A More Cautious Approach to Security Against Mass Surveillance

Jean Paul Degabriele, *Pooya Farshim, and Bertram Poettering*

Royal Holloway, Queen's University Belfast, Ruhr University Bochum

FSE - 11th March 2015

ROYAL HOLLOWAY UNIVERSITY OF LONDON

Outline of this Talk

The Snowden Revelations

- Since June 2013 Edward Snowden has been disclosing classified documents about mass surveillance programs carried by the NSA and GCHQ.
- Until now, there has been no indication that these agencies are capable of breaking any of the main cryptographic primitives/assumptions which we believe to be secure/hard.
- Instead these agencies have resorted to more devious means:
  - Manoeuver standardisation bodies to advance the backdoored EC DRBG and the TLS Ext Random.
  - Secretly pay RSA to make the EC DRBG the default option in their cryptographic library.
  - Forcing vendors and service providers (through secret courts) to provide user data, secret keys, access to infrastructure, etc.
  - Intercept postal shipping to replace networking hardware.
  - Inject malware in network data carrying executable files.

Guarding Against Surveillance

- In light of these events it is natural to ask what other means could be employed by such entities.
- Following the Snowden revelations, a first step in this direction is the recent work of Bellare, Paterson and Rogaway from CRYPTO 2014 [BPR14].
- The focus of their study is Algorithm Substitution Attacks (ASA) with respect to symmetric encryption.

Algorithm Substitution Attacks

- Consider some type of closed-source software that makes use of a standard symmetric encryption scheme.
- In an ASA the code of the standard encryption scheme is replaced with that of an alternative scheme that the attacker has authored.
- Following the terminology of [BPR14] we call this latter scheme a **subversion** and we refer to the attacker as **big brother**.
- If the code is obfuscated can we protect against this?

Algorithm Substitution Attacks

- Consider some type of closed-source software that makes use of a standard symmetric encryption scheme.
- In an ASA the code of the standard encryption scheme is replaced with that of an alternative scheme that the attacker has authored.
- Following the terminology of [BPR14] we call this latter scheme a **subversion** and we refer to the attacker as **big brother**.
- If the code is obfuscated can we protect against this?

Algorithm Substitution Attacks

- Note that ASAs are different from backdoors, as in the case of the Dual EC DRBG.
- The focus here is whether an **implementation** of the scheme offers the claimed security. The original scheme is assumed to be secure and free from backdoors.
- ASAs have been considered in the past in the works of Young and Yung, and others, under the name of Kleptography. In addition ASAs often rely on constructing subliminal channels.
- However [BPR14] is the first to provide a formal treatment of ASAs and also provides a more general analysis.

Algorithm Substitution Attacks

- Note that ASAs are different from backdoors, as in the case of the Dual EC DRBG.
- The focus here is whether an **implementation** of the scheme offers the claimed security. The original scheme is assumed to be secure and free from backdoors.
- ASAs have been considered in the past in the works of Young and Yung, and others, under the name of Kleptography. In addition ASAs often rely on constructing subliminal channels.
- However [BPR14] is the first to provide a formal treatment of ASAs and also provides a more general analysis.

Subversions

- For a symmetric encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ its subversion is a pair $\widetilde{\Pi} = (\widetilde{\mathcal{K}}, \widetilde{\mathcal{E}})$.
- In an ASA the attacker samples a subversion key $\widetilde{K}$ and substitutes $\mathcal{E}$ with $\widetilde{\mathcal{E}}_{\widetilde{K}}$, where $\widetilde{\mathcal{E}}$ takes the same inputs as $\mathcal{E}$ together with $\widetilde{K}$.
- Since the code is assumed to be obfuscated, the subversion key $\widetilde{K}$ is inaccessible to the user.
- This gives big brother much more power to reach his goal.

Main Results From BPR14

- Propose two complementary security definitions:
  - A notion of **surveillance resilience** to prove positive results.
  - A notion of **undetectability** to prove negative results.
- The **biased ciphertext attack**, consisting of an undetectable subversion, applicable to any probabilistic scheme, which allows the attacker to recover the user's key.
- Identify a property of symmetric encryption schemes, called **unique ciphertexts**, that is sufficient to guarantee surveillance resilience.
- They show that most nonce-based schemes can be used to build schemes with unique ciphertexts.

Surveillance Resilience [BPR14]

Game $\mathrm{SURV}^{\mathscr{B}}_{\Pi,\widetilde{\Pi}}$

$b \leftarrow\!\!\$ \{0,1\}, \widetilde{K} \leftarrow\!\!\$ \widetilde{\mathcal{K}}, b' \leftarrow \mathscr{B}^{\mathrm{KEY},\mathrm{ENC}}(\widetilde{K})$
return $(b = b')$

$\mathrm{KEY}(i)$

if $K_i = \perp$ then $K_i \leftarrow\!\!\$ \mathcal{K}, \sigma_i \leftarrow \varepsilon$
return ε

