# ImMuNE : Improved Multilateration in Noisy Environments

Mathieu Champion†, Marc Dacier†, Elisa Chiapponi*
† RC3, CEMSE - KAUST, Kingdom of Saudi Arabia, * EURECOM, France
{mathieu.champion, marc.dacier}@kaust.edu.sa, elisa.chiapponi@eurecom.fr

***Abstract*—Identifying an attacker is a key factor to mitigate ongoing attacks. To evade localization, a single compromised machine can hide for months behind millions of available residential IP proxies. Without knowing the IP address of the machine, registration-based geolocation methods cannot be applied.**

**Measurement-based methods have been proposed to estimate the location of a target without using its IP address. These methods use Round Trip Time (RTT) values and network speed modeling. They estimate a distance between the target and other observation points with known locations, called landmarks. However, most of these methods require additional information, whether it is on the topology of the network or the characteristics of the landmarks.**

**In this paper, we present ImMuNE, a measurement-based technique which can estimate a location with only a few Round Trip Time measurements between a target and landmarks, even when some of these measures are inflated by temporary network congestion.**

**Leveraging a previously made measurement campaign, we present promising results based on 11 millions TCP connections collected over a period of 4 months.**

***Index Terms*—RTT, multilateration, Geolocalisation, Internet measurement**

## I. Introduction

Scraping bots are a plague for online companies. They continuously query websites, increasing the costs for their owners without generating any revenue [1]. Commercial anti-bot solutions exist to counter this threat. They recognize and block mischieving IPs. However, in recent years, scrapers have started using the services of Residential IP proxy (RESIP) providers. This enables them to hide their machine behind millions of IPs belonging to real residential users, helping them to increase stealthiness and avoid detection.

The goal of our work is to locate such scraping machine hiding behind a RESIP provider. This will help us in better understanding the ecosystem of the scrapers, their main actors. A precise localization is not required at this stage. Knowing from which part of the world they launch their attacks would already greatly help those fighting against them.

There are two types of internet geolocalization techniques: registration-based and measurement-based [2]. The first method uses a database. This database contains previously known data which link a block of IP addresses to a location. The second method uses Round-Trip Time (RTT) measurements to apply a process called multilateration.

The first approach is not applicable to our problem since we do not know the IP address of the scrapers. We must thus consider measurement-based ones.

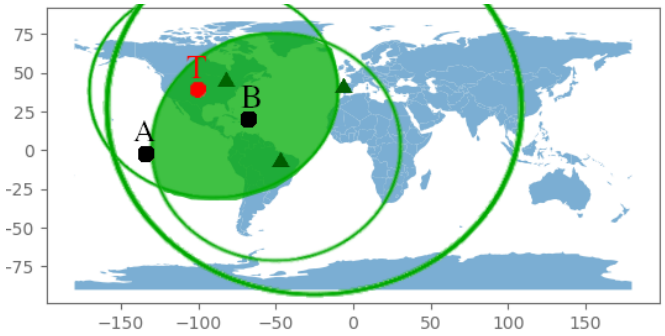


Fig. 1. Examples of multilateration. The point T is the ground-truth location of the target. The triangles are landmarks, the centers of circles. The radius of a circle is the distance between the landmark and the target, considering an average packet speed of 80 km/ms. Methods seeking the common intersection of all circles lead to point A as a solution; CBG [3], another classical method, returns instead the whole green area, from which its centroid B can be derived.

The idea of multilateration is to measure a distance between a target and several other observation points, called landmarks, whose location are known. From these, we can infer the location of the target (see Fig. 1). Instead of distances, we can only measure how much time it takes for an exchange of packets between two machines, i.e. a round-trip-time (RTT). If we know the speed of the packet, we can derive a distance from this RTT. Unfortunately, the RTT can be affected by many elements, such as the path taken, and so does the speed (e.g. because of network congestion). We must thus find a solution that would be resilient to, possibly, large errors in the estimation of the distances used.

