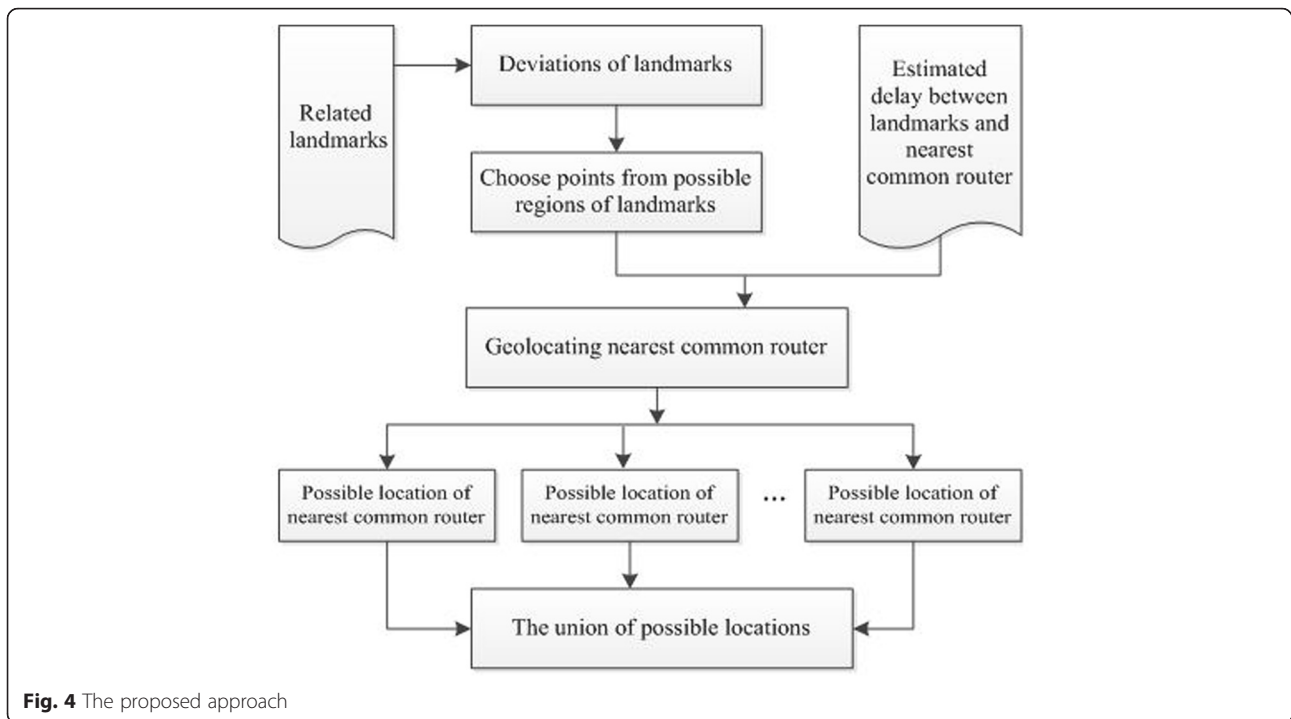
so we do not give more details about them, and after finding out the nearest common router, the following steps of our approach are shown as below:

Step 1: estimating the landmark's deviation. For the landmarks connected to the nearest common routers, assign a deviation to each landmark according to the corresponding organization and the network connectivity of the landmark, then we can get a circle which uses the claimed location of the landmark as the center, the deviation as the radius, and the covered area of this circle is the possible region (denoted as D) of this landmark.

Step 2: sampling the point of the landmark's location. For a landmark L, choose a point from D (obtained from step 1) as the landmark's real location according to a certain rule. For instance, take the claimed location as the center and 1 km as radius; choose one point in this circle for each 36°, and then ten points can be obtained for one circle. Increase the radius successively, such as 2 and 3 km, and the radius must be smaller than the deviation. All those chosen points are the sample points of the landmark's real location.

Step 3: geolocating the nearest common router. Taking the sample points as the landmark's location, combined with conversion coefficient of delay and distance, and delay between the nearest common router and landmark, the distance constraint from the landmark to the router can be calculated, as well as a possible location of the router.

