Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 14, 2014

C. Huitema

March 13, 2014

Unique Identifiers in DHCP options enable device tracking
draft-huitema-perpass-dhcp-identifiers-00.txt

Abstract

Some DHCP options carry unique identifiers. These identifiers can enable device tracking even if the device administrator takes care of randomizing other potential identifications like link-layer addresses or IPv6 addresses. This document reviews these options and proposes solutions for better management.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 14, 2014.

Table of Contents

1. Introduction 2
1.1. Requirements 2
2. Listening to DHCP traffic is easy 3
3. Managing Identifiers in DHCP Client Options 3
3.1. Client IP address field 5
3.2. Client hardware address 5
3.3. Client Identifier Option 5
3.4. Host Name Option 6
3.5. Client FQDN Option 7
3.6. UUID/GUID-based Client Identifier Option 7
4. Security Considerations 7
5. IANA Considerations 8
6. Acknowledgments 8
7. References 8
7.1. Normative References 8
7.2. Informative References 8
Author's Address 8

1. Introduction

Reports surfaced recently of systems that would monitor the wireless connections of passengers at Canadian airports [CNBC]. We can assume that these are either fragments or trial runs of a wider system that would attempt to monitor Internet users as they roam through wireless access points and other temporary network attachments. We can also assume that privacy conscious users will attempt to evade this monitoring, for example by ensuring that low level identifiers like link-layer addresses are "randomized," so that the devices do not broadcast a unique identifier in every location that they visit.

The IETF may or may not be in charge of the actual process for changing link-layer addresses as the mobile nodes move. But the IETF is in charge of protocols like DHCP or IPv6 neighbor discovery that use low level identifers and associate them with IPv4 or IPv6 addresses. The present note identifies specific issues with DHCP options that could be used to identify and track a device, even if the link-layer identifiers were randomized.

1.1. Requirements

The keywords MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL, when they appear in this document, are to be interpreted as described in [RFC2026].

2. Listening to DHCP traffic is easy

The use of unique identifiers in DHCP is very much "by design." DHCP enables administrators to easily configure network parameters of connected devices, and efficient management requires that the devices be well identified. However, this requirement of easy management conflicts with the desire to maintain the privacy of the users when they visit "untrusted" networks.

DHCP traffic is carried over multicast groups, and is thus very easy to monitor by anyone with access to the local link. A node that joins multicast group 224.0.0.12 and listen to traffic over port 68 will receive a copy of all packets sent by clients to DHCP servers. Joining multicast group 224.0.0.12 and listening to port 67 will provide access to all packets sent by DHCP servers to clients.

In theory, it might be possible to modify the DHCP protocol to use unicast instead of multicast. Clients could perhaps discover the link layer address of the DHCP servers or relays, and servers or relays could send traffic directly to the link layer address of clients. This would make the traffic somewhat harder to listen to, but would probably only be a minimal speed bump for pervasive monitoring agencies, who can simply listen to all the link layer traffic, whether unicast or anycast.

In theory, it might also be possible to negotiate some form of encryption between DHCP servers and DHCP clients. This would protect the privacy of DHCP clients, as long as pervasive monitoring agencies do notpressure network operators to give them a copy of their DHCP traffic. But this would probably have a high deployment cost, as standard solutions like TLS or IPSEC can only be deployed after the client has acquired an IP address through DHCP.

Instead of proposing changes to the DHCP transport, or proposing encryption of the DHCP traffic, we consider here a simpler solution. The monitoring risk derives mostly from a set of DHCP options that carry unique identifiers of the clients. It can be mitigated by careful management of the information sent in these options, as explained in the next section.

3. Managing Identifiers in DHCP Client Options

The DHCP protocol consists of message excahnges between clients and servers. In the following discussion, we will concentrate on the analysis of messages sent by the DHCP clients, which contain identifying information provided by the client to the server.

