Privacy considerations for DHCPv4
draft-ietf-dhc-dhcp-privacy-02

Abstract

   DHCP is a protocol that is used to provide addressing and
   configuration information to IPv4 hosts. This document discusses the
   various identifiers used by DHCP and the potential privacy issues.

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 July 2, 2016.

Copyright Notice

Table of Contents

1. Introduction

   Dynamic Host Configuration Protocol (DHCP) [RFC2131] is a protocol
   that is used to provide addressing and configuration information to
   IPv4 hosts. The DHCP protocol uses several identifiers that could
   become a source for gleaning information about the IPv4 host. This
   information may include device type, operating system information,
   location(s) that the device may have previously visited, etc. This

document discusses the various identifiers used by DHCP and the
potential privacy issues [RFC6973]. In particular, it also takes
into consideration the problem of pervasive monitoring [RFC7258].

Future works may propose protocol changes to fix the privacy issues
that have been analyzed in this document. It is out of scope for
this document.

The primary focus of this document is around privacy considerations
for clients to support client mobility and connection to random
networks. The privacy of DHCP servers and relay agents are
considered less important as they are typically open for public
services. And, it is generally assumed that relay agent to server
communication is protected from casual snooping, as that
communication occurs in the provider's backbone. Nevertheless, the
topics involving relay agents and servers are explored to some
degree. However, future work may want to further explore privacy of
DHCP servers and relay agents.

2. Requirements Language and Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
document are to be interpreted as described in [RFC2119]. When these
words are not in ALL CAPS (such as "should" or "Should"), they have
their usual English meanings, and are not to be interpreted as
[RFC2119] key words.

In addition the following terminology is used:

Stable identifier - Any property disclosed by a DHCP client that
        does not change over time or changes very infrequently and is
        unique for said client in a given context. Examples may
        include MAC address, client-id or a hostname. Some
        identifiers may be considered stable only under certain
        conditions, for example one client implementation may keep
        its client-id stored in stable storage while other may
        generate it on the fly and use a different one after each
        boot. Stable identifier may or may not be globally unique.

3. DHCP Options Carrying Identifiers

In DHCP, there are a few options which contain identification
information or can be used to extract the identification information
about the client. This section enumerates various options and
identifiers conveyed in them, which can be used to disclose client
identification.

3.1. Client Identifier Option

   The Client Identifier Option [RFC2131] is used to pass an explicit
   client identifier to a DHCP server.

   The client identifier is an opaque key, which must be unique to that
   client within the subnet to which the client is attached. It
   typically remains stable after it has been initially generated. It
   may contain a hardware address, identical to the contents of the
   'chaddr' field, or another type of identifier, such as a DNS name.
   [RFC3315] in Section 9.2 specifies DUID-LLT (Link-layer + time) as
   the recommended DUID (DHCP Unique Identifier) type. [RFC4361],
   Section 6.1 introduces this concept to DHCPv4. Those two document
   recommend that client identifiers be generated by using the permanent
   link-layer address of the network interface that the client is trying
   to configure. [RFC4361] updates the recommendation of Client
   Identifiers to be "consists of a type field whose value is normally
   255, followed by a four-byte IA_ID field, followed by the DUID for
   the client as defined in RFC 3315, section 9". This does not change
   the lifecycle of the Client Identifiers. Clients are expected to
   generate their Client Identifiers once (during first operation) and
   store it in a non-volatile storage or use the same deterministic
   algorithm to generate the same Client Identifier values again.

   This means that most implementations will use the available link-
   layer address during its first boot. Even if the administrator
   enables link-layer address randomization, it is likely that it was
   disabled during the first device boot. Hence the original,
   unobfuscated link-layer address will likely end up being announced as
   client identifier, even if the link- layer address has changed (or
   even if being changed on a periodic basis). The exposure of the
   original link-layer address in the client identifier will also
   undermine other privacy extensions such as [RFC4941].

3.2. Address Fields & Options

   The 'yiaddr' field [RFC2131] in DHCP message is used to allocate
   address from the server to the client.

   The DHCPv4 specification [RFC2131] provides a way to specify the
   client link-layer address in the DHCPv4 message header. A DHCPv4
   message header has 'htype' and 'chaddr' fields to specify the client
   link-layer address type and the link-layer address, respectively.
   The 'chaddr' field is used both as a hardware address for
   transmission of reply messages and as a client identifier.

   The 'requested IP address' option [RFC2131] is used by client to
   suggest that a particular IP address be assigned.