$\mathrm{ENC}(M, A, i)$

if $K_i = \perp$ then return $\perp$
if $b = 1$ then $(C, \sigma_i) \leftarrow \mathcal{E}(K_i, M, A, \sigma_i)$
else $(C, \sigma_i) \leftarrow \widetilde{\mathcal{E}}(\widetilde{K}, K_i, M, A, \sigma_i, i)$
return C

$$\mathbf{Adv}^{\mathrm{srv}}_{\Pi,\widetilde{\Pi}}(\mathscr{B}) := 2 \cdot \Pr\left[\,\mathrm{SURV}^{\mathscr{B}}_{\Pi,\widetilde{\Pi}}\,\right] - 1$$

Undetectability [BPR14]

Game $\mathrm{DETECT}^{\mathscr{U}}_{\Pi,\widetilde{\Pi}}$

$b \leftarrow\$ \{0,1\}, \widetilde{K} \leftarrow\$ \widetilde{\mathcal{K}}, b' \leftarrow \mathscr{U}^{\textsc{Key},\textsc{Enc}}$
return $(b = b')$

$\textsc{Key}(i)$

if $K_i = \perp$ then $K_i \leftarrow\$ \mathcal{K}, \sigma_i \leftarrow \varepsilon$
return K_i

$\textsc{Enc}(M, A, i)$

if $K_i = \perp$ then return $\perp$
if $b = 1$ then $(C, \sigma_i) \leftarrow \mathcal{E}(K_i, M, A, \sigma_i)$
else $(C, \sigma_i) \leftarrow \widetilde{\mathcal{E}}(\widetilde{K}, K_i, M, A, \sigma_i, i)$
return C

$$\mathbf{Adv}^{\mathrm{det}}_{\Pi,\widetilde{\Pi}}(\mathscr{U}) := 2 \cdot \Pr\left[\,\mathrm{DETECT}^{\mathscr{U}}_{\Pi,\widetilde{\Pi}}\,\right] - 1$$

The Decryptability Condition

- Whithout additional restrictions it is always possible to find a subversion $\widetilde{\Pi}$ such that $\mathscr{B}$ can win the SURV game with probability one.
- Accordingly BPR require the following 'minimal' condition of undetectability that **every** subversion must satisfy.

Definition (Decryptability)

A subversion $\widetilde{\Pi} = (\widetilde{\mathcal{K}}, \widetilde{\mathcal{E}})$ is said to satisfy decryptability with respect to the scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ if the encryption scheme $(\widetilde{\mathcal{K}} \times \mathcal{K}, \widetilde{\mathcal{E}}, \mathcal{D}')$ is perfectly correct, where $\mathcal{D}'((\widetilde{K}, K), C, A, \varrho) = \mathcal{D}(K, C, A, \varrho)$.

Analysis of The BPR Model

- The first thing to note is that:

$$\text{Undetectability} \not\Longrightarrow \text{Decryptability}$$

- Undetectability allows $\mathscr{U}$ a small success probability but the same is not true for Decryptability.
- This is overly restrictive on $\mathscr{B}$. There is no reason why $\mathscr{B}$ would only consider subversions that have zero probability of being detected.

Analysis of The BPR Model

- The first thing to note is that:

$$\text{Undetectability} \not\Longrightarrow \text{Decryptability}$$

- Undetectability allows $\mathscr{U}$ a small success probability but the same is not true for Decryptability.

- This is overly restrictive on $\mathscr{B}$. There is no reason why $\mathscr{B}$ would only consider subversions that have zero probability of being detected.

- So why not relax the decryptability condition by allowing a small probability of error?

Analysis of The BPR Model

- The first thing to note is that:

$$\text{Undetectability} \not\Rightarrow \text{Decryptability}$$

- Undetectability allows $\mathscr{U}$ a small success probability but the same is not true for Decryptability.
- This is overly restrictive on $\mathscr{B}$. There is no reason why $\mathscr{B}$ would only consider subversions that have zero probability of being detected.
- So why not relax the decryptability condition by allowing a small probability of error?

Input-Triggered Subversions

- This slight relaxation renders the notion of surveillance resiliance **unsatisfiable**!

- For **any** scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ there exists a subversion $\widetilde{\Pi} = (\widetilde{\mathcal{K}}, \widetilde{\mathcal{E}})$ defined by:

$$\boxed{\begin{array}{l} \underline{\text{Algorithm } \widetilde{\mathcal{E}}_{\widetilde{K}}(K, M, A, \sigma, i)} \\ \quad C \leftarrow \mathcal{E}_K(M, A, \sigma) \\ \quad \text{if } \mathbf{R}(\widetilde{K}, K, M, A, \sigma, i) = \text{true} \\ \qquad \text{then return } (C \parallel K, \sigma) \\ \quad \text{else return } (C, \sigma) \end{array}}$$

- This subversion is decryptable (with negligible error) and is in fact undetectable, but there exists an adversary $\mathscr{B}$ such that $\mathbf{Adv}^{\mathrm{srv}}_{\Pi,\widetilde{\Pi}}(\mathscr{B}) = 1$.

Input-Triggered Subversions

- This slight relaxation renders the notion of surveillance resiliance **unsatisfiable**!