Section II motivates the research done. Section III discusses why existing solutions are not satisfactory for our problem. Section IV shows that the distribution of packet speeds in our dataset is too volatile to enable a satisfying modeling, which deters the use of existing techniques. Section V describes our new measurement-based geolocalization method, called ImMuNE, and how it addresses our needs. Section VI presents our results. Section VII concludes the paper and offers some ideas for future work.

## II. Motivation

In [1], Chiapponi et al. present the impact of residential IP proxies on web scraping campaigns. In [5], the same authors use the semantics of the received queries to group searches issued by different IP addresses, leading to the conclusion that

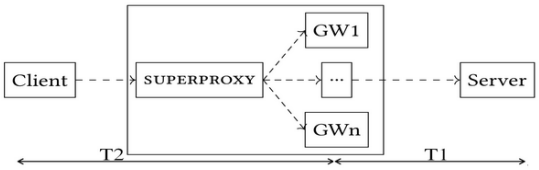


Fig. 2. RESIP proxies infrastructure. In this picture, the client is the web scraper. (taken from Chiapponi et al [4])

a single machine is likely behind specific campaigns but its location remains a mystery, which we want to solve.

The connections between the scrapers, the RESIP providers and the web servers are represented in Fig. 2. In [4], Chiapponi et al. explain how a web server can measure the values T1 and T2 represented in that Figure. It is very important to understand that there are a very large number of gateways (several millions), located all over the world and that, for each, we have a very small amount of T1 and T2 measures. The machines acting as gateways are diverse, ranging from servers to cell phones. The quality and stability of their access to the network are extremely variable. Furthermore, the packets sent (resp. received) by a gateway to a client must pass through the RESIP provider infrastructure which can be far away from the gateway and/or the client. In other words, the imposed trajectory through the RESIP infrastructure is likely to increase the value T2 and, thus, the error in the estimated distance between the gateway and the client.

Our end goal is to use the numerous gateways ($GW_i$) as landmarks to find the location of the scraper (client). We face two major difficulties to solve that problem: i) the poor quality of our landmarks, ii) the unknown latency introduced by the RESIP infrastructure.

As we will see, existing solutions cannot cope with the kind of errors introduced by our landmarks. We must thus address that first challenge before considering the second one. Therefore, in this work, we seek a robust method capable of geolocalizing a machine thanks to a number of measures of varying quality. To do so, instead of using the T2 measures, we will use the T1 measure, not impacted by the RESIP infrastructure. We rephrase our problem as follows: how to use the gateways as landmarks to geolocalize a given server? Since we know the real location of the servers, we can assess the quality and stability of our method. This is what we do in this paper and we plan on leveraging the produced new method to address the second challenge in a future work.

To validate experimentally our new method, we use the dataset built by Chiapponi et al [4]. It has been built thanks to 22 servers located in eight different places, all around the world (India, Australia, Japan, Europe, Canada, USA, South Africa, and Brazil). Connections have been collected over almost a 4 months period ((12/01/2022-01/05/2022), leading to a dataset of over 11 millions of unique gateway/server pair. The locations of the RESIP gateways were obtained from the MaxMind Ip-to-location database [6]. [7] [8] have shown that such database can encounter serious errors, which add to the uncertainty of our measurements.

## III. State of the art

Several measurement-based location methods have been proposed in the past. In [9] and [10], the authors aim to use measurement-based location to complement IP block-to-location database methods and need the IP address. In our use-case scenario we do not assume to have access to the IP address. Other methods are restricted to a limited geographical area (such as China [11] [12], or Europe/US [13] [14] [15] [16]) and need information like network topology to model the speed. All these information are not accessible in the context of targets hidden behind a proxy, as ours.

Other works do not work on the global scale because they need close landmarks [17] [18], sometimes inside the very city of the target, whereas our landmarks can be all over the world. The further away the landmarks are, the larger the RTT will be, and the more cumulative the error will be. The radii of the circles will then have a very large error and the results will then be affected.