3.3. Client FQDN Option

The Client Fully Qualified Domain Name (FQDN) option [RFC4702] is
used by DHCP clients and servers to exchange information about the
client's fully qualified domain name and about who has the
responsibility for updating the DNS with the associated AAAA and PTR
RRs.

A client can use this option to convey all or part of its domain name
to a DHCP server for the IP-address-to-FQDN mapping. In most case a
client sends its hostname as a hint for the server. The DHCP server
MAY be configured to modify the supplied name or to substitute a
different name. The server should send its notion of the complete
FQDN for the client in the Domain Name field.

3.4. Parameter Request List Option

The Parameter Request List option [RFC2131] is used to inform the
server about options the client wants the server to send to the
client. The content of a Parameter Request List option are the
option codes for an option requested by the client.

3.5. Vendor Class and Vendor-Identifying Vendor Class Options

The Vendor Class option [RFC2131], the Vendor-Identifying Vendor
Class option and Vendor-Identifying Vendor Information option
[RFC3925] are used by the DHCP client to identify the vendor that
manufactured the hardware on which the client is running.

The information contained in the data area of this option is
contained in one or more opaque fields that identify the details of
the hardware configuration of the host on which the client is
running, or of industry consortium compliance, for example, the
version of the operating system the client is running or the amount
of memory installed on the client.

3.6. Civic Location Option

DHCP servers use the Civic Location Option [RFC4776] to deliver of
the location information (the civic and postal addresses) to the DHCP
clients. It may refer to three locations: the location of the DHCP
server, the location of the network element believed to be closest to
the client, or the location of the client, identified by the "what"
element within the option.

3.7. Coordinate-Based Location Option

The GeoConf and GeoLoc options [RFC6225] is used by DHCP server to provide the coordinate-based geographic location information to the DHCP clients. It enables a DHCP client to obtain its geographic location.

3.8. Client System Architecture Type Option

The Client System Architecture Type Option [RFC4578] is used by DHCP client to send a list of supported architecture types to the DHCP server. It is used by clients that must be booted using the network rather than from local storage, so the server can decide which boot file should be provided to the client.

3.9. Relay Agent Information Option and Sub-options

A DHCP relay agent includes a Relay Agent Information [RFC3046] to identify the remote host end of the circuit. It contains a "circuit ID" sub-option for the incoming circuit, which is an agent-local identifier of the circuit from which a DHCP client-to-server packet was received, and a "remote ID" sub-option which provides a trusted identifier for the remote high-speed modem.

Possible encoding of "circuit ID" sub-option includes: router interface number, switching hub port number, remote access server port number, frame relay DLCI, ATM virtual circuit number, cable data virtual circuit number, etc.

Possible encoding of the "remote ID" sub-option includes: a "caller ID" telephone number for dial-up connection, a "user name" prompted for by a remote access server, a remote caller ATM address, a "modem ID" of a cable data modem, the remote IP address of a point-to-point link, a remote X.25 address for X.25 connections, etc.

The link-selection sub-option [RFC3527] is used by any DHCP relay agent that desires to specify a subnet/link for a DHCP client request that it is relaying but needs the subnet/link specification to be different from the IP address the DHCP server should use when communicating with the relay agent. It contains an IP address, which can identify the client's subnet/link. Also, assuming network topology knowledge, it also reveals client location.

A DHCP relay includes a Subscriber-ID option [RFC3993] to associate some provider-specific information with clients' DHCP messages that is independent of the physical network configuration through which the subscriber is connected. The "subscriber-id" assigned by the provider is intended to be stable as customers connect through

different paths, and as network changes occur. The Subscriber-ID is
an ASCII string, which is assigned and configured by the network
provider.

4. Existing Mechanisms That Affect Privacy

This section describes deployed DHCP mechanisms that affect privacy.

4.1. DNS Updates

The Client FQDN (Fully Qualified Domain Name) Option [RFC4702] used
along with DNS Updates [RFC2136] defines a mechanism that allows both
clients and server to insert into the DNS domain information about
clients. Both forward (A) and reverse (PTR) resource records can be
updated. This allows other nodes to conveniently refer to a host,
despite the fact that its IP address may be changing.

This mechanism exposes two important pieces of information: current
address (which can be mapped to current location) and client's
hostname. The stable hostname can then be used to correlate the
client across different network attachments even when its IP
addresses keep changing.

