**Allowing Monitoring of the Census Trusted Internet Connection with Einstein 3 Accelerated and TLS Decryption While Preserving Respondent Confidentiality**

# Background

As part of its Trusted Internet Connection program[[1]](#footnote-1), the Department of Homeland Security (DHS) is requiring that civilian government agencies have their internet connections monitored by the Einstein 3 network monitoring system.

According to published reports and other sources, Einstein 3 Accelerated[[2]](#footnote-2) (E3A) has the following capabilities:

* Monitoring IP source & destination addresses (flow analysis)
* Monitoring of email headers (e.g. To: and From: fields) (email analysis)
* Decryption of Hypertext Transfer Protocol Secure (HTTPS) and other protocols encrypted with Transport Layer Security (TLS)[[3]](#footnote-3)

These capabilities are used to provide the following functions:

* Protection of agency information systems from known attackers and ongoing attacks
* Discovery of new attackers and attacks
* Government-wide situational awareness

### SSL/TLS Decryption

According to the May 6, 2016 Privacy Impact Assessment Update, “WCF [Web Content Filtering] capabilities also include in-line Secure Socket Layer (SSL) decryption.”[[4]](#footnote-4) SSL is an obsolete term for the Transport Layer Security encryption protocol used by the HTTPS internet protocol to protect communications between web browsers and web servers. By design, TLS provides for confidentiality and integrity of the connection: it assures that a third party cannot eavesdrop on the contents of the connection, and the data sent by the browser is received by the server without modification.

TLS is based on public key cryptography. Each web server is configured with a *private key* and a *public key*, which is contained in a public key certificate. Information encrypted with the private key can only be decrypted with the public key. When the web browser connects to the web server, the server sends the browser its public key certificate. The browser then creates a randomly generated session key, encrypts the key with the public key and sends it to the server. The server uses its private key to decrypt the session key so that the TLS-encrypted communications can take place. In its final deployment, there is a possibility that DHS will require that E3A be provided with a copy of each web servers’ private key so that the TLS connections can be decrypted.

There are compelling reasons for DHS to decrypt the data sent by TLS from the browser to the server. Many attacks and exploits are based on the idea of sending unexpected values to the server. For example, consider a web form with two fields, age and gender:

Age: \_\_\_\_\_\_\_\_\_  
 Gender: \_\_\_\_\_\_\_\_\_\_  
 [SUBMIT]

In normal operation, this information is encoded as an HTTP request and sent to the web server:

Age=52&gender=male

However, an attacker might attempt an SQL injection attack, for example:

Age=52&gender=male’;drop table RESPONSES;

In theory, E3A could detect this SQL injection attack and break the TLS connection, then blacklist the IP address associated with the web browser so that the system could not attack any other US Government web server protected by E3A. *Even if Census web applications are not vulnerable to SQL injection attacks, would-be attackers will attempt SQL injection attacks to probe the servers. Detecting these attacks will allow DHS to blacklist the attackers; analyzing these attacks will allow DHS to improve the defenses provided by E3A.*

As described, E3A appears to violate the Census Bureau’s confidentiality pledge. For example, census publication BC-1428RV states:

“Our highest priority at the Census Bureau is using your answers only for determining how the country is doing. By law, we cannot share your answers with anyone—including other federal agencies and state and local law enforcement agencies. In 2010, the Justice Department determined that no provision of the U.S. Patriot Act overrides the provisions of Title 13 that protects the confidentiality of Census Bureau data.”[[5]](#footnote-5)

The Department of Homeland Security is both another federal agency and a law enforcement agency. By using a TIC that incorporates E3A, Census would appear to be sharing respondent data with DHS. Even more troubling, there is a chance that, while analyzing an E3A alert, a DHS employee might directly uncover evidence of illegal activity on the part of the respondent. In such a circumstance, the DHS employee might be legally obligated to respond. Thus, it is vital that DHS be able to verify that data sent from the web browser to the web server not contain SQL injection attacks (or other exploits) while being shielded from seeing respondent data.

Acknowledging this possibility, the Department of Commerce negotiated an Addendum to its Memorandum of Agreement with DHS that

1. requires immediate notification of by CS&C (DHS) anytime it appears that there has been unauthorized access or use of DOC information obtained through EINSTEIN, and defines unauthorized access as “access or use for reasons other than protecting information and information systems from cybersecurity risks...”
2. requires prompt notification if DHS disseminates DOC information obtained through EINSTEIN that is “visibly marked as protected by statute, ...”
3. requires quarterly reporting by DHS to DOC of how frequently DHS personnel “manually access” DOC's statutorily protected information; and
4. exempts Census from the extremely onerous splash screen announcing the presence of DHC screening on government internet portals.

As a result of the memorandum, Census requested emergency clearance for a change in its general pledge from:

The U.S. Census Bureau is required by U.S. law to keep your answers confidential. This means that the Census Bureau cannot give out information that identifies you <this business/organization/government entity>, including to other government agencies. Federal law protects your privacy and keeps your answers confidential (Title 13, United States Code, Sections 9 and 214).

to

The U.S. Census Bureau is required by law to protect your information. The Census Bureau is not permitted to publicly release your responses in a way that could identify you <this business/organization/government entity>. We are conducting this survey under the authority of Title 13, United States Code, Sections 141 and 193. Federal law protects your privacy and keeps your answers confidential (Title 13, United States Code, Sections 9 and 214). Per the Federal Cybersecurity Enhancement Act of 2015, your data are protected from cybersecurity risks through screening of the systems that transmit your data.

