FHA Directed Exchange Headers Issue Paper

Version 1.0
July, 2014

DISTRIBUTION STATEMENT
Distribution authorized to Federal Health Architecture (FHA) representatives.

Federal Health Architecture Program Management Office | Office of the National Coordinator for Health IT | Department of Health and Human Services

Capital View Office Building
425 3rd Street SW, Washington, DC 20004

Website: http://healthIT.gov/fha General Email: federal.health@hhs.gov

Document Change Control

Version	Release Date	Summary of Changes	Addendum Number	Name
1.0	3/5/14			Scott Rea (DigiCert)
1.1	4/14/14	Editorial reorganization for clarity		John "Mike" Davis (DVA)
1.2	6/2/14	-Reference rfc822 -Edits to introduction		Russ Davis (DoD)
1.3	6/26/14	Update to Summary Item 3 from Greg Meyer (CERNER) – no MDN to avoid DoS attempts		Scott Rea (DigiCert)

Issue

Headers used to address Directed Exchange messages may be compromised. Additionally, in a proxy attack, the bad actor could take a legitimate message header and replace the attached payload.

BACKGROUND

When sending Direct messages, Header fields are by necessity sent in the clear, so that the message can be routed to the appropriate destination, and validation of source information can occur. However, this opens up the possibility for exploitation of Direct messages, especially if a domain-based credential (e.g. Direct Organizational Certificate (Org cert)), is used in securing the transmission of the message (this Direct exploit was identified and demonstrated in 2011).

PROBLEM DESCRIPTION

The exploitation occurs when a malicious party intercepts the message and modifies the Healthcare Organization (HCO) end point referenced in the Header to point to ANY other end point within that domain. The message would be deliverable and readable by the party at that modified address. ***NOTE:*** this does restrict the scope of this redirected attack to only those end points (Direct addresses) provisioned in the recipient HCO domain.

A Sender of a message secured by an Org cert does not authoritatively identify the source HCO end point either; it only cryptographically binds the domain from which it comes. Any end point in that HCO domain could create the same message, opening up the system for potential spoofing of the source end point of the message. Again, the scope of this spoofing attack is restricted to only those HCO end points (Direct addresses) provisioned in the sending HCO domain, and the attack must be carried out BEFORE the Org cert is used to create the Direct Message (the attack surface for this exploit is therefore restricted to the internal HCO messaging system). This might be a significant risk for Electronic Health Records (EHR) that prepare base message components and then utilize 3rd parties to format and send the Direct Messages.

Solution 1: Direct Address Certificates

Direct Address certificates at the receiving HCO will also mitigate the redirect attack identified above. Incidentally, the use of Direct Address certificates also protect against source spoofing. Direct Address certificates to secure the sending HCO end points (source), along with a reliable account provisioning and strong account authentication process deployed ensures that source end points are cryptographically bound to the security certificates used. Use of Direct Address certificates to secure both ends of a Direct message will mitigate both the source spoofing and destination redirect risk introduced by the use of Org certs.

Solution 2. S/MIME message/rfc822 wrapper

RFC 5751[1], Section 3.1 describes Message Wrapping[2] as a means to mitigate the risk of an Org cert redirect attack as identified above. The sending client MAY wrap a full MIME message in a message/rfc822 wrapper resulting in S/MIME security services to the header. Messages received will be identified by Content-Type message/rfc822. Message Wrapping entails taking the original message (Headers included) and signing it with the Senders Direct Certificate and encrypting it with the Receiver's Direct Certificate to create a valid Internet Message Format (IMF) document (as if creating a standard Direct message). This IMF document is then used as a new message body, and Headers matching those encrypted in that document are added to create a new Direct message by wrapping the entire original message. The resulting Direct message has Headers in the clear, as is required for processing, but also contains the same Headers encrypted within the body of the message. A receiving agent is then able to verify that Headers were not tampered with by decrypting the message and comparing the outer "in-the-clear" Headers to the inner "encrypted" Headers. It should be noted that Message Wrapping with comparison of inner and outer headers, can alert a receiving Security and Trust Agent (STA) when an exploit is being attempted, allowing for appropriate remediation.

The message/rfc822 wrapper is NOT a requirement within the Direct Applicability Statement, meaning STAs may exist that do not support this functionality. Some Direct trust communities, such as DirectTrust.org, may in future require Message Wrapping support within the STAs of Accredited HISPs in order to be accredited; however there is no such enforcement of this today.