Step 4: calculating the possible region of the nearest common router. The union of all locations obtained

from step 3 is the geolocation result of the router, and the smallest region that can cover all possible locations is the possible region of this router.

4 The implementation of landmark calibration

The implementation of the proposed approach can be divided into coarse-grained and fine-grained positioning, and in these two kinds of geolocation ways, there are two situations: the measured and unmeasured target IP. According the steps described in the last section, the possible region of the nearest common router can be obtained with inaccurate landmarks and delay. For the measured target IP (the delay and path from a probe point to a target IP can be detected), using the nearest common router as the landmark, we can calculate the geolocation result of the target IP. For the unmeasured target IP, we take the possible region of the nearest common router as the estimated geolocation result of the target IP; this is because in the real Internet environment, the entities are usually close to the last-hop router, and then target IP can be supposed located in the possible region of the nearest common router.

4.1 Coarse-grained positioning

In this paper, coarse-grained positioning is the calculation of a region-level geolocation result for the target IP. In the second geolocation tier of the SLG approach, taking 4/9 C (C is light speed) as conversion coefficient of delay and distance, combined with the relative delay between the landmark and target IP, the distance between

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 6 of 11

the landmark and the target can be calculated, and then we can get the coarse-grained geolocation result of the target. While the landmark is not located at its claimed location, the geolocation result of the SLG approach will be inaccurate, and the target may not be located in the geolocation region.

If the landmark is inaccurate, it is no longer located at the claimed location, taking this landmark as a region (denoted as PR). When calculating the geolocation result for the target IP by this landmark, instead of a circle which uses the claimed location of the landmark as center and distance constraint as the radius, the corresponding geolocation region is a set of circles with the same radius, and the centers are the points located at PR. For instance, there are three landmarks, denoted as A, B, and C, and the corresponding region is PRA, PRB, and PRC, respectively. While geolocating the target IP by those landmarks, choose one point from PRA, PRB, and PRC, respectively, and take those three points as centers, and then, combined with the conversion coefficient and relative delay between the three landmarks and the target, calculate the three distance constraints from the landmarks to the target IP. Take those three constraints as radius, with three centers and three radii, then the intersection of the three circles is the geolocation region of the target IP. While assigning different points of PRA, PRB, and PRC to A, B, and C, respectively, an intersection can be obtained, and the union of all intersections is the final geolocation result.

4.2 Fine-grained positioning

Fine-grained positioning refers is the calculation of a point decided by a pair of latitude and longitude, while a small region of the target IP is known. According to the estimated delay between landmarks and the target, the geolocation approach based on landmark calibration can take the landmark as the probe point. After introducing the deviation to an inaccurate landmark, IP geolocation is converted into an optimization problem, and the target's location can be achieved by solving this problem. Therefore, the approach includes three steps: delay estimation, converting delay to geographic distance, and solving the optimization problem.

4.2.1 The delay estimation

For the delay between two network entities, it usually refers to the RTT (round trip time) from a source to a destination, which is composed of four types of delay, such as transmission delay, propagation delay, processing delay, and queuing delay. Transmission delay is the time needed to place a packet onto a link; propagation delay represents the time that is needed for a packet to reach from the source end of a link to the destination end; processing delay is the time needed for an

intermediate router (on the path from the source to a destination) or a destination to do data extraction, error checking, and forwarding; and the queuing delay is the time that the packet is waiting for process on the intermediate router. Thus, it can be seen, for the above four types of delay, that only the propagation delay is related to distance. While the propagation delay cannot be directly measured and the RTT is easy to be measured, geolocation algorithm usually uses half of the smallest RTT between two network entities as the propagation delay. This is because small RTT usually means small processing delay and send delay, and propagation delay is a larger proportion of RTT.