4.2. Allocation strategies

A DHCP server running in typical, stateful mode is given a task of
managing one or more pools of IP address. When a client requests an
address, the server must pick an address out of configured pool.
Depending on the server's implementation, various allocation
strategies are possible. Choices in this regard may have privacy
implications. Note that the constraints in DHCPv4 and DHCPv6 are
radically different, but servers that allow allocation strategy
configuration may allow configuring them in both DHCPv4 and DHCPv6.
Not every allocation strategy is equally suitable for DHCPv4 and for
DHCPv6.

Iterative allocation - a server may choose to allocate addresses one
by one. That strategy has the benefit of being very fast, thus can
be favored in deployments that prefer performance. However, it makes
the allocated addresses very predictable. Also, since the addresses
allocated tend to be clustered at the beginning of available pool, it
makes scanning attacks much easier.

Identifier-based allocation - some server implementations may choose
to allocate an address that is based on one of available identifiers,
e.g. client identifier or MAC address. It is also convenient, as
returning client is very likely to get the same address. Those
properties are convenient for system administrators, so DHCP server

implementors are often requested to implement it. The downside of
such allocation is that the client has a very stable IP address.
That means that correlation of activities over time, location
tracking, address scanning and OS/vendor discovery apply. This is
certainly an issue in DHCPv6, but due to much smaller address space
is almost never a problem in DHCPv4.

Hash allocation - it's an extension of identifier based allocation.
Instead of using the identifier directly, it is being hashed first.
If the hash is implemented correctly, it removes the flaw of
disclosing the identifier, a property that eliminates susceptibility
to address scanning and OS/vendor discovery. If the hash is poorly
implemented (e.g. can be reverted), it introduces no improvement over
identifier-based allocation.

Random allocation - a server can pick a resource randomly out of
available pool. That strategy works well in scenarios where pool
utilization is small, as the likelihood of collision (resulting in
the server needing to repeat randomization) is small. With the pool
allocation increasing, the collision is disproportionally large, due
to birthday paradox. With high pool utilization (e.g. when 90% of
available resources being allocated already), the server will use
most computational resources to repeatedly pick a random resource,
which will degrade its performance. This allocation scheme
essentially prevents returning clients from getting the same address
again. On the other hand, it is beneficial from privacy perspective
as addresses generated that way are not susceptible to correlation
attacks, OS/vendor discovery attacks or identity discovery attacks.
Note that even though the address itself may be resilient to a given
attack, the client may still be susceptible if additional information
is disclosed other way, e.g. client's address can be randomized, but
it still can leak its MAC address in client-id option.

Other allocation strategies may be implemented.

Given the limited size of most IPv4 public address pools, allocation
mechanisms in IPv4 may not provide much privacy protection or leak
much useful information, if misused.

5. Attacks

5.1. Device type discovery

The type of device used by the client can be guessed by the attacker
using the Vendor Class Option, the 'chaddr' field, and by parsing the
Client ID Option. All of those options may contain an
Organizationally Unique Identifier (OUI) that represents the device's

vendor. That knowledge can be used for device-specific vulnerability
exploitation attacks.

5.2. Operating system discovery

The operating system running on a client can be guessed using the
Vendor Class option, the Client System Architecture Type option, or
by using fingerprinting techniques on the combination of options
requested using the Parameter Request List option.

5.3. Finding location information

The location information can be obtained by the attacker by many
means. The most direct way to obtain this information is by looking
into a message originating from the server that contains the Civic
Location, GeoConf, or GeoLoc options. It can also be indirectly
inferred using the Relay Agent Information option, with the remote ID
sub-option, the circuit ID option (e.g. if an access circuit on an
Access Node corresponds to a civic location), or the Subscriber ID
Option (if the attacker has access to subscriber info).

5.4. Finding previously visited networks

When DHCP clients connect to a network, they attempt to obtain the
same address they had used before they attached to the network. They
do this by putting the previously assigned address in the requested
IP address option. By observing these addresses, an attacker can
identify the network the client had previously visited.

5.5. Finding a stable identity

An attacker might use a stable identity gleaned from DHCP messages to
correlate activities of a given client on unrelated networks. The
Client FQDN option, the Subscriber ID Option and the Client ID
options can serve as long lived identifiers of DHCP clients. The
Client FQDN option can also provide an identity that can easily be
correlated with web server activity logs.

5.6. Pervasive monitoring

This is an enhancement, or a combination of most aforementioned
mechanisms. Operator who controls non-trivial number of access
points or network segments, may use obtained information about a
single client and observer client's habits.

