Email Transport Encryption STARTTLS vs. DANE vs. MTA-STS

Delimitation

This document deals with the encrypted transport of messages between two email servers. Encryption during transport is crucial for a basic level of security for the exchange of messages.

The exchange of messages between an email client and a server or the end-to-end encryption of messages are not covered in this article. If the aim is to create a secure overall system, these aspects should be considered in addition to the recommendations in this article.

Initial situation

When the first standard for the Simple Mail Transfer Protocol (in short: SMTP) was adopted, encryption of the transport was not included. All messages were exchanged as plain text between the servers. They were thus basically readable for anyone who could tap into the data exchange between the two servers.

This only changed with the introduction of the SMTP Service Extensions for Extend SMTP (ESMTP for short), which made it possible to choose encrypted transport as an opportunistic feature of data exchange. Opportunistic, because it had to be assumed that not all servers would be able to handle transport encryption. They should only encrypt when it is opportune; encryption is not mandatory.

To indicate the basic ability to encrypt to a sending server, it was agreed that the receiving server should send the keyword STARTTLS at the beginning of an ESMTP transport session.

STARTTLS

STARTTLS can encrypt the transport between two email servers. The receiving server signals the sending server that it is capable of encrypting after a connection is established and in the ESMTP session. If the sending server is basically able to perform encrypted

transport, it can accept the offer and request the receiving server to start a TLS-encrypted connection.

The two servers then negotiate under which conditions they want to encrypt the transport. If both servers can agree on the criteria for an encrypted ESMTP session, the receiving server creates a session key. This is only used between these two servers – and only for a certain time – to encrypt the communication between the two parties.

STARTTLS allows secure encrypted transport when an encrypted session has been established, but it has critical flaws in session setup and session key generation.

The security researchers Diffie and Hellman demonstrated early on how easy it was to decrypt an encrypted transport subsequently, even if the attacker were only in possession of the session key and a record of the encrypted communication of both email servers. They also showed how this attack, called Session Replay, could be prevented by the use of additional keys, thus achieving Perfect Forward Secrecy.

For a long time, no countermeasures were established for other attack vectors. For example, the sending server only learns during the session with the other email server that the latter is capable of encrypted transport. An attacker who can control the data stream of the two servers can suppress the ESMTP capability or the STARTTLS keyword and thus force the sending server to "downgrade" the session. The communication then takes place without encryption and the attacker can record in plain text all messages exchanged.

Another attack vector is the Man-in-the-Middle Attack. In this form of attack, the attacker places their own system between the two email servers and pretends to be the target system that can also offer encrypted transport.

A sending server which can "only" do STARTTLS is not able to recognize the Man-in-the-Middle without further protective measures, because SMTP does not offer any methods with which servers can obtain the identity characteristics of the counterpart before establishing a connection and then compare these after establishing a connection.

When deceived in this manner by the attacker, a server can therefore communicate with the wrong server and transmit data that should be protected, in the belief that it is communicating with the right server. The communication can even be encrypted.

DANE remedies these two vulnerabilities.

DANE

The method "DNS-Based Authentication of Named Entities" (short: DANE) fixes the above-mentioned vulnerabilities of STARTTLS. DANE describes an abstract procedure which can basically be used in many protocols, not only in SMTP. This document is limited to the implementation of DANE for SMTP.

DANE requires the storage of information in the DNS of the target domain, which, even before the connection to the target server is established, shows that the server supports STARTTLS and by which characteristics the target server can be identified. If the sending server knows in advance that the target system supports STARTTLS, it cannot become the victim of a session downgrade attack. And if it knows by which characteristics it can identify its counterpart, it cannot fall victim to a Man-in-the-Middle attack, because only the right server has those specific characteristics.

Responses from traditional DNS servers are not suitable for DANE because they can be faked. An attacker could attempt to redirect the sending server to the wrong destination. Therefore, the DANE method requires that DNS servers which want to issue DANE-relevant responses use the DNS Security Extensions (DNSSEC). DNSSEC is used to cryptographically sign DNS server responses. The sending server can check if the "correct DNS server" has responded and if the DNS response has arrived uncompromised.

DANE changes the trust model that is the basis for encryption with certificates. It no longer relies on the chain of trust of a Certification Authority (CA), but it anchors the trust in the DNS zone of the target system. With the DANE method, it is irrelevant whether the present certificate is a self-signed or expired certificate. As long as the DNSSEC-signed zone shows the certificate to be valid, it can be used.

DANE requires DNSSEC and this hinders its distribution on the recipient side, because the proportion of DNSSEC-signed DNS zones is small in global terms.

The low penetration of DNSSEC does not prevent senders from using DANE. Their email servers only have to check if the target system offers DNSSEC and DANE.

If both are available, it can fully use DANE security to protect transport encryption.

Don Owens of Cisco Talos Intelligence Group on DANE:

"DANE is an excellent way of encrypting data traffic while avoiding MITM (Man-in-the-Middle) attacks. If DANE is not supported by the recipient, it is possible to make MITM attacks more difficult with MTA-STS. Another option is to use opportunistic TLS to provide a layer of security, even though it is more vulnerable to MITM attacks. But really, recipients should set up DANE wherever possible to ensure the best possible security for email traffic."

Critics of DANE, who object to the fact that it requires DNSSEC, have therefore specified MTA-STS.

MTA-STS

SMTP MTA Strict Transport Security (MTA-STS) improves STARTTLS without requiring DNSSEC. It leaves the trust with the CA and establishes a "trust on first use" (TOFU) model. Guidelines for "correct sending" are obtained from the domain of the target system when the first connection is established and are followed from then on. In this way, MTA-STS can signal to a server that it is to expect STARTTLS, protecting it from a session downgrade attack.

This only works if an attacker has not "poisoned" the sender's DNS with false information (cache poisoning). When DNSSEC is used, this type of attack is excluded on principle.

Furthermore, it presupposes that an attacker is not in possession of a certificate suitable for the target domain. MTA-STS relies on CAs to implement control mechanisms that prevent multiple issuance of a certificate for a target system.

MTA-STS cannot protect against a Man-in-the-Middle attack, because it does not provide the sender with criteria with which it can uniquely identify the target system. Here too, MTA-STS relies on the trust model of the CA system.

Summary and conclusion

Traditional SMTP does not protect messages in transit. STARTTLS encrypts the transport, but it provides attack points that make it possible for transport encryption to be skipped or enable encrypted communication with the wrong system.

DANE encrypts the transport and is immune to the attacks mentioned above on principle. This requires a DNSSEC-signed zone on the receiver side. Because DNSSEC is perceived as complex and complicated, it stands in the way of the dissemination of DANE. MTA-STS overcomes this hurdle.

MTA-STS encrypts the transport and makes it difficult to exploit the vulnerabilities mentioned above. However, it does not remedy these, because it is on principle unable to do so.

If you want to secure the transmission of email, you should use DANE for sending, because it does not require your own DNS zone to be secured with DNSSEC. Whenever your own system encounters a DANE-activated target system during transmission, the advantages of DANE can come into play.

For your own domain, it makes sense to offer MTA-STS for senders as a first step, even though there are hardly any MTA-STS-capable email servers available so far. It can increase the level of security compared to traditional STARTTLS.

In the medium to long term, your own domain should be secured with DNSSEC. The next step to becoming a DANE-secured domain is easy. Then senders can take advantage of DANE's unmatched security gain and the encrypted transport of email is effectively and comprehensively protected.

Author: Patrick Koetter,
Leader Competence Group E-Mail,
Board member sys4 AG