Internet identification options based on IP address

Summary

Project number:
2018.115

Publication number
2018.115-1907-MS v1.1

Date:
Utrecht, 18 October 2019

Authors:
Ir. Tommy van der Vorst
Jessica Steur MSc
Ir. Nick Jelicic
Ir. Jan van Rees

Management summary

Research question

In light of proposed legislation introducing a limited retention obligation on telecommunications data for detection and prosecution purposes, we examined whether it is technically feasible to identify individual users based on a public IP address. The question addressed in this study is:

> *How do (mobile) internet providers identify an individual user of a public IP address, up until 12 months after use, for investigation and prosecution purposes, and what are the relevant (social) considerations?*

We define social considerations as: (1) usability for investigation and prosecution, (2) citizens' privacy, and (3) the costs for internet providers. To address this research question, we reviewed the literature and held interviews with (mobile) internet providers, the police, and other stakeholders/experts. The findings enabled us to formulate strategy options.

Background

When a criminal offence is committed, but the offender is not caught in the act, they have to be found through traces, in order to proceed with prosecution. For instance, the perpetrator of a speed violation can be found if the vehicle licence plates are registered. A similar principle applies to online crime. When communication takes place via the internet, the recipient usually knows the sender's *IP address* – which is necessary for communication in the other direction. This IP address thus provides a direct indication of the connection and/or system used to commit an offence. IP addresses are issued by internet service providers (ISPs). Investigation services can request ISPs to disclose to which subscriber they have issued a certain IP address.

As a result of internet developments, the link between an individual and an IP address is no longer as evident as in the past. Due to the scarcity of IPv4 addresses (4th version of the internet protocol), these are *assigned dynamically*: subscribers have to share the same IP address but never simultaneously. Based on date, time and public IP address, an individual subscriber can, however, be identified. This is similar to searching for a driver of a rental vehicle based on the licence plate: this registration number belongs to the rental company, which, by checking its administration, can of course trace the hirer.

In situations where the number of available IPv4 addresses is *much* lower than the number of devices online at the same time, it is necessary to share IPv4 addresses *simultaneously* among users. This can be done by applying CG-NAT (*carrier grade network address translation*). In the analogy with rental vehicles, CG-NAT means that *various* hirers drive around the country in different rental vehicles, but all with *the same* licence plate. If the police want to identify a driver right away, along with the date and time, they need either more information about the car (for example the type and colour) or about the route (where was the car seen, or what was its destination?).

Currently, CG-NAT is mostly used in mobile networks. Depending on the operator, a public IP address is simultaneously shared with a handful to a thousand other subscribers. In CG-NAT cases, an IP address alone does not give the police enough information to identify the person they are trying to track down. More details are required to narrow down this group of people.

Findings

The current situation

- Based on an IP address and date/time, it is usually possible to identify subscribers on fixed-line networks in the Netherlands.

- The identification capabilities based on IP addresses differ considerably between (mobile) operators. Only in specific cases (depending on available port information and the operator) is it possible to identify one single address. Otherwise the size of the group ranges from 84 to 84,000 subscribers, which creates problems for investigation authorities.

Potential for improvement

We see several ways to improve identification. The net costs will barely differ for the operators but can make a considerable difference to the usability for detection and keeping track of (potential) privacy violation/legal proportionality information.

The most obvious solution for achieving 1:1 identification is IPv6. The sector is in agreement that (for more reasons than just identification) it must switch to IPv6. Currently the Dutch mobile internet providers have hardly any motivation to do so. Only one ISP has publicly announced the roll-out of IPv6 on their network. Potentially the government (as telecommunication services 'customer') can give the final push in the right direction. Although it will take time to adopt IPv6, which may never replace IPv4 traffic entirely, there will be less pressure on CG-NAT, and thus improved capabilities for identification based on an IPv4 address.

There are also options that allow identifying a smaller group size without IPv6, that could be implemented over the next few years. We think that adding IPv4 addresses is the easiest solution. The ISPs seem to have enough to be able to (re)use in their mobile network (an estimated total of 4.2 million). Deploying CG-NAT for all subscribers will result in a group size of about five. If an ISP cannot otherwise or does not want to deploy its own IPv4 addresses, it could purchase some. Affecting this decision are the (one-off) costs and the uncertainty if the right amounts of IPv4 addresses are available.

One alternative solution is to log the allocation of *source ports* to subscribers. This enables 1:1 identification if port information is available for investigation purposes. However, as this information is only available in a minority of cases, source port logging does not improve the general situation.