Yet a landmark does not have the ability to measure RTT to the target IP actively, once the delay between landmark and target IP can be measured, the landmark is exactly the probe point. Wang et al. [7] introduced the concept of relative delay. For two landmark nodes A and B, the nearest common router between A and B is R, while the RTT from probe P to A, B, and R is RTT(P,A), RTT(P,B), and RTT(P,R); the relative delay between A and B is shown in formula (1).

$$\begin{aligned} \mathrm{RlttRTT}(A, B) = {} & (\mathrm{RTT}(P, A) - \mathrm{RTT}(P, R)) \\ & + (\mathrm{RTT}(P, B) - \mathrm{RTT}(P, R)) \end{aligned} \quad (1)$$

4.2.2 Converting delay to geographic distance

In the Internet, hosts which share the last router are usually distributed around this router. Because the processing ability, material, and congestion of links from the nearest common router to landmarks and target are similar to each other, the relative delay between landmarks and the target which shares the nearest common router is supposed to be proportional to the geographic distance between them. For instance, denote T as the target IP, and A, B, and C as three related landmarks, while t_1, t_2, and t_3 are the corresponding relative delays between A, B ,and C; and $d(A, T)$, $d(B, T)$, and $d(C, T)$ are the geographic distances; and then $d(A, T) : d(B, T) : d(C, T) = t_1 : t_2 : t_3$.

4.2.3 The optimal solution

When introducing the deviations to landmarks, with known delay between landmarks and target, finding the target's location can be transformed into solving an optimization problem, and the objective function of this problem is minimum mean square error of deviations; two conditions are as follows: (1) the distance between the landmark's claimed position and real position are no larger than the corresponding deviation and (2) the distance between the target and landmark's real position of landmarks is proportional to the relative delay between them.
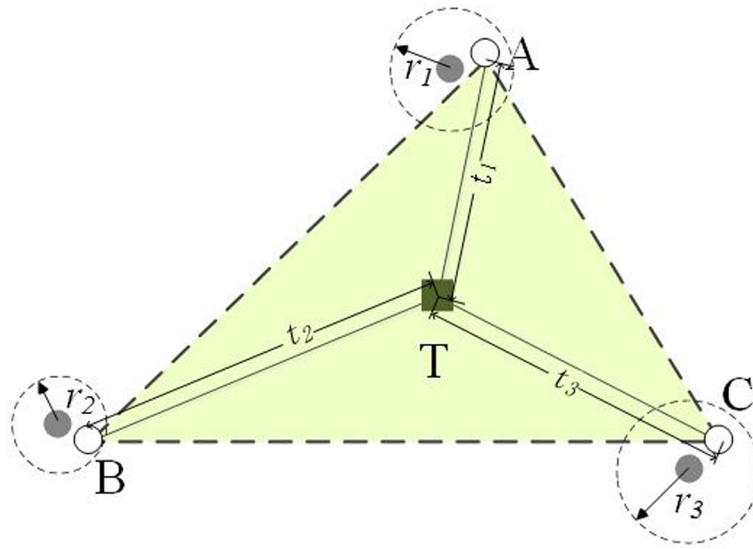
Fig. 5 Geolocation with three landmarks

Taking three related landmarks as an example, geolocation schematic is shown in Fig. 5. Denoting A, B, and C as the claimed positions (●) of three landmarks; A′, B′, and C′ as the real positions (○); the deviations of the three landmarks as r_1, r_2, and r_3, respectively; and the relative delay between landmarks and the target as t_1, t_2, and t_3, then the objective function of optimization problem is shown in formula (2). Two conditions are shown in formulas (3) and (4), respectively. Among those formulas, $d(\cdot)$ is the function of geographical distance.

$$min\left((r_1{}^2 + r_2{}^2 + r_3{}^2)/3\right) \tag{2}$$

$$d\left(A, A^{'}\right) \leq r_1$$
$$d\left(B, B^{'}\right) \leq r_2 \tag{3}$$
$$d\left(C, C^{'}\right) \leq r_3$$

$$d\left(A^{'}, T\right) \times t_2 = d\left(B^{'}, T\right) \times t_1$$
$$d\left(A^{'}, T\right) \times t_3 = d\left(C^{'}, T\right) \times t_1 \tag{4}$$
$$d\left(C^{'}, T\right) \times t_2 = d\left(B^{'}, T\right) \times t_3$$