Some methods which only need the RTT exist and find ways to estimate a distance from this metric [19] [3]. [20] conducted a study to determine an average global packet speed, giving a result of about 80km/ms with a good coefficient of determination (0.9794). However, as [21] pointed out, this study has several biases :

- The landmarks were taken from PlanetLabs which nodes have better connectivity than mainstream nodes.
- Most of PlanetLabs nodes are located inside the US and Europe, which could lead to a geographical bias.
- It was possible to conduct many measures between a single server-landmark pair, in order to closely approximate the minimum value, i.e. the one without congestion delay.

In their paper [21], Weinberg et al compared the most prominent measurement-based location methods from the literature, plus two variations of their own design, at the scale of the whole world. They clearly show that the most efficient solution was Constraint-Based Geolocation (CBG). CBG consists in an overestimation of the speed, and the intersection of all circles should give the area of the target. From this, they implemented CBG++, to fit CBG on a global scale.

The goal of their work is different from ours: they want to check that a proxy is within the country it is claimed to be in, and therefore see if it is in the "solution zone" proposed by CBG++. This promising solution cannot be used in our framework because of several issues :

- In CBG++, the targets are assumed to be within datacenters. This information is used for an important step of their analysis, called the disambiguation phase. That is not necessarily true for our final use case.
- The landmarks they use are RIPE anchors [22], known to be very stable, reliable and well connected. In contrast, our gateways are often, according to RESIP providers, mobile phones with unknown connectivity in varying conditions.

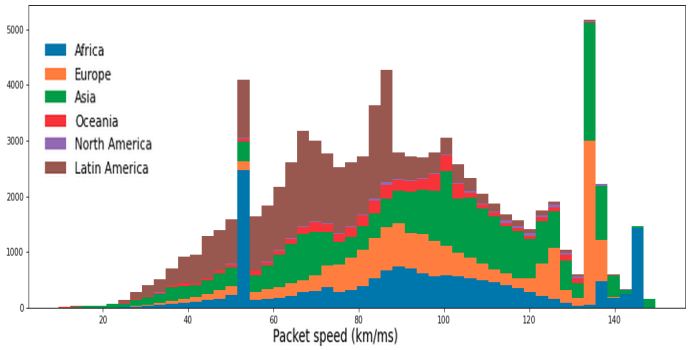


Fig. 3. Distribution of the packet speed, for connections from different continents to a server located in India (Y axis is the amount of connections observed with a given speed)

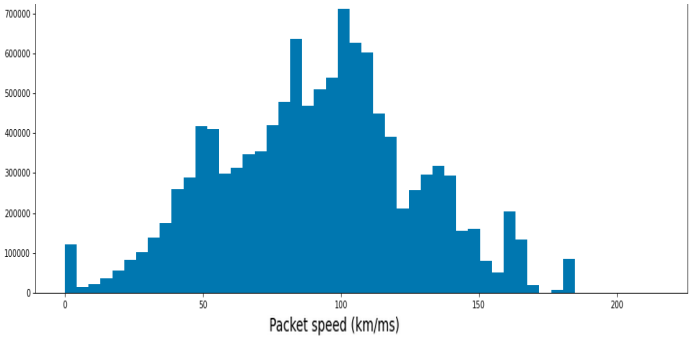


Fig. 4. Distribution of packet speed, from the connections among all servers (Y axis is the amount of connections observed with a given speed)

- It is important to notice that unlike in previously published experimental environments, the gateways (our landmarks) establish a single TCP connection of short duration to the target (the server). In the very few cases where we observe the same IP more than once, we do not know if it is associated to the same device, in the same operational conditions. In other words, we cannot accumulate a large number of measurements for each landmark to reduce the error possibly caused by, e.g., temporary network congestion.
- They obtain very good results at the global scale thanks to a 2 phases method. They first identify the target continent, then rely on local landmarks to refine the geolocation. This is only possible if all the RTTs are reliable and none grossly overestimated. In our study, this is unfortunately not the case.
- They rely on an "average speed" which, as we will show in the Section IV, does not really exist in our use case.

Enriched by the past results, we have produced an original method, called ImMuNE, which can cope with all the constraints described above and give us a good approximation of the position of the servers, while using very poor quality landmarks.

