# On the Use, Challenges, Alternatives of IP Geolocation Data

**Authors:** Paul Gao, Enock Lee, Yaozhong Zhang

**Organization:** Beijing Tiantexin Tech. Co., Ltd.(IPIP.NET)

**Date:** Sep, 2025

1. **How to use IP-geo data?**
    - User analysis and targeting: Respond to users based on their geographic location and language. Similar to sites like TikTok and YouTube, IP-geo data can be used to push regionally preferred videos to specific users.
    - Advertising and marketing: IP-geo helps advertisers analyze the effectiveness of ad campaigns and can also be used to some extent to prevent ad fraud and invalid ad traffic.
    - CDN / DNS / load balancing routing: With accurate IP-geo data, you can return the most appropriate resolution for a user or direct them to the nearest server, thereby speeding up access and reducing latency.
    - Geofencing: You can allow access only from certain geographic locations or block access from specific geographic locations. Alternatively, different treatments can be applied for different locations.
2. **What problems exist with the industry's current production methods?**
    - Poor data quality.
    - The problem of duplicate names between a nation's administrative divisions and place names, as well as issues of language translation.
    - Lack of active detection mechanisms.
    - Neglecting the creation of the backbone network.
3. **What problems exist in the IP geolocation data currently used in the industry?**
    - Data not provided in a timely manner; update cycles are too long.
    - Special cases such as ANYCAST are not handled.
    - Issues with translating geographic locations.
    - There is a proliferation of databases that are secondarily provided by copying data, resulting in ambiguity about users' commercial licensing.
    - The impact of accuracy and precision on user experience, especially for IPv6.

4. **Is it necessary to establish a unified format and standards for publishing and updating IP geolocation data?**

   Yes. A unified file type can reduce user difficulty in adoption and migration and provide better compatibility. At minimum, such a file type should:

   - Support IPv6: As IPv6's share on the Internet continues to grow, IP databases should support it.
   - Database read performance: IP databases should achieve the fastest query speeds while using as few system resources as possible.
   - Support for ANYCAST: The special IP type ANYCAST has been widely used on the Internet. Because it cannot be assigned to a single city, the IP database should also be compatible with this special form.
   - Support for application scenario field: Mark in the database under what circumstances the IP is used (base station? broadband? data center? etc.).
   - Reserve space for multiple languages: provide users with support for multiple languages if needed.
   - Maintain extensibility: keep compatible fields for future requirement expansion.

5. **How is IP geolocation different from other geolocation methods?**

   - IP geolocation: It is based on the network layer. It looks up your IP address information in the database. It is the only geolocation method compatible with all situations, but its drawback is that it is not necessarily accurate.
   - GPS positioning: is based on physical-world location. By actively measuring, it directly determines the device's position in physical space and is not affected by network connection quality. The drawback is that it is limited to devices that support GPS positioning and requires user authorization.
   - Wi-Fi positioning and cell-tower positioning: are mappings between the network and the physical world. They locate via Wi-Fi hotspot databases and cell tower IDs. Depending on the operating system, they may not be available.

6. **Are there alternative methods besides IP-based positioning?**

   - In advertising and marketing, positioning can be done using the aforementioned GPS positioning and Wi-Fi.
   - For needs like DNS routing, because the server can only learn the client's IP address, it can only use the IP address to determine the user's location.

7. **In an IP database, besides geographic information, is it valuable to convey other information?**

   Other information should include: ISP information, user connection type, IP usage scenario, user network connection speed, etc.

- ISP information: should be included in the IP database. Some CDNs or DNS providers need this data for precise routing to direct users to the provider's caches. Also, different ISPs allocate IP address blocks in different ways. Therefore distinguishing ISPs can make maintenance work more straightforward.
- IP usage scenario: needs to be included in the IP database. It generally includes fields such as: data center, broadband, base station, organization, school, and so on. This field is very important in areas like advertising and anti-fraud.
- User connection type: should not be included in the IP database. Scenarios like advertising, CDN routing, etc., do not care whether a user is on fiber or coax. Moreover, multiple access methods may exist under the same IP, so it cannot be quantified from a single dimension.
- User network connection speed. Each operator's network conditions differ; under the same IP there may be users on high-speed 5G phones and others on old 3G phones. At the same time, connection speed is affected by signal coverage, interference, the number of simultaneous users, website load, bandwidth, and other factors, so it cannot be quantified from a single dimension; therefore, even if labeled, it is of little significance.

**8. How to define coverage?**

There are two ways to define coverage. The coverage referred to here is for the IPv4 network. The IPv6 network is too large to determine coverage.

- Simply looking at the ratio of IPs in a country that are labeled to province/city versus the total national IPs: this method is inaccurate. Some countries have large amounts of unused IP blocks, and some base station IP blocks are used at the national level. Chasing this paper-based coverage and insisting on labeling down to the city level not only fails to reflect the real local network situation, but may also cause problems for future work.
- Determine this based on sampling of the client's or one's own data: this is the correct approach. The ratio of actually sampled IP data to the IP data in the database already labeled at the provincial/city level can more truthfully reflect the state of the network.

**9. How to define accuracy?**

Accuracy refers to the ratio between IP addresses labeled to the provincial/city level and the real-world physical locations of those IP addresses. Given correctness, accuracy should be as high as possible.

**10. What kind of commercial licensing model should an IP database adopt?**

- Paid annually. An offline version of the database is provided.
- API interface, billed by number of queries.
- Customizable for clients; fields can be specified.