5 Algorithm analysis

The advantage and estimation of relative delay and its application in IP geolocation are elaborated in [7]; we do not give more analysis in this paper. Specifically,
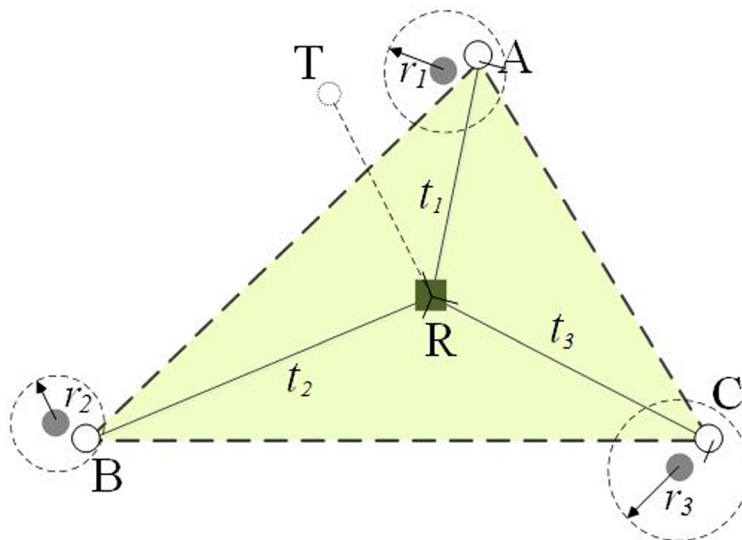

Fig. 6 Router geolocation

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 8 of 11

network coordinate system [16, 17] calculates the coordinates for each node through a small amount of delay measurements; using the coordinates, predicted delay between any two nodes can be obtained without direct measurement. The following gives analysis on converting delay to distance and optimal solution.

5.1 Converting delay to geographic distance

Considering that landmarks obtained by Web mining may have certain degree of deviation, when applying landmarks to geolocation, this paper introduces the deviation to landmarks, which is the distance between the landmark's claimed position and real position. In order to calculate the constraint distance between the landmark and target, we also need to compute the conversion relationship between the delay and the distance, after getting the estimated delay and possible areas of the landmark. As known from the existing studies, the correlation between the delay and distance of PlanetLab nodes is rather strong, and taking 2/3 and 4/9 C or bestline (such as CBG) as the conversion coefficient, the geolocation algorithms based on delay measurement can obtain effective constraints for the target and get the geolocation result. In the Internet region we studied, the correlation between delay and distance is very weak [18], as well as the relative delay and distance. So it is difficult to construct distance constraints for target IP from the probes and landmarks.

SLG [7] argued that, while there are a sufficient numbers of traceroute servers, the path between landmark and target IP connected through the nearest common routers can represent the direct path between them. While estimating the relative delay between the landmark and target, the link between the two connected by a nearest common router is seen as the real link between them. If the target and landmark are terminal nodes in the network, or very close to the terminal nodes, the corresponding network status and material of those links are similar. Therefore, for those links that are hard to find a fixed linear relationship between delay and distance, we can still think that the delay is proportional to distance.

Table 2 Paths from the probe point to 1.192.158.178 and 171.8.225.141

Probe_IP	Router_IP	Target_IP	RouterHop
218.29.102.104	10.69.30.33	1.192.158.178	3
	171.8.240.213		4
	222.85.124.50		5
	1.192.158.178		6
	10.69.30.33	171.8.225.141	3
	171.8.240.213		4
	171.8.240.18		5
	171.8.225.141		6

5.2 The optimal solution

When verifying the reliability of landmarks, the institutions utilize shared hosting or CDN will be removed, as well as the institutions with "multi-branch." However, the reserved landmarks' claimed location may still be inconsistent with the real location. For the geolocation approach proposed in this paper, a deviation is introduced to the landmark, and in combination with estimated relative delay between landmarks and the target, the target IP geolocation is converted to solve an optimization problem. In the solving process, the scope of deviation can be set according to the type or credibility of a corresponding institution, and taking the claimed position as the initial value of the real location and adopting the idea of optimization to find a set of optimal solutions, we can get the real positions of landmarks, as well as the geolocation result of the target.