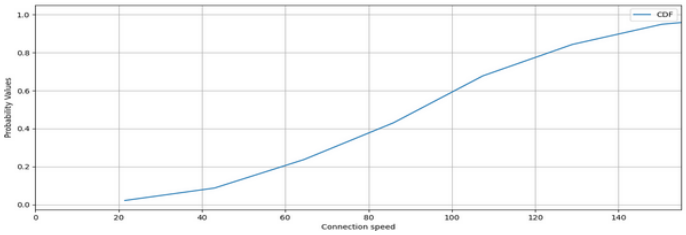


Fig. 5. CDF of the packet speed

## IV. Assessment of the packet speed

As mentioned in section III, several methods were proposed to model the *speed* of packets between 2 points. All these techniques assume that such mean value exists for the speed of a packet between each pair of landmark and targets. Our experimental results, unfortunately, dispute this claim.

Fig. 3 shows the distribution of the average speed between a server located in India and landmarks located everywhere in the world. This distribution was built by taking all the RTT measurements acquired between this server and the landmarks in our dataset. Using MaxMind IP-to-Location database [6], we gathered the location of each landmark. The relative *speed* of each connection was simply taken by dividing half of the RTT by the ground-truth distance between the landmark and the server.

We can see a very large range of values. The CBG++ solution proposed by [21] makes the assumption that there are no packet speed below 84km/ms but we have a lot of speeds under it! Fig. 5 shows that 40% of the computed speeds are below this value. Most speeds from Africa to India are even below 60km/ms, as shown in Fig. 3.

Other servers, even in Europe or in the USA, lead to a similar packet speed distribution. Fig. 4 shows the overall distribution computed over all our servers.

Due to this uncertainty, the final error on the distance is far too large to consider one value alone as acceptable. The resulting error prevents existing algorithms from performing a correct multilateration, and gives unsatisfactory results (see section VI).

There is no average speed, because each speed depends on the landmark-target pair. Thus, using an algorithm that requires to use the same speed for all landmark-target pairs is deemed to fail.

## V. Our ImMuNE solution

### A. Iterative Least Square Multilateration

The area estimation proposed by CBG++ is often too big to narrow the research to a specific point, covering more than half the surface of the earth in more than 40% of the time [21]. The multilateration method used by CBG/CBG++ is therefore not optimal for our study framework.

The multilateration solution we choose is based on the Iterative Least Square (ILS) method [23] [24]. The principle of the Least Square method is to find a point that minimizes

the squared error between this point and our different circles. The idea is to find a point ”close enough” to all the circles.

The solution is found iteratively. The detailed method of the resolution is explained in [23] .

This method has been used before in other geolocation settings, where the scale was reduced and the error was much smaller [23] . The challenges imposed by our study setting led to an improvement of ILS presented later in this section.

### B. 3D ILS

To our knowledge, ILS-based multilateration was only performed in a context where the landmarks and the target are relatively close [23] or at least close enough to neglect the roundness of the earth and to assimilate it to a plane.

At our scale, it is logical to consider the spherical aspect of the earth. Thus, we implemented a 3D version of ILS. However, the addition of a third dimension was discouraging, both in terms of lower accuracy (see Fig. 10) and much higher complexity. The reason is that the intersection of the spheres is not necessarily on the surface of the earth. Reprojection onto the earth's surface then introduces an additional error that can be very large (on the order of 5000kms in our experiments). Based on these results, we made the choice to do the multilateration on a 2D projection of the earth. The impact of this projection is discussed in Subsection V-D.

### C. Multiple ILS : Using several speeds per landmark

This section presents the modifications of the ILS-based multilateration that we have made to adapt it to our problem.

We started using ILS with the speeds proposed by previous authors. Our related distances were sometimes close, sometimes very far from the ground-truth. As we saw in section IV, this happens because the speed is dependent on the landmark-server pair. In our final problem, this will be the web scraper-landmark pair which is, clearly, unknown since the scraper is the unknown element that we try to identify.

