Abstract

Protecting Internet of Things (IoT) devices against attacks is not enough. IoT deployments need to make sure that they are not used for Distributed Denial-of-Service (DDoS) attacks. DDoS attacks are typically done with compromised devices or with amplification attacks using a spoofed source address. This document gives examples of different theoretical amplification attacks using the Constrained Application Protocol (CoAP). The goal with this document is to raise awareness and to motivate generic and protocol-specific recommendations on the usage of CoAP. Some of the discussed attacks can be mitigated by not using NoSec or by using the Echo option.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at https://t2trg.github.io/t2trg-amplification-attacks/draft-irtf-t2trg-amplification-attacks.html. Status information for this document may be found at https://datatracker.ietf.org/doc/draft-irtf-t2trg-amplification-attacks/.

Discussion of this document takes place on the Thing-to-Thing Research Group mailing list (mailto:t2trg@irtf.org), which is archived at https://mailarchive.ietf.org/arch/search/?email_list=t2trg. Subscribe at https://www.ietf.org/mailman/listinfo/t2trg/.

Source for this draft and an issue tracker can be found at https://github.com/t2trg/t2trg-amplification-attacks.

Status of This Memo

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

One important protocol used to interact with Internet of Things (IoT) sensors and actuators is the Constrained Application Protocol (CoAP) [RFC7252]. CoAP can be used without security in the so called NoSec mode but any Internet-of-Things (IoT) deployment valuing security and privacy would use a security protocol such as DTLS [RFC9147], TLS [RFC8446], or OSCORE [RFC8613] to protect CoAP, where the choice of security protocol depends on the transport protocol and the presence of intermediaries. The use of CoAP over UDP and DTLS is specified in [RFC7252] and the use of CoAP over TCP and TLS is specified in
[RFC8323]. OSCORE protects CoAP end-to-end with the use of COSE
[RFC8152] and the CoAP Object-Security option [RFC8613] and can
therefore be used over any transport. Group OSCORE
[I-D.ietf-core-oscore-groupcomm] can be used to protect CoAP Group
Communication [I-D.ietf-core-groupcomm-bis].

Protecting Internet of Things (IoT) devices against attacks is not
enough. IoT deployments need to make sure that they are not used for
Distributed Denial-of-Service (DDoS) attacks. DDoS attacks are
typically done with compromised devices or with amplification attacks
using a spoofed source address. DDoS attacks is a huge and growing
problem for services and critical infrastructure [DDoS-Infra] and
mitigations are costly.

The document gives examples of different theoretical amplification
attacks using CoAP. When transported over UDP, the CoAP NoSec mode is
susceptible to source IP address spoofing and as a single request can
result in multiple responses from multiple servers, CoAP can have
very large amplification factors. The goal with this document is to
raise awareness and understanding of amplification attacks and to
motivate mitigations suitable for constrained devices and networks.
The intent is not to suggest that CoAP is more vulnerable to
amplification attacks than other protocols.

Some of the discussed attacks can be mitigated by not using NoSec or
by using the Echo option [RFC9175].

2. Amplification Attacks using CoAP

In a Denial-of-Service (DoS) attack, an attacker sends a large number
of requests or responses to a target endpoint. The denial-of-service
might be caused by the target endpoint receiving a large amount of
data, sending a large amount of data, doing heavy processing, or
using too much memory, etc. In a Distributed Denial-of-Service (DDoS)
attack, the request or responses come from a large number of sources.

In an amplification attack, the amplification factor is the ratio
between the total size of the data sent to the target and the total
size of the data received from the attacker. Note that in the
presence of intermediaries, the size of the data received by the
target might be different than the size of the data sent to the
target and the size of the data received from the attacker might be
different than the size of the data sent from the attacker.

In the attacks described in this section, the attacker sends one or
more requests, and the target receives one or more responses. By
spoofing the source IP address of the targeted victim and requesting
as much information as possible from several servers an attacker can
multiply the amount of traffic and create a distributed denial-of-
service attack on the target. When transported over UDP, the CoAP NoSec mode is susceptible to source IP address spoofing.