In the last section, we point out that our geolocation approach needs to estimate relative delay between landmarks and target, and this requires the target IP should be measurable. For the unmeasurable target IP (there is no RTT from a probe point to this target), we can geolocalize the nearest common routers (R), taking this geolocation result as an estimation of the target's location. The router geolocation schematic is shown in Fig. 6. For different nearest common routers detected from different

Table 1 Landmarks and measurable targets for test

No.	IP address	Type	Latitude	Longitude
1	1.192.147.69	Primary school	34.72892	113.611044
2	222.88.59.236	Primary school	34.79583	113.67322
3	123.161.204.46	Middle school	34.67615	113.633818
4	1.192.158.178	Middle school	34.78597	113.690709
5	1.192.156.104	Travel agency	34.79784	113.673154
6	171.8.225.141	Government	34.81322	113.576988

Table 3 Relative delay between targets and landmarks

Target_IP	Landmark_IP	Relative_delay (ms)
1.192.158.178	1.192.147.69	2.314
1.192.158.178	222.88.59.236	1.733
1.192.158.178	123.161.204.46	2.863
1.192.158.178	1.192.156.104	2.180
171.8.225.141	1.192.147.69	1.813
171.8.225.141	222.88.59.236	1.232
171.8.225.141	123.161.204.46	2.362
171.8.225.141	1.192.156.104	1.679

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 9 of 11

Table 4 Geolocation results of the two measurable targets

Target_IP	r1	r2	r3	Latitude	Longitude	Error 1 (km)	Error 2 (km)
1.192.158.178	0.022	0.0219	0.0219	34.9557	113.603	20.5219	1.9393
171.8.225.141	0.0218	0.0218	0.0218	34.7517	113.6882	12.2591	9.0066

probe points, and if there are landmarks connected to those routers, geolocating each R, the position of target IP can be estimated according to the same strategy, such as finding the smallest area covered by each R. When the geolocation result is limited to a single point, we can take the centroid of the above area as the result location.

6 Experimental results

Usually, the IP address and location of the Web server is stable, so extracting the IP address and location of the Web servers by Web mining is an effective way to obtain landmarks [7], and because there are large numbers of Web servers in the network, a large number of landmarks can be obtained. In the process of landmark acquisition, while the landmarks which are far from its real position have been removed, such as the institution using CDN and shared hosting, the landmarks with small deviation (as shown in Figs. 1 and 2) are still reserved, and existing studies have not yet done further processing. For those landmarks, existing geolocation methods do not consider landmark's deviation, and the number of those landmarks is large in practice, so we cannot remove them directly. In addition, it is difficult to judge whether there is a measurement error in the measured delay, and we cannot give the experimental result of the proposed approach in view of an inaccurate delay now. So assuming that the measured delay is correct, this approach is verified by geolocating the measurable and unmeasurable targets. In the experiment, both the delay and nearest common routers are detected from a single probe point (located at 34.816129 N, 113.535455° E).

6.1 Geolocating the measurable target

There are six landmarks (located in Zhengzhou City), and the IP address, institution's type, latitude, and longitude are shown in Table 1. Among them, due to small coverages of the corresponding institutions, the deviations of the first three should be small, and usually, travel agencies and the district government do not maintain servers independently, so the last two landmarks are likely a little far from its claimed locations. Therefore, we take the last two landmarks as the target IP and give geolocation results for them using the first four landmarks with small deviations.

The paths from the probe point to the target (1.192.158.178 and 171.8.225.141) are shown in Table 2, and the first three columns in this table are the IP address

of the probe point, intermediate router, and target; the last column is the hop of the intermediate router on the path to the target. Combined with the paths of the four landmarks, we can get the nearest common routers (171.8.240.213) between targets and landmarks. Then we measure the RTT from the probe point to landmarks, the target IP, and the router (171.8.240.213) delay, and the relative delay between the target and landmarks is shown in Table 3 (the unit of the relative delay is millisecond).