Here is an example of DHCP message sent by a client, as captured by a network monitor.

```
......,........l  1 1 6 0 dd112cdb  0 080 0 c0a8 06c
................  0 0 0 0  0 0 0 0  0 0 0 0 e89a8fb3
K............... 4bad 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
................  0 0 0 0  0 0 0 0  0 0 0 0  0 0 0 0
............c.Sc  0 0 0 0  0 0 0 0  0 0 0 0 63825363
5..=......K...ic 35 1 83d  7 1e89a 8fb34bad  c 66963
ebox<.MSFT 5.07. 65626f78 3c 84d53 46542035 2e3037 d
....,./.!y.+....  1 f 3 6 2c2e2f1f 2179f92b fcff 0 0
............      0 0 0 0  0 0 0 0  0 0 0 0
```

This can be parsed as:

```
ohhh:  1 1 6 0
xid: dd112cdb
secs:  0 0
flags: 80 0
ciaddr: 192.168.0.108
yiaddr: 0.0.0.0
siaddr: 0.0.0.0
giaddr: 0.0.0.0
chaddr: e89a8fb34bad 0 0 0 0 0 0 0 0 0 0
sname:
file:
Message Type(53): 8
Client Id(61):  1e89a8fb34bad
Host name(12): icebox
Vendor Class Id(60): MSFT 5.0
Req. List(55): 1, 15, 3, 6, 44, 46, 47, 31, 33, 121, 249, 43, 252
```

In this example, four fields can be considered client identifiers: the "client IP address," the "channel address" (chaddr) which carries the link layer identifier, the "client ID," and the "host name." The vendor class identifier and the request list provide information about the type of device and the software version that it uses, but do not directly identify the client.

The DHCP message above was sent by a computer running Windows 7, and is fairly typical. Analysis of the published RFC shows a list of other options that may be used more rarely, but do also carry unique identifiers, such as the Client FQDN option and the UUID/GUID-based Client Identifier Option.

3.1. Client IP address field

Four bytes in the header of the DHCP messages carry the "Client IP address" (ciaddr) as defined in [RFC2131]. In DHCP, this field is used by the clients to indicate the address that they used previously, so that as much as possible the server can allocate them the same address.

There is very little privacy implication of sending this address in the DHCP messages, except in one case, on a new connection to a new network. If the DHCP client somehow repeated the address used in a previous network attachment, monitoring services might use the information to tie the two network locations. DHCP clients should ensure that the field is cleared when they know that the network attachment has changed, and in particular of the link layer address is reset by the device's administrator.

3.2. Client hardware address

Sixteen bytes in the header of the DHCP messages carry the "Client hardware address" (chaddr) as defined in [RFC2131]. The presence of this address is necessary for the proper operation of the DHCP service.

Hardware addresses, called "link layer address" in many RFC, can be used to uniquely identify a device, especially if they follow the IEEE 802 recommendations. These unique identifiers can be used by monitoring services to track the location of the device and its user. The only plausible defense is to somehow reset the hardware address to a random value before visiting an untrusted location. If the hardware address is reset to a new value, or randomized, the DHCP client should use the new randomized value in the DHCP messages.

3.3. Client Identifier Option

The client identifier option is defined in [RFC2132] with option code 61. It is discussed in details in [RFC4361]. The purpose of the client identifier option is to identify the client in a manner independent of the link layer address. This is particularly useful if the DHCP server is expected to assign the same address to the client after a network attachment is swapped and the link layer address changes. It is also useful when the same nodes issues

requests through several interfaces, and expect the DHCP server to provide consistent configuration data over multiple interfaces.

The considerations for hardware independence and strong client identity have an adverse effect on the privacy of mobile clients, because the hardware independent unique identifier obviously enables very efficient tracking of the client's movements.

The recommendations in [RFC4361] are very strong, stating for example that "DHCPv4 clients MUST NOT use client identifiers based solely on layer two addresses that are hard-wired to the layer two device (e.g., the ethernet MAC address). These strong recommendations are in fact a tradeoff between ease of management and privacy, and the tradeoff should depend on the circumstances.

In contradiction to [RFC4361], when privacy considerations trump management considerations, DHCP clients should use client identifiers based solely on the link layer address that will be used in the underlying connection. This will ensure that the DHCP client identifier does not leak any information that is not already available to entities monitoring the network connection. It will also ensure that a strategy of randomizing the link layer address will not be nullified by DHCP options.