On December 14, 2016, a new generic statistical confidentiality pledge for the Census Bureau’s data collections was published in the Federal Register:

The U.S. Census Bureau is required by law to protect your information. The Census Bureau is not permitted to publicly release your responses in a way that could identify you. Per the Federal Cybersecurity Enhancement Act of 2015, your data are protected from cybersecurity risks through screening of the systems that transmit your data.[[6]](#footnote-6)

# Possible Solutions

This paper proposes the framework for a new solution to the problem of DHS monitoring based on *web application level cryptography* to shield respondent data.

## Web Application Level Cryptography

Modern web applications consist of JavaScript code that is downloaded into the web browser that communicates with code that is running on the web server. Typically, JavaScript is used for animation, form validation, responsive layout, and the like. The JavaScript can also be used to implement cryptography. The W3C Web Cryptography API[[7]](#footnote-7) is a proposed recommendation for integrating some cryptographic libraries into web browsers, but there exist many JavaScript libraries for performing a wide range of cryptographic primitives within the browser today, even without the adoption of the Cryptography API.

The framework of the solution is to have the web browser and server derive an application-level per-session encryption key that will be used to encrypt the specific respondent answers sent from the web browser to the server. For example, JavaScript running on the browser and at the server might engage in a Diffie-Hellman key exchange protocol to derive a shared encryption key. Alternatively, the server and the browser and employ the Paillier cryptosystem to transmit a session key from the browser to the server in a probabilistic manner that would not be susceptible to third-party monitoring. Once a per-session encryption key is derived between the client and the server, this key would be used to provide field-level encryption of respondent values.

To use application-level cryptography, it is first necessary to convert the responses into a numeric form. For example, original response becomes:

Age=52&gender=1

With application-level cryptography, the encrypted response might become:

Age=343126314&gender=236134

This message can be viewed by E3A without compromising the respondent’s information. When it is received by the server, which has knowledge of the shared encryption key, the values are decrypted and handled as before.

This simplistic approach can be improved:

* Instead of preserving the original fields, the JavaScript on the web browser could prepare and encrypt a block of numbers that includes checksums and other error detecting codes. For example, the response might become:

Response=6a6a550621f3d360dbabb15a4644e72ajjU

The server would then decrypt the response and verify the checksums, ensuring that the client-side cryptography is performing properly.

## Deployment requirements

To deploy a web application-level cryptography solution to an existing survey, the following would be needed:

* A JavaScript library providing the necessary cryptographic primitives. Currently there is no FIPS-140 validated JavaScript library, so a waiver might be required. On the other hand, existing browsers do not have FIPS-140 validated cryptography either, so this may not be a concern.
* Modifications to the browser-side JavaScript that displays the survey forms so that the contents would be encrypted prior to being transmitted.
* Modifications to the survey web server so that the application-level public key would be provided to the web browser and the survey results would be decrypted.
* Although probably not required, it would be useful to have DHS approve the use of application-level cryptography that they cannot decrypt, just so that DHS does not later argue that E3A needs to be able to decrypt survey responses for some reason.

## Providing Respondent Anonymity with Proxy Servers

One aspect of respondent anonymity is protection against an outsider knowing if a person has responded to a survey. Because E3A records IP addresses of incoming HTTPS connections and because many IP addresses can be linked to individuals, a DHS analyst could determine for many respondents if they had participated in a survey. This may violate the Census confidentiality pledge.

The only way to hide source IP addresses is by using a proxy. There are two approaches that Census might employ:

1. Census could engage a single-hop proxy such as Anonymizer.[[8]](#footnote-8)
2. Census could make its services available as a “hidden service” using the Tor anonymity service.[[9]](#footnote-9) Respondents could access the Census Tor server using either the Tor client software, which they would download to their computers, or using a tor2web bridge. There are several such tor2web bridges currently in operation, and Census could engage others on a contractual basis.

## Conclusion

There exist models that could be used to allow DHS to monitor the Census web servers, using SSL/TLS decryption, while still providing for respondent confidentiality and anonymity. Although there is currently no off-the-shelf software for accomplishing this goal, Census could create such software with minimal engineering effort.

1. https://www.dhs.gov/trusted-internet-connections [↑](#footnote-ref-1)
2. See <https://www.dhs.gov/publication/einstein-3-accelerated> [↑](#footnote-ref-2)
3. Note that Transport Layer Security (TLS) was originally called Secure Socket Layer (SSL), and many documents still call it by that name. [↑](#footnote-ref-3)
4. https://www.dhs.gov/sites/default/files/publications/privacy-pia-nppd-einstein3a-may2016.pdf [↑](#footnote-ref-4)
5. https://www.census.gov/prod/2001pubs/bc-1428rv.pdf [↑](#footnote-ref-5)
6. Federal Register December 14, 2016 https://www.federalregister.gov/documents/2016/12/14/2016-30014/confidentiality-pledge-revision-notice. [↑](#footnote-ref-6)
7. https://www.w3.org/TR/WebCryptoAPI/ [↑](#footnote-ref-7)
8. https://www.anonymizer.com/ [↑](#footnote-ref-8)
9. https://www.torproject.org/ [↑](#footnote-ref-9)