5.7. Finding client's IP address or hostname

Many DHCP deployments use DNS Updates [RFC4702] that put client's information (current IP address, client's hostname) into DNS, where it is easily accessible by anyone interested. Client ID is also disclosed, albeit in not easily accessible form (SHA-256 digest of the client-id). As SHA-256 is considered irreversible, DHCID can't be converted back to client-id. However, SHA-256 digest can be used as an unique identifier that is accessible by any host.

5.8. Correlation of activities over time

As with other identifiers, an IP address can be used to correlate the activities of a host for at least as long as the lifetime of the address. If that address was generated from some other, stable identifier and that generation scheme can be deducted by an attacker, the duration of correlation attack extends to that identifier. In many cases, its lifetime is equal to the lifetime of the device itself.

5.9. Location tracking

If a stable identifier is used for assigning an address and such mapping is discovered by an attacker, e.g. a hostname being put into DNS, it can be used for tracking user. In particular both passive (a service that the client connects to can log client's address and draw conclusions regarding its location and movement patterns based on address it is connecting from) and active (attacker can send ICMP echo requests or other probe packets to networks of suspected client locations) methods can be used. To give specific example, by accessing a social portal from tomek-laptop.coffee.somecity.com.example, tomek-laptop.mycompany.com.example and tomek-laptop.myisp.example.com, the portal administrator can draw conclusions about tomek-laptop's owner current location and his habits.

5.10. Leasequery & bulk leasequery

Attackers may pretend as an access concentrator, either DHCP relay agent or DHCP client, to obtain location information directly from the DHCP server(s) using the DHCP leasequery [RFC4388] mechanism.

Location information is information needed by the access concentrator to forward traffic to a broadband-accessible host. This information includes knowledge of the host hardware address, the port or virtual circuit that leads to the host, and/or the hardware address of the intervening subscriber modem.

Furthermore, the attackers may use DHCP bulk leasequery [RFC6926] mechanism to obtain bulk information about DHCP bindings, even without knowing the target bindings.

Additionally, active leasequery [RFC7724] is a mechanism for subscribing to DHCPv4 lease update changes in near real-time. The intent of this mechanism is to update operator's database, but if misused, an attacker could defeat server's authentication mechanisms and subscribe to all updates. He then could continue receiving updates, without any need for local presence.

6. Security Considerations

In current practice, the client privacy and the client authentication are mutually exclusive. The client authentication procedure reveals additional client information in their certificates/identifiers. Full privacy for the clients may mean the clients are also anonymous for the server and the network.

7. Privacy Considerations

This document at its entirety discusses privacy considerations in DHCP. As such, no dedicated discussion is needed.

8. IANA Considerations

This draft does not request any IANA action.

9. Acknowledgements

The authors would like to thank the valuable comments made by Stephen Farrell, Ted Lemon, Ines Robles, Russ White, Christian Huitema, Bernie Volz, Jinmei Tatuya, Marcin Siodelski, Christian Schaefer and other members of DHC WG.

This document was produced using the xml2rfc tool [RFC2629].

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

   [RFC2131] Droms, R., "Dynamic Host Configuration Protocol",
              RFC 2131, DOI 10.17487/RFC2131, March 1997,
              <http://www.rfc-editor.org/info/rfc2131>.

   [RFC2136] Vixie, P., Ed., Thomson, S., Rekhter, Y., and J. Bound,
              "Dynamic Updates in the Domain Name System (DNS UPDATE)",
              RFC 2136, DOI 10.17487/RFC2136, April 1997,
              <http://www.rfc-editor.org/info/rfc2136>.

   [RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J.,
              Morris, J., Hansen, M., and R. Smith, "Privacy
              Considerations for Internet Protocols", RFC 6973,
              DOI 10.17487/RFC6973, July 2013,
              <http://www.rfc-editor.org/info/rfc6973>.

   [RFC7258] Farrell, S. and H. Tschofenig, "Pervasive Monitoring Is an
              Attack", BCP 188, RFC 7258, DOI 10.17487/RFC7258, May
              2014, <http://www.rfc-editor.org/info/rfc7258>.

10.2. Informative References

   [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629,
              DOI 10.17487/RFC2629, June 1999,
              <http://www.rfc-editor.org/info/rfc2629>.

   [RFC3046] Patrick, M., "DHCP Relay Agent Information Option",
              RFC 3046, DOI 10.17487/RFC3046, January 2001,
              <http://www.rfc-editor.org/info/rfc3046>.

   [RFC3315] Droms, R., Ed., Bound, J., Volz, B., Lemon, T., Perkins,
              C., and M. Carney, "Dynamic Host Configuration Protocol
              for IPv6 (DHCPv6)", RFC 3315, DOI 10.17487/RFC3315, July
              2003, <http://www.rfc-editor.org/info/rfc3315>.