Instead of attributing one unique speed to a given landmark-server pair, we can attribute several ones. The idea is that if we are able to determine the perfect speed for each connection between each landmark and each server, then we will have the right solution. Then, if we add bad speeds (i.e. more circles centered around the same landmark), the good circles will still be there, and the others will only add marginal noise. The intuition is that the optimal solution, as defined by the ILS method in Subsection V-A, remains the same even with the addition of the “erroneous” circles. The procedure is an iterative one and does not lend itself well to an analytical proof of correctness though. Instead, we provide a thorough experimental evaluation of the application of the method with real world data representative of the problem we want to solve.

### D. Impact of the projection centering on the estimation

A projection on a flat surface necessarily introduces errors, but this error is different depending on the chosen projection.

In our case, we must minimize the distortion around the equator, because most landmarks are located between 40S and 60N, as pointed out by [25] and also done in [21]. We also need to project the whole earth. We chose the Equirectangular projection. This projection distorts the distances at the poles, so we expect to have a latitude error higher than the longitude error. Since the landmarks are more likely to be close the equator than very close to the poles, this is acceptable in our case.

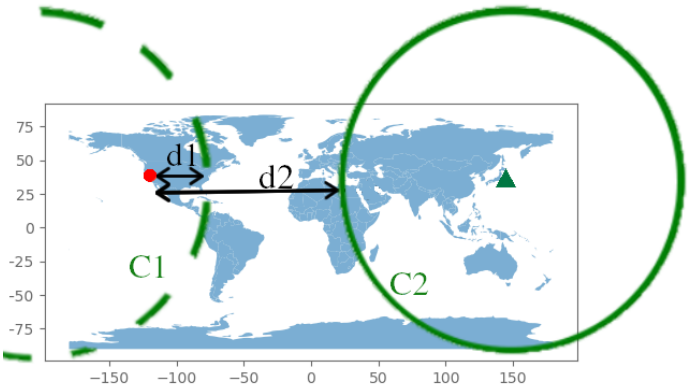


Fig. 6. Example of a misplaced circle. Because of the centering of the projection, the solution will try to minimize d2 instead of d1.

The centering of the projection also affects the result. Fig. 6 illustrates the problem of the 2D projection. In this figure, the red dot in the USA is the target, and the green triangle in Japan is one landmark. The circle C2 is the circle representing the distance between this landmark and the target. C2 is projected, and we know that in reality it should end where C1 is represented. Unfortunately, C1 is lost in the 2D projection. The target is at the distance d1 of that circle, but the ILS will see it at the distance d2, which introduces a large error.

Fortunately, we can find a way to select the correct projection. As explained, ILS searches the point with the least squared distance to the circles. The implementation of ILS we choose [24] returns several values : the estimated location, but also its error (i.e. least squared distance to the circles), and a radius which is correlated to the uncertainty of the result. We can use these last two values as a confidence value, called $\kappa$.

For a given set of landmarks and a given target, we have used three distinct projections. In the first one, Europe is in the middle of the rectangle whereas the American continent (resp. China) is in the center of the second (resp. third). We apply the ILS method to each case and discover that the confidence value $\kappa$ is best when the ground-truth location of the target is located in the center of the projection. That experiment has been repeated a very large amount of times and Fig. 9 highlights how, by using this indicator, we can select the right projection and minimize the error with respect to the ground truth of the target.

## VI. Results

To test our model, we apply our method for each of the targets (the servers) at our disposal. The method is as follows. First, we select the amount of landmarks to be used.