The amplification factor and the bandwidth depend on the layer in the protocol stack that is used for the calculation. The amplification factor and bandwidth can e.g., be calculated using whole IP packets, UDP payloads, or CoAP payloads. The bandwidth decreases and the amplification factor typically increases higher up in the protocol stack. The bandwidth should be calculated using the layer that is considered to be under attack.

The following sections give examples of different theoretical amplification attacks using CoAP.

2.1. Simple Amplification Attacks

An amplification attack using a single response is illustrated in Figure 1. If the response is $c$ times larger than the request, the amplification factor is $c$.

Victim  Foe  Server

```
| GET | Code: 0.01 (GET) |
|     | Uri-Path: random quote |
| 2.05 | Code: 2.05 (Content) |
|      | Payload: "just because you own half the county doesn't mean that you have the power to run the rest of us. For twenty-three years, I've been dying to tell you what I thought of you! And now... well, being a Christian woman, I can't say it!"
```

Figure 1: Amplification attack using a single response

An attacker can increase the bandwidth by sending several GET requests. If the server supports PUT/POST and doesn't limit the payload size, an attacker may be able to increase the amplification factor by creating or updating a resource. By creating new resources, an attacker may also increase the size of /.well-known/core. An amplification attack where the attacker influences the amplification factor is illustrated in Figure 2.
2.2. Amplification Attacks using Observe

Amplification factors can be significantly worse when combined with observe [RFC7641]. As a single request can result in multiple responses from the server, the amplification factors can be very large.

An amplification attack using observe is illustrated in Figure 3. If each notification response is $c$ times larger than the registration request and each request results in $n$ notifications, the amplification factor is $c \cdot n$. 

![Amplification attack using several requests and a chosen amplification factor](image-url)
A more advanced amplification attack using observe is illustrated in Figure 4. By registering the same client several times using different Tokens or port numbers, the bandwidth can be increased. By updating the observed resource, the attacker may trigger notifications and increase the size of the notifications.

If the server supports the pmax conditional attribute \[\text{I-D.ietf-core-conditional-attributes}\] an attacker may increase the frequency of notifications and therefore the amplification factor. The maximum period attribute pmax indicates the maximum time, in seconds, between two consecutive notifications (whether or not the resource state has changed). Servers should put limits on the pmax values they accept.

If it is predictable when notifications are sent as confirmable and which Message ID are used the acknowledgements may be spoofed.
2.3. Amplification Attacks using Group Requests

Amplification factors can be significantly worse when combined with observe [RFC7641]. As a single request can result in responses from multiple servers, the amplification factors can be very large.

As the group request is sent over multicast or broadcast the request has to be sent by a node on the local network. An attacker on a local network (e.g., a compromised node) might use local CoAP servers to attack targets on the Internet or on the local network. The servers
might also be accessible through a gateway (GW) that transforms unicast requests into multicast request. In such architectures, amplification attacks using group requests might be possible to launch from the Internet.

An amplification attack using a group request is illustrated in Figure 5. A single unicast request results is transformed into a multicast group request by the gateway and results in m responses from m different servers. If each response is c times larger than the request, the amplification factor is $c \cdot m$. Note that the servers usually do not know the variable m.

![Diagram of Amplification Attack Using Multicast](image)

Figure 5: Amplification attack using multicast

Even higher amplification factors can be achieved when combining group requests [I-D.ietf-core-groupcomm-bis] and observe [RFC7641]. In this case a single request can result in multiple responses from multiple servers.

An amplification attack using a multicast request and observe is illustrated in Figure 6. In this case a single request results in n responses each from m different servers giving a total of $n \cdot m$ responses. If each response is c times larger than the request, the amplification factor is $c \cdot n \cdot m$. 
Figure 6: Amplification attack using multicast and observe

An attacker can use the same techniques as in Figure 4 to increase the number of notifications.

2.4. MITM Amplification Attacks

TLS and DTLS without Connection ID [RFC9146][RFC9147] validate the IP address and port of the other peer, binds them to the connection, and do not allow them to change. DTLS with Connection ID allows the IP address and port to change at any time. As the source address is not protected, an MITM attacker can change the address. Note that an MITM attacker is a more capable attacker then an attacker just spoofing the source address. It can be discussed if and how much such an attack is reasonable for DDoS, but DTLS 1.3 states that "This attack is of concern when there is a large asymmetry of request/response message sizes." [RFC9147].

