Privacy considerations for IP broadcast and multicast protocol designers

draft-winfaa-intarea-broadcast-consider-01

Rolf Winter
rolf.winter@hs-augsburg.de

Origin of this work

- Experiment on a campus network with 5000+ users
  - Recording of multicast and broadcast messages
  - Anomymization of PII-relevant data
- Same experiment was conducted on selected SSIDs during the Yokohama meeting during a 24h period

Why is broadcast/multicast special

- Often the only way to implement certain functions efficiently (e.g. local service discovery)
  - Many non-IETF specified protocols are based on broadcast/multicast
- Large receiver group by design
  - Makes it trivial for anybody on a LAN to collect the information without special privileges or a special location in the network
- Encryption is more difficult when broadcasting/ multicasting messages

Observations from these experiments that are related to privacy - I

- Some apps broadcast frequently
  - Observable online times and known location (on the same LAN). Also a performance problem on wireless
  - Observed frequencies of a couple of broadcasts per minute/per app
- Example: one (popular) app observed accounted for 7% of all broadcast traffic

Observations from these experiments that are related to privacy - II

- Use of persistent identifiers
  - E.g. used to identify an installation of a certain app
  - This can effectively destroy all efforts to randomize IP and MAC addresses
  - Also allows correlating different interfaces to belong to the same device

Observations from these experiments that are related to privacy - III

- Some protocol carry user-specified data such as hostnames
  - Relates to ietf-intarea-hostname-practice
- Prevalent behavior is to use the device owners name in e.g. hostnames
- Example: During the experiment on the IETF network, for over 240 (of 2600) devices the owner could be doubtlessly identified (data was anonymized but we could say that a name or name combination was used and it was unique based on the attendees list, which was anonymized the same way and the anonymization keys were thrown away)
- Control experiments with students revealed that without anonymization, the figures above would be higher and names reveal a lot of additional information

Observations from these experiments that are related to privacy - IV

- Lots of protocols with lots of different pieces of information
- Correlation is possible and that allows to construct user and user group profiles

Observations from these experiments that are related to privacy - V

- Lack of configurability
- On/Off only (if at all)
  - If the app implements a desired functionality (and they typically do) then the decision is typically always on better everywhere on

Why does this matter

- For IETF protocols interesting but these are well-known
  - E.g. there are operational measures for protection such as DHCP-snooping
  - WG scrutiny, sec reviews etc.
  - OS developers and device manufactuers aware of it
- Non-IETF protocols
  - Designed in isolation
  - No operation support
  - Privacy consideration are useful as guidelines

How does this fit in with the other privacy related work

- DHCP-related work
- mDNS/DNS-SD-related work
- IP address randomization
- Hostname draft
- Potentially others…