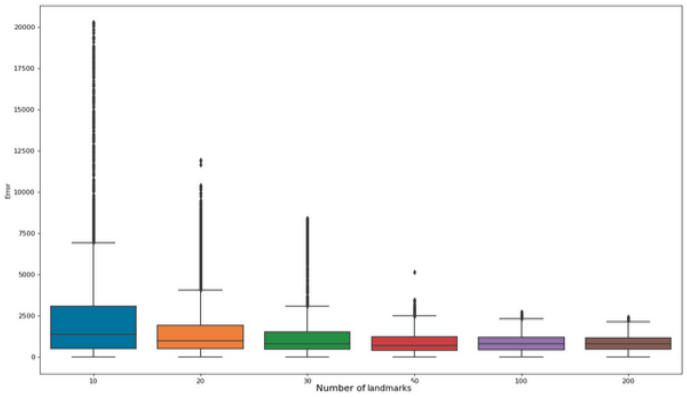


Fig. 7. Box plot showing the statistics of the overall location estimation error

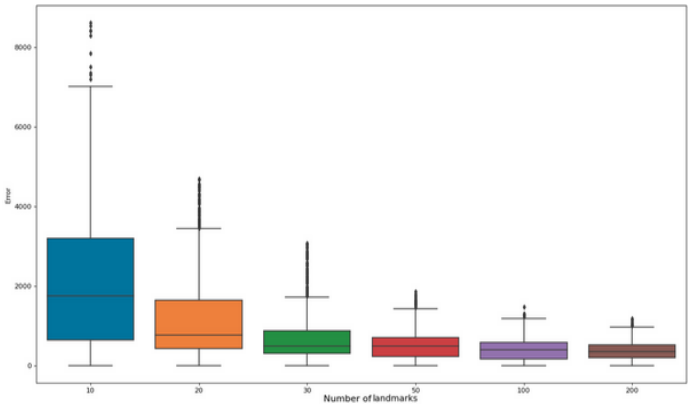


Fig. 8. Box plot showing the statistics of the location estimation error, when the target is in India

This number can be among the following values : [10, 20, 30, 50, 100, 200]. We use 3 different speeds per landmark, namely 40, 80 and 120km/ms. Then, we do the multilateration using IMMUNE with 3 projections each time, centered around Europe, USA and China, to tackle all cases discussed in Subsection V-D. Relying on the $\kappa$ estimation explained in section V-D, we select the assumed best location estimation. Since we know the exact location of the target, we compute the error between the estimated and the real location.

We repeat this 10.000 times, for each server and for each number of chosen landmarks. We manage to locate every target with an error of less than 1000kms in 50% of cases using only 30 landmarks (see Fig.e 7). Some servers like Japan or India (see Fig. 8) are located with an accuracy of about 500kms to their ground-truth location. This enables us to clearly identify these countries with less than 30 landmarks, no matter where they are placed on the planet.

As expected in Subsection V-D, the longitude errors are smaller than the latitude errors. This difference increases as one moves away from the equator. Thus, the errors in South Africa are larger than the errors in India, closer to the equator. The errors in longitude are small enough for the problem we try to solve, with less than 1000km in more than 60% of cases and less than 2000km in 95% of cases. Fig 9 shows the Cumulative Distribution Function (CDF) of the longitudinal error for all servers, with 50 landmarks per multilateration and using 1000 multilaterations per server, with randomly taken landmarks for each multilateration and in three situations. The top line represents the ideal case when we always chose the best projection to find the location, knowing the ground truth location. The line underneath it, close to it, is computed when we use our $\kappa$ estimator to identify which of the 3 projections to use to identify the location. The third line, below, is computed with a single projection always centered on Europe. This highlights the impact of the projection and the effectiveness of the $\kappa$ estimator, as the results are far better with an adaptive centering using the $\kappa$ estimator than without.

A 1000 kms approximation might seem a big error for a geolocalization solution but it is perfectly acceptable for our use case. Indeed, as a first step, we want to know in which

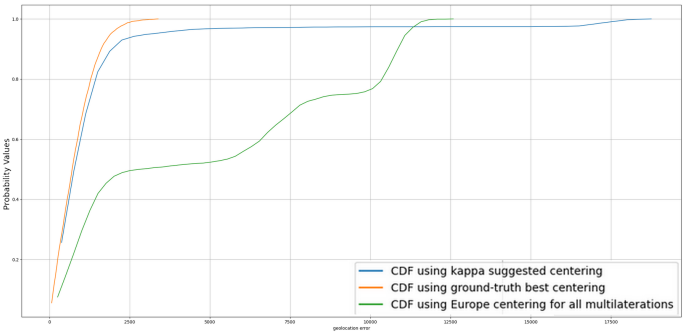


Fig. 9. CDF of the longitudinal location error, using centering suggested by the $\kappa$ estimator, the best ground-truth centering, and the same centering (Europe) for all targets.