DTLS 1.2 with Connection ID [RFC9146] requires that "the receiver MUST NOT replace the address" unless “there is a strategy for ensuring that the new peer address is able to receive and process
DTLS records” but does not give more details than that. It seems like
the receiver can start using the new peer address and test that it is
able to receive and process DTLS records at some later point. DTLS
1.3 with Connection ID [RFC9147] requires that "implementations MUST
NOT update the address" unless “they first perform some reachability
test” but does not give more details than that. OSCORE [RFC8613] does
not discuss address updates, but it can be assumed that most servers
send responses to the address it received the request from without
any reachability test. A difference between (D)TLS and OSCORE is that
in DTLS the updated address is used for all future records, while in
OSCORE a new address is only used for responses to a specific
request.

An MITM amplification attack updating the client's source address in
an observe registration is illustrated in Figure 7. This attack is
possible in OSCORE and DTLS with Connection ID. The server will send
notifications to the Victim until it at some unspecified point
requires an acknowledgement [RFC7641]. In DTLS 1.2 the reachability
test might be done at a later point. In OSCORE a reachability test is
likely not done.

![Diagram](image)

**Figure 7**: MITM Amplification attack by updating the client's source
address in an observe registration request

Where 'S' means the MITM attacker is changing the source address of
the message and 'D' means the MITM attacker is changing the
destination address of the message.
An MITM amplification attack updating the server's source address is illustrated in Figure 8. This attack is possible in DTLS with Connection ID. In DTLS 1.2 the reachability test might be done at a later point.

![Client Foe Victim Server Diagram]

Figure 8: MITM Amplification attack by updating the server's source address in a response

3. Summary

CoAP has always considered amplification attacks, but most of the requirements in [RFC7252], [RFC7641], [RFC9175], and [I-D.ietf-core-groupcomm-bis] are "SHOULD" instead of "MUST", it is undefined what a "large amplification factor" is, [RFC7641] does not specify how many notifications that can be sent before a potentially spoofable acknowledgement must be sent, and in several cases the "SHOULD" level is further softened by "If possible" and "generally". [I-D.ietf-core-conditional-attributes] does not have any amplification attack considerations.

QUIC [RFC9000] mandates that "an endpoint MUST limit the amount of data it sends to the unvalidated address to three times the amount of data received from that address" without any exceptions. This approach should be seen as current best practice for non-constrained devices.

While it is clear when a QUIC implementation violates the requirement in [RFC9000], it is not clear when a CoAP implementation violates the
requirement in [RFC7252], [RFC7641], [RFC9175], and [I-D.ietf-core-groupcomm-bis].

In CoAP, an address can be validated with a security protocol or by using the Echo Option [RFC9175]. Restricting the bandwidth per server is not enough as the number of servers the attacker can use is typically unknown. For multicast requests, anti-amplification limits and the Echo Option do not really work unless the number of servers sending responses is known. Even if the responses have the same size as the request, the amplification factor from m servers is m, where m is typically unknown. While DoS attacks from CoAP servers accessible over the Internet pose the largest threat, an attacker on a local network (e.g., a compromised node) might use local CoAP servers to attack targets on the Internet or on the local network.

4. Security Considerations

The whole document can be seen as security considerations for CoAP.

5. IANA Considerations

This document has no actions for IANA.

6. Informative References


Acknowledgements

The authors would like to thank Sultan Alshehri, Carsten Bormann, Klaus Hartke, Jaime Jiménez, Ari Keränen, Matthias Kovatsch, Achim Kraus, Sandeep Kumar, and András Méhes for their valuable comments and feedback.

Authors' Addresses

John Preuß Mattsson  
Ericsson AB  
SE-164 80 Stockholm  
Sweden  
Email: john.mattsson@ericsson.com

Göran Selander  
Ericsson AB  
SE-164 80 Stockholm  
Sweden  
Email: goran.selander@ericsson.com

Christian Amsüss  
Energy Harvesting Solutions  
Email: c.amsuess@energyharvesting.at