While geolocating the target, we can choose different numbers of landmarks to construct the corresponding optimization problem. For the above two targets, using the first three landmarks in Table 1, two kinds of constraints could be obtained according to formulas (3) and (4), solving the optimization function as shown in (2), and we can get the geolocation results of the target IP. In the process of solving the optimization problem, the landmarks' deviation can be assigned different ranges, and each range may correspond to a different geolocation result. The possible ranges of the three landmarks' deviation are set as $[0 \sim 1]$, and the geolocation results of two targets are shown in Table 3. In Table 4, r1, r2, and r3 are the real deviations of the three landmarks; the latitude and longitude is the geolocation result of the target IP; and error 1 is the distance between the targets' claimed position and geolocation position, and error 2 is the geolocation error of the SLG method. (However, it seems that the geolocation error of SLG is more accurate; the geolocation is still meaningless without accurate landmarks). Table 4 shows that, when the target

Table 5 Landmarks and unmeasurable targets for test

No.	IP address	Type	Latitude	Longitude
1	203.171.231.106	Middle school	34.721486	113.67701
2	203.171.233.19	Middle school	34.713082	113.67333
3	218.28.177.173	Middle school	34.722917	113.7421
4	116.255.138.232	University	34.810269	113.68857
5	116.255.166.127	Government	34.774541	113.75879
6	116.255.207.74	Hospital	34.761226	113.70679
7	123.15.32.242	University	34.808231	113.6844
8	202.102.249.115	Primary school	34.802163	113.57886
9	218.28.221.195	Library	34.71182	113.5165
10	222.143.36.48	Government	34.755977	113.63185
11	61.163.101.18	Government	34.789172	113.6887

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 10 of 11

Table 6 Geolocation results of unmeasurable targets

Target_IP	r1	r2	r3	Latitude	Longitude	Error 1 (km)	Error 2 (km)
116.255.138.232	0.8385	0.6052	0.2667	34.7272	113.7232	4.2737	10.8867
116.255.166.127						4.8261	5.9461
116.255.207.74						1.7937	5.3498
123.15.32.242						9.7744	10.8642
202.102.249.115						6.1943	17.3399
218.28.221.195						4.0743	20.6796
222.143.36.48						9.6931	10.7361
61.163.101.18						15.6163	8.846

is measurable, the proposed geolocation approach could get a geolocation position with inaccurate landmarks. In addition, this experiment also shows the landmarks obtained by Web mining do exist and that the real positions are inconsistent with the claimed position.

6.2 Geolocating the unmeasurable target

There are 11 landmarks (located in Zhengzhou City), and the detailed information is shown in Table 5. The first three institutions with a small coverage are used as landmarks, and the deviations between their real position and claimed position are small, too. Taking the last eight institutions as targets, and assuming that those targets are unmeasurable, this means that there are no RTT from the probe point to those targets. Using the paths from the probe point to the landmarks and targets, the nearest common routers (61.168.251.69) between landmarks and the target can be achieved.

The possible ranges of the three landmarks' deviation are set as [0 ~ 1], and we can get the geolocation position of the router (61.168.251.69). After mapping the eight targets to this router, the geolocation results of the eight targets are shown in Table 6; r1, r2, and r3 are the real deviations of the three landmarks, and the latitude and longitude is the geolocation result of the router. Specially, while error 1 is the distance between the targets' claimed position and router's position, error 2 is the geolocation error of the SLG method.

It is shown in Table 6 that, geolocating the nearest common router and mapping the target to the position of this router, our approach can get similar geolocation errors as SLG, and this means that, while the delay from the probe point to the target is unmeasurable, the geolocation approach could still give the estimated position with acceptable error.

In addition, this experiment also shows that among those landmarks (Web servers and the corresponding locations) which were collected by Web mining, there are inaccurate landmarks, and the claimed locations are inconsistent with the real location.

7 Conclusions

For an IP geolocation system, lots of factors can affect the geolocation validity, such as the deviation of the landmark, measurement error of delay, and the link intricacy. To improve the effectiveness of geolocation system, a landmark calibration-based IP geolocation approach is proposed in this paper. Through paths detection, we find out the nearest common router and the related landmarks for the target IP; introduce deviations to landmarks and regarded their locations as possible areas; and calculate the relative delay between landmarks and the target; taking the distance between landmarks and the target is proportional to the relative delay as constraint conditions and estimates the real deviations of landmarks and the location of the target using optimization idea. Algorithm analysis and experimental results show that our approach can be used in IP geolocation for the measureable target IP and, especially, for the target which is unmeasurable (no RTTs can be measured); its location can be achieved by estimating the nearby router on the corresponding path. In the next study, we will focus on the deviation range and the effect on geolocation error and then give the statistical geolocation results using large numbers of landmarks.