part of the world the web scraper is located (e.g. Russia VS USA, Europe VS China, etc.). For regions where mitigation is possible, then, in a second phase, a selection of landmarks from that region can be selected to get much better results, as proposed in previous approaches ( [21] [18] [16])

We also observe a disparity between the different hemispheres. Indeed, despite an equivalent distance to the equator between South Africa and Japan, we observe a bigger latitude error for South Africa (-500km for Japan and about 3000km for South Africa). A possible explanation is the less rectilinear internet link between the continents of the Southern Hemisphere (e.g. regarding submarine cables). For example, a US-South Africa link via the Atlantic passes through Europe, resulting in a larger RTT and thus a far larger associated circle than it should be theoretically.

[26] points out this issue in detail. This problem takes place especially for African location. Packets sent to these locations take large detours and this results in large measurement errors. The same problem exists as well for other types of transcontinental links, such as the Australia-South America route which pass by the USA. However, this should not be to much of a problem in our case since these regions have not historically been known as safe harbors for scrapers machines.

Last but not least, Figure 10 confirms that with the iterative

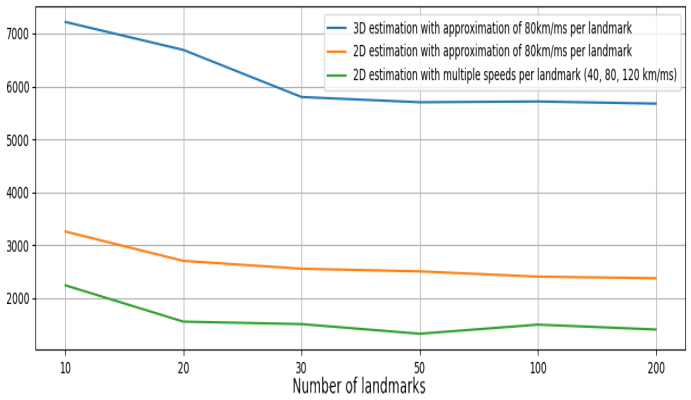


Fig. 10. Median error on the location estimation (considering the combination of all target servers) Y axis is the longitudinal error in kms.

least square technique, a 2D projection (bottow lines) leads to better result than with a 3D approach (top line). It also shows the consistently better results obtained when using several speeds per landmark as opposed to the average 80km/msec.

## VII. Conclusions and future work

To our knowledge, our method is the first to achieve the kind of geolocalisation we need in terms of connectivity and stability with very unreliable landmarks about which we have no additional information. We are now confident that we can use the gateways, as described in our RESIP environment, in order to localize the servers they talk to.

It remains to be seen whether we can leverage this newly produced method in order to find the location of the scrapers. Whereas we have no doubt about our method itself, the complete absence of knowledge of the inside architecture of the RESIP infrastructure does not enable us to ensure that we will be successful in tackling the second challenge. We leave this as further work. If we are successful, we will then be able to minimize our geolocalization error by enriching our solution with a multi steps approach, as proposed in some earlier work in other contexts.

## References


[1] E. Chiapponi, M. Dacier, O. Catakoglu, O. Thonnard, M. Fangar, M. Mattsson, and V. Rigal, "An industrial perspective on web scraping characteristics and open issues," in *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume (DSN-S)*, 2022, pp. 5–8.

[2] J. Bendale and J. R. Kumar, "Review of different ip geolocation methods and concepts," 2014.

[3] 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.

[4] E. Chiapponi, M. Dacier, O. Thonnard, M. Fangar, and V. Rigal, "Badpass: Bots taking advantage of proxy as a service," in *Under submission*, 2022.

[5] E. Chiapponi, M. Dacier, O. Catakoglu, O. Thonnard, and O. Todisco, "Scraping airlines bots: Insights obtained studying honeypot data," *Intl. Journal of Cyber Forensics and Advanced Threat Investigations*, vol. 2, no. 1, pp. 3–28, 2021.