   [RFC3527] Kinnear, K., Stapp, M., Johnson, R., and J. Kumarasamy,
              "Link Selection sub-option for the Relay Agent Information
              Option for DHCPv4", RFC 3527, DOI 10.17487/RFC3527, April
              2003, <http://www.rfc-editor.org/info/rfc3527>.

   [RFC3925] Littlefield, J., "Vendor-Identifying Vendor Options for
              Dynamic Host Configuration Protocol version 4 (DHCPv4)",
              RFC 3925, DOI 10.17487/RFC3925, October 2004,
              <http://www.rfc-editor.org/info/rfc3925>.

   [RFC3993] Johnson, R., Palaniappan, T., and M. Stapp, "Subscriber-ID
              Suboption for the Dynamic Host Configuration Protocol
              (DHCP) Relay Agent Option", RFC 3993,
              DOI 10.17487/RFC3993, March 2005,
              <http://www.rfc-editor.org/info/rfc3993>.

   [RFC4361] Lemon, T. and B. Sommerfeld, "Node-specific Client
              Identifiers for Dynamic Host Configuration Protocol
              Version Four (DHCPv4)", RFC 4361, DOI 10.17487/RFC4361,
              February 2006, <http://www.rfc-editor.org/info/rfc4361>.

   [RFC4388] Woundy, R. and K. Kinnear, "Dynamic Host Configuration
              Protocol (DHCP) Leasequery", RFC 4388,
              DOI 10.17487/RFC4388, February 2006,
              <http://www.rfc-editor.org/info/rfc4388>.

   [RFC4578] Johnston, M. and S. Venaas, Ed., "Dynamic Host
              Configuration Protocol (DHCP) Options for the Intel
              Preboot eXecution Environment (PXE)", RFC 4578,
              DOI 10.17487/RFC4578, November 2006,
              <http://www.rfc-editor.org/info/rfc4578>.

   [RFC4702] Stapp, M., Volz, B., and Y. Rekhter, "The Dynamic Host
              Configuration Protocol (DHCP) Client Fully Qualified
              Domain Name (FQDN) Option", RFC 4702,
              DOI 10.17487/RFC4702, October 2006,
              <http://www.rfc-editor.org/info/rfc4702>.

   [RFC4776] Schulzrinne, H., "Dynamic Host Configuration Protocol
              (DHCPv4 and DHCPv6) Option for Civic Addresses
              Configuration Information", RFC 4776,
              DOI 10.17487/RFC4776, November 2006,
              <http://www.rfc-editor.org/info/rfc4776>.

   [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy
              Extensions for Stateless Address Autoconfiguration in
              IPv6", RFC 4941, DOI 10.17487/RFC4941, September 2007,
              <http://www.rfc-editor.org/info/rfc4941>.

   [RFC6225] Polk, J., Linsner, M., Thomson, M., and B. Aboba, Ed.,
              "Dynamic Host Configuration Protocol Options for
              Coordinate-Based Location Configuration Information",
              RFC 6225, DOI 10.17487/RFC6225, July 2011,
              <http://www.rfc-editor.org/info/rfc6225>.

   [RFC6926] Kinnear, K., Stapp, M., Desetti, R., Joshi, B., Russell,
              N., Kurapati, P., and B. Volz, "DHCPv4 Bulk Leasequery",
              RFC 6926, DOI 10.17487/RFC6926, April 2013,
              <http://www.rfc-editor.org/info/rfc6926>.

   [RFC7724] Kinnear, K., Stapp, M., Volz, B., and N. Russell, "Active
              DHCPv4 Lease Query", RFC 7724, DOI 10.17487/RFC7724,
              December 2015, <http://www.rfc-editor.org/info/rfc7724>.

Authors' Addresses

   Sheng Jiang
   Huawei Technologies Co., Ltd
   Q14, Huawei Campus, No.156 Beiqing Road
   Hai-Dian District, Beijing, 100095
   P.R. China

   Email: jiangsheng@huawei.com

   Suresh Krishnan
   Ericsson
   8400 Decarie Blvd.
   Town of Mount Royal, QC
   Canada

   Phone: +1 514 345 7900 x42871
   Email: suresh.krishnan@ericsson.com

   Tomek Mrugalski
   Internet Systems Consortium, Inc.
   950 Charter Street
   Redwood City, CA 94063
   USA

   Email: tomasz.mrugalski@gmail.com