Authors' contributions
JC conceived and designed the steps of the proposed approach, performed the experiments, and drafted the manuscript. FL and XL contributed to the conceiving of the proposed approach. FZ and GZ participated in the design of the experiment and experimental analysis. All authors read and approved the final manuscript.

Chen *et al. EURASIP Journal on Information Security* (2016) 2016:4

Page 11 of 11

References

1. VN Padmanabhan, L Subramanian, An investigation of geographic mapping techniques for Internet hosts. ACM SIGCOMM Comp. Commun. Rev. **31**(4), 173–185 (2001)
2. JA Muir, PCV Oorschot, Internet geolocation: evasion and counterevasion. ACM Comput. Surv. **42**(1), 1–22 (2009)
3. M Gondree, ZN Peterson, *Geolocation of data in the cloud, proceedings of the third ACM conference on data and application security and privacy, ACM*, 2013, pp. 25–36
4. ZN Peterson, M Gondree, R Beverly, *A position paper on data sovereignty: the mportance of geolocating data in the cloud, proceedings of the 8th USENIX conference on networked systems design and implementation*, 2011
5. B Gueye, A Ziviani, M Crovella, S Fdida, Constraint-based geolocation of Internet hosts. IEEE/ACM Trans. Networking **14**(6), 1219–1232 (2006)
6. E Katz-Bassett, JP John, A Krishnamurthy, D Wetherall, T Anderson, Y Chawathe, *Towards IP geolocation using delay and topology measurements, proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, 2006, pp. 71–84
7. B Wong, I Stoyanov, EG Sirer, *Octant: a comprehensive framework for the geolocation of Internet hosts, proceedings of USENIX NSDI conference*, 2007, pp. 23–36
8. Y Wang, D Burgener, M Flores, A Kuzmanovic, C Huang, *Towards street-level client-independent IP geolocation, proceedings of the 8th USENIX conference on networked systems design and implementation*, 2011, pp. 27–36
9. B Eriksson, P Barford, J Sommers, R Nowak, *A learning-based approach for IP geolocation, proceedings of passive and active measurements conference*, 2010, pp. 171–180
10. P Guo, J Wang, B Li, S Lee, A variable threshold-value authentication architecture for wireless mesh networks. J. Internet Technol. **15**(6), 929–936 (2014)
11. S Xie, Y Wang, Construction of tree network with limited delivery latency in homogeneous wireless sensor networks. Wirel. Pers. Commun. **78**(1), 231–246 (2014)
12. J Shen, H Tan, J Wang, J Wang, S Lee, A novel routing protocol providing good transmission reliability in underwater sensor networks. J. Internet Technol. **16**(1), 171–178 (2015)
13. Y Shavitt, N Zilberman, A geolocation databases study. IEEE J. Select Areas Commun. **29**(10), 2044–2056 (2011)
14. SS Siwpersad, B Gueye, S Uhlig, *Assessing the geographic resolution of exhaustive tabulation for geolocating Internet hosts, proceedings of passive and active measurements conference*, 2008, pp. 11–20
15. Technology report (2011) http://networks.cs.northwestern.edu/technicalreport.pdf
16. F Dabek, R Cox, F Kaashoek, R Morris, Vivaldi: a decentralized network coordinate system. ACM SIGCOMM Comput. Commun. Rev. ACM **34**(4), 15–26 (2004)
17. Y Chen, Y Xiong, X Shi, J Zhu, B Deng, X Li, Pharos: accurate and decentralised network coordinate system. IET Commun. **3**(4), 539–548 (2009)
18. D Li, J Chen, C Guo, Y Liu, J Zhang, Z Zhang, Y Zhang, IP-geolocation mapping for moderately connected Internet regions. IEEE Trans. Parallel Distribut. Syst. **24**(2), 381–391 (2013)
19. Planetlab (2007). http://www.planet-lab.org