A second alternative is to resort to a form of traffic data logging that masks the data. Although it is not possible to discover precisely who has made a connection, given a specific IP address, you can identify a (smaller) group of subscribers. Considering the costs involved, it is doubtful if this solution would be a sufficient improvement. The infringement of privacy is on the one hand lower (on account of the smaller group size identified), but can increase, depending on how the data masking is implemented.

The following table compares the various options.

Option	Usability for detection and prosecution	Amount of personal data stored	Degree of privacy infringement due to investigation (group size)	Costs for provider
1. Roll-out / adopt IPv6	High. Less effort required for identification. More cases can be detected. It will take time before all services utilise IPv6. Until then, many traces will be IPv4, with no improvement.	Minimal. Information on (semi)static allocation of IPv6 address block to subscriber (analogue to IPv4 on fixed networks) according to date/time.	Minimal. An IPv6 address is always specific to a single subscriber / connection. Other subscribers can be directly excluded.	A few million euros maximum. Investment in IPv6 seems (also for other reasons than detection) ultimately inevitable. There are differences between operators regarding their previous investments.
2. Increase number of public IPv4 addresses	Average to high, depends on group size (maximum with 1:1 allocation). In most cases an IPv4 leads directly to identification. More traces to identify and more issues can be picked up.	Limited. Information on (semi)static allocation of IPv4 address to subscriber (already tracked as such on fixed networks).	Average. Depends on ratio between number of public IPv4 addresses and number of subscribers. If one address is available for each subscriber, breach is minimal. Group sizes from 15 are achievable.	Maximum of a few million euros. The cost of purchasing (scarce) IPv4 addresses and adjusting configuration.
3. Source port logging	Limited, unless service providers increase their tracking of information on source ports.	Limited. Information on allocating public IPv4 address and port sequences to subscriber according to date/time. Service providers must log port numbers.	Minimal, if investigation reveals source port number, IP address, date, time. Average to high in all other cases, depending on number of subscribers sharing a public IPv4 address.	Maximum of a few million euros. The information is already (briefly) tracked to enable CG-NAT. Investment needed to log/store data and make it accessible.
4. Masked logging of traffic data	Average to high, depending on the size of the group and the form of masking.	High. Information on each established connection must be stored. To a limited extent you can deduce who is communicating with who.	Average. Depends on ratio between number of public IPv4 addresses and subscribers and the way masking is done.	Maximum of a few million euros. Involves storing large amounts of data.
5. Increase Police workload	Low. Some cases cannot be solved without identification via an IP address. Other cases require a significant time investment to narrow down a suspect.	Minimal. Information on (semi)static allocation of IPv4 address to subscriber (already tracked as such on fixed networks). One mobile ISP keeps track of source port sequences.	Minimal (with fixed IP addresses), average (with mobile if source port numbers can be used), to high (if no source port number is available – the majority of cases).	None, other than the current costs for monitoring, storing and making the data accessible.

International comparison

From the international comparisons we can learn three main lessons for the situation in the Netherlands:

1. The Netherlands has a very high amount of IPv4 addresses in relation to its total population. Consequently, the incentive for ISPs to roll out IPv6 there is lower than in other countries.

2. Factors such as national policies play a greater role in ISPs' decision whether or not to adopt Ipv6 than international ISP conglomerate strategy.

3. IPv6 is well established on mobile networks and so operators will be able to roll it out within the foreseeable future.

Policy options

We see a number of policy options:

1. **A functional 1:1 identification obligation for ISPs**. Given the growing numbers (of devices/subscribers versus amount of available IPv4 addresses), in practice this solution ultimately means rolling out IPv6. Nevertheless, internet providers will be able to apply their own strategy for the short term. By implementing logging or adding IPv4 addresses, operators can postpone making major changes for several years and do not need to write off investments in CG-NAT.

2. **Encourage or oblige ISPs to roll out IPv6.** Considering the number of devices that will be connected to the internet in the future, Ipv6 will inevitably be rolled out and adopted. Although an obligation to roll out Ipv6 could be imposed, this would not be in line with the general strategy that ISPs are responsible for their own technical decisions, and might exclude other technical solutions.

3. **A functional 1:N identification requirement for ISPs**. As internet providers are probably not able to comply right away with 1:1 identification (their technology is not ready anyway), the decision can be made to implement this gradually or over a number of years.

4. **No specific new strategy; 'nudging'.** Potentially, apply 'softer' tools, such as encouraging websites/online services to store source ports, and drawing internet providers' attention to their (moral) responsibility. It is likely that ISPs will deploy IPv6 autonomously but this will be a very slow process.