The Reference Implementation (RI) for Direct implements Message Wrapping by default for sending messages, and automatically detects and "unwraps" any wrapped messages that are received. However, the RI does not do a comparison of inner and outer Headers to detect if tampering has occurred, it simply discards the outer Headers and takes the inner Headers as authoritative. This has the effect of only ever delivering to the originally intended recipient as required (thus mitigating the Org cert redirect attack), but it misses the opportunity to identify and respond to potentially malicious environments where these attacks are being attempted.

It is strongly recommended that ALL agents processing Direct messages be able to support message/rfc822 wrapping (sending and receiving), and only accept inner Header values as authoritative when wrapping is used. Agents should identify and log any Direct messages where the inner and outer Headers of a Wrapped Message do not correlate in order to facilitate response to malicious tampering (e.g. an MDN (Message Disposition Notification) back to the authenticated sender notifying of the tamper attempt would allow both sides the opportunity to respond to potential malicious parties).

---

[1] See http://tools.ietf.org/pdf/rfc5751.pdf

[2] See also http://wiki.directproject.org/Protecting+Transport+Headers for Direct explanation on Wrapping

ROLE OF TRUST BUNDLES

A Trust Bundle for Direct is a collection of Trust Anchors from HISPs that meet an established security policy. FHA is able to enforce its own specific implementation policies upon HISPs, CAs, and RAs in the Direct community, by defining a security policy that encapsulates its desired security posture related to Direct, and then creating or sponsoring a Trust Bundle for those TAs that represent those HISPs/CAs/RAs that have been verified as meeting the published security posture. Relying parties e.g., the FHA community of participants, who wish the security posture to be enforced across their operations, simply import the FHA Trust Bundle to be used in trust validation. If a Direct end entity certificate chains to one of the FHA TAs', then the relying party can be assured that the Direct participants represented by that certificate, are operating in an environment that FHA has prescribed.

It is suggested that FHA recommend a security policy for Direct, that enables FHA specific policies to be enforced for Direct e.g., end entity certificates issued under an FBCA cross-certified CA/policy, mandatory wrapping of messages sent to addresses protected by Organizational certificates, and mandatory unwrapping of messages received from organizational certificate protected accounts with comparison between internal and external "To" address and mandatory logging of disparities etc. Due to the expected broad adoption of this FHA specific policy set, it is recommended that FHA invite commercial trust frameworks to host and manage the FHA Trust Bundle to enforce demonstration of capabilities for meeting the controls before a HISP: TA tuple is added to the bundle. FHA and any other relying party seeking to be compatible with FHA can then adopt the FHA Trust Bundle for trust purposes, which ensures the FHA controls are being met.

Summary

The use of Direct Address certificates to secure both ends of a Direct message will mitigate both the source spoofing and destination redirect risk introduced by the use of Org certs at sending and receiving HCOs. If use of Address certificates is not available or practical for the participating parties, then the redirect attack may be mitigated by Message Wrapping, and an explicit agreement between the two parties to only use wrapping when Org certs are utilized. In absence of such an agreement, reliance upon trust frameworks that enforce wrapping for accredited parties (e.g. in future for DirectTrust) may also mitigate this risk. The following recommendations are provided as fundamental to mitigating the risk of compromise to Directed Exchange Message Headers:

1. ALL agents processing Direct messages should support message/rfc822 wrapping (sending and receiving), and only accept inner Header values as authoritative when wrapping is used.

2. FHA Directed Exchange should recommend a security policy for Direct, that enables FHA specific policies to be enforced for Direct , specifically, certificates (address or domain bound) issued under an FBCA cross-certified CA/policy, mandatory wrapping of messages sent to addresses protected by Organizational certificates, and mandatory unwrapping of messages received from organizational certificate protected accounts with comparison between internal and external "To" address and mandatory logging of disparities etc.
3. Agents should identify and log any Direct messages where the inner and outer Headers of a Wrapped Message do not correlate in order to facilitate response to malicious tampering. An out of band notification to the sender is recommended to allow both sides the opportunity to respond to potential malicious events in accordance to your existing agency policy on breech. NOTE: an MDN (Message Disposition Notification) should NOT be returned to the sender if a tamper attempt is discovered. A tamper attempt should be considered an untrusted message and therefore sending MDNs or any type of systemic message to any source of tampered or untrusted messages opens the door to DoS attempts.
4. FHA should invite commercial trust frameworks to host and manage a FHA Federal Common Trust Bundle. This provides opportunity to enforce demonstration of capabilities for meeting the controls before a HISP: Trust Anchors and metadata tuple is added to the bundle. FHA and any other relying party seeking to be compatible with FHA can then adopt the FHA Trust Bundle for trust purposes, which ensures the FHA controls are being met.