3.4. Host Name Option

The Host Name option is defined in [RFC2132] with option code 12. Depending on implementations, the option value can carry either a fully qualified domain name such as "node1984.example.com," or a simple host name such as "node1984." The host name is commonly used by the DHCP server to identify the host, and also to automatically update the address of the host in local name services.

Fully qualified domain names are obviously unique identifiers, but even simple host names can provide a significant amount of information on the identity of the device. They are typically chosen to be unique in the context where the device is most often used. If that context is wide enough, in a large company or in a big university, the host name will be a pretty good identifier of the device. Monitoring services could use that information in conjunction with traffic analysis and quickly derive the identity of the device's owner.

When privacy considerations trump management considerations, DHCP clients should always send a non-qualified host name instead of a fully qualified domain name, and should consider randomizing the host name value. A simple solution would be to set the host name value to

an hexadecimal representation of the link layer address that will be used in the underlying connection.

3.5. Client FQDN Option

The Client FQDN option is defined in [RFC4702] with option code 81. The option allows the DHCP clients to advertize to the DHCP their fully qualified domain name (FQDN) such as "mobile.example.com." This would allow the DHCP server to update in the DNS the PTR record for the IP address allocated to the client. Depending on circumstances, either the DHCP client or the DHCP server could update in the DNS the A record for the FQDN of the client.

Obviously, this option uniquely identifies the client, exposing it to the DHCP server or to anyone listening to DHCP traffic. In fact, if the DNS record are updated, the location of the client becomes visible to anyone with DNS lookup capabilities.

When privacy considerations trump management considerations, DHCP clients should not include the Client FQDN option in their DHCP requests.

3.6. UUID/GUID-based Client Identifier Option

The UUID/GUID-based Client Machine Identifier option is defined in [RFC4578], with option code 97. The option is part of a set of options for Intel Preboot eXecution Environment (PXE). The purpose of the PXE system is to perform management functions on a device before its main OS is operational. The Client Machine Identifier carries a 16-octet Globally Unique Identifier (GUID), which uniquely identifies the device.

The PXE system is clearly designed for devices operating in a controlled environment, and its functions are not meant to be used by mobile nodes visiting untrusted networks. If only for privacy reasons, nodes visiting untrusted networks should disable the PXE functions, and not send the corresponding options.

4. Security Considerations

This draft does not introduce new protocols. It does present a series of attacks on existing protocols, and proposes an assorted set of mitigations.

5. IANA Considerations

This draft does not require any IANA action.

6. Acknowledgments

The inspiration for this draft came from discussions in the Perpass mailing list.

7. References

7.1. Normative References

[RFC2026] Bradner, S., "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, March 1997.

[RFC2132] Alexander, S. and R. Droms, "DHCP Options and BOOTP Vendor Extensions", RFC 2132, March 1997.

[RFC4361] Lemon, T. and B. Sommerfeld, "Node-specific Client Identifiers for Dynamic Host Configuration Protocol Version Four (DHCPv4)", RFC 4361, February 2006.

[RFC4578] Johnston, M. and S. Venaas, "Dynamic Host Configuration Protocol (DHCP) Options for the Intel Preboot eXecution Environment (PXE)", RFC 4578, November 2006.

[RFC4702] Stapp, M., Volz, B., and Y. Rekhter, "The Dynamic Host Configuration Protocol (DHCP) Client Fully Qualified Domain Name (FQDN) Option", RFC 4702, October 2006.

7.2. Informative References

[CNBC] Weston, G., Greenwald, G., and R. Gallagher, "CBC News: CSEC used airport Wi-Fi to track Canadian travellers", Jan 2014, <http://www.cbc.ca/news/politics/csec-used-airport-wi-fi-to-track-canadian-travellers-edward-snowden-documents-1.2517881>.

Author's Address

Christian Huitema
Clyde Hill, WA 98004
U.S.A.

Email: huitema@huitema.net