- For **any** scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ there exists a subversion $\widetilde{\Pi} = (\widetilde{\mathcal{K}}, \widetilde{\mathcal{E}})$ defined by:

$$\boxed{\begin{array}{l} \underline{\text{Algorithm } \widetilde{\mathcal{E}}_{\widetilde{K}}(K, M, A, \sigma, i)} \\ \quad C \leftarrow \mathcal{E}_K(M, A, \sigma) \\ \quad \text{if } \mathbf{R}(\widetilde{K}, K, M, A, \sigma, i) = \text{true} \\ \qquad \text{then return } (C \parallel K, \sigma) \\ \quad \text{else return } (C, \sigma) \end{array}}$$

- This subversion is decryptable (with negligible error) and is in fact undetectable, but there exists an adversary $\mathscr{B}$ such that $\mathbf{Adv}^{\mathrm{srv}}_{\Pi,\widetilde{\Pi}}(\mathscr{B}) = 1$.

The Proposed Surveillance Resilience Definition

- Perfect decryptability implicitly excludes this important class of subversions thereby imposing artificial limitations on big brother.
- We propose a security definition that builds on ideas from [BPR14] but disposes of the the decryptability requirement altogether.
- A one-time detection strategy does not suffice, instead it seems that a continuous detection strategy is necessary.
- In addtition our security definition provides quantifiably stronger guarantees of detecting an ASA.

The Proposed Surveillance Resilience Definition

Game $\overline{\mathrm{SURV}}^{\mathscr{B}}_{\Pi,\widetilde{\Pi}}$

$$
\begin{array}{l}
b \leftarrow_{\$} \{0,1\}, \widetilde{K} \leftarrow_{\$} \widetilde{\mathcal{K}} \\
b' \leftarrow \mathscr{B}^{\mathrm{KEY},\mathrm{ENC}}(\widetilde{K}) \\
\text{return } (b = b')
\end{array}
$$

$\underline{\mathrm{KEY}(i)}$ // called at most once

$$
\begin{array}{l}
\text{if } K_i = \perp \text{ then } K_i \leftarrow_{\$} \mathcal{K}, \sigma_i \leftarrow \varepsilon \\
\text{return } \varepsilon
\end{array}
$$

$\underline{\mathrm{ENC}(M, A, i)}$

$$
\begin{array}{l}
\text{if } K_i = \perp \text{ then return } \perp \\
\text{if } b = 1 \text{ then } (C, \sigma_i) \leftarrow \mathcal{E}(K_i, M, A, \sigma_i) \\
\text{else } (C, \sigma_i) \leftarrow \widetilde{\mathcal{E}}(\widetilde{K}, K_i, M, A, \sigma_i, i) \\
\text{return } C
\end{array}
$$

This is the SURV game from [BPR14] formulated in the single-user setting.

The Proposed Surveillance Resilience Definition

Game $\overline{\mathrm{DETECT}}^{\mathscr{B},\mathscr{U}}_{\Pi,\widetilde{\Pi}}$

$b \leftarrow_\$ \{0,1\}, \widetilde{K} \leftarrow_\$ \widetilde{\mathcal{K}}$
$b' \leftarrow \mathscr{B}^{\mathrm{Key},\mathrm{Enc}}(\widetilde{K}), b'' \leftarrow \mathscr{U}(T)$
return $(b = b'')$

$\underline{\mathrm{Key}(i)}$ // called at most once

if $K_i = \perp$ then $K_i \leftarrow_\$ \mathcal{K}, \sigma_i \leftarrow \varepsilon$
$T \leftarrow (K_i, i)$
return ε

$\underline{\mathrm{Enc}(M, A, i)}$

if $K_i = \perp$ then return $\perp$
if $b = 1$ then $(C, \sigma_i) \leftarrow \mathcal{E}(K_i, M, A, \sigma_i)$
else $(C, \sigma_i) \leftarrow \widetilde{\mathcal{E}}(\widetilde{K}, K_i, M, A, \sigma_i, i)$
$T \leftarrow T \parallel (M, A, C)$
return C

The Proposed Surveillance Resilience Definition

The advantages corresponding to each game are defined as:

$$\mathbf{Adv}^{\overline{\mathrm{srv}}}_{\Pi,\widetilde{\Pi}}(\mathscr{B}) := 2 \cdot \Pr\left[\,\overline{\mathrm{SURV}}^{\mathscr{B}}_{\Pi,\widetilde{\Pi}}\,\right] - 1\,,$$

and

$$\mathbf{Adv}^{\overline{\mathrm{det}}}_{\Pi,\widetilde{\Pi}}(\mathscr{B},\mathscr{U}) := 2 \cdot \Pr\left[\,\overline{\mathrm{DETECT}}^{\mathscr{B},\mathscr{U}}_{\Pi,\widetilde{\Pi}}\,\right] - 1\,.$$

Definition

The pair $(\Pi, \mathscr{U})$ is said to be surveillance resilient if for all subversions $\widetilde{\Pi}$ and all adversaries $\mathscr{B}$ it hold that $\mathbf{Adv}^{\overline{\mathrm{det}}}_{\Pi,\widetilde{\Pi}}(\mathscr{B},\mathscr{U}) \geq \