[6] "Maxmind," https://www.maxmind.com/, accessed: 2022-7-27.

[7] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos, "A look at router geolocation in public and commercial databases," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 463–469.

[8] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, "Ip geolocation databases: Unreliable?" *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 2, pp. 53–56, 2011.

[9] Z. Dong, R. D. Perera, R. Chandramouli, and K. Subbalakshmi, "Network measurement based modeling and optimization for ip geolocation," *Computer Networks*, vol. 56, no. 1, pp. 85–98, 2012.

[10] J. Chen, F. Liu, X. Luo, F. Zhao, and G. Zhu, "A landmark calibration-based ip geolocation approach," *EURASIP Journal on Information Security*, vol. 2016, no. 1, pp. 1–11, 2016.

[11] T. Wang, K. Xu, J. Song, and M. Song, "An optimization method for the geolocation databases of internet hosts based on machine learning," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[12] D. Li, J. Chen, C. Guo, Y. Liu, J. Zhang, Z. Zhang, and Y. Zhang, "Ip-geolocation mapping for involving moderately-connected internet regions," *Project participation from Microsoft Research*, 2009.

[13] B. Gueye, S. Uhlig, A. Ziviani, and S. Fdida, "Leveraging buffering delay estimation for geolocation of internet hosts," in *International Conference on Research in Networking*. Springer, 2006, pp. 319–330.

[14] B. Wong, I. Stoyanov, and E. G. Sirer, "Octant: A comprehensive framework for the geolocalization of internet hosts." in *NSDI*, vol. 7, 2007, pp. 23–23.

[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 *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, 2006, pp. 71–84.

[16] M. J. Arif, S. Karunasekera, S. Kulkarni, A. Gunatilaka, and B. Ristic, "Internet host geolocation using maximum likelihood estimation technique," in *2010 24th IEEE International Conference on Advanced Information Networking and Applications*. IEEE, 2010, pp. 422–429.

[17] F. Zhao, Y. Song, F. Liu, K. Ke, J. Chen, and X. Luo, "City-level geolocation based on routing feature," in *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*. IEEE, 2015, pp. 414–419.

[18] M. Grey, D. Schatz, M. Rossberg, and G. Schaefer, "Towards distributed geolocation by employing a delay-based optimization scheme," in *2014 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2014, pp. 1–7.

[19] S. Laki, P. Mátray, P. Hága, T. Sebők, I. Csabai, and G. Vattay, "Spotter: A model based active geolocation service," in *2011 Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 3173–3181.

[20] O. Krajsa and L. Fojtova, "Rtt measurement and its dependence on the real geographical distance," in *2011 34th International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2011, pp. 231–234.

[21] Z. Weinberg, S. Cho, N. Christin, V. Sekar, and P. Gill, "How to catch when proxies lie: Verifying the physical locations of network proxies with active geolocation," in *Proceedings of the Internet Measurement Conference 2018*, 2018, pp. 203–217.

[22] R. N. Staff, "Ripe atlas: A global internet measurement network," *Internet Protocol Journal*, vol. 18, no. 3, pp. 2–26, 2015.

[23] S. G. Tzafestas, "12 - mobile robot localization and mapping," in *Introduction to Mobile Robot Control*, S. G. Tzafestas, Ed. Oxford: Elsevier, 2014, pp. 479–531. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780124170490000122

[24] "easy-trilateration," https://github.com/agusalex/easy-trilateration, accessed: 2022-7-27.

[25] B. Eriksson, P. Barford, J. Sommers, and R. Nowak, "A learning-based approach for ip geolocation," in *International Conference on Passive and Active Network Measurement*. Springer, 2010, pp. 171–180.

[26] R. Landa, J. T. Araújo, R. G. Clegg, E. Mykoniati, D. Griffin, and M. Rio, "The large-scale geography of internet round trip times," in *2013 IFIP Networking Conference*. IEEE, 2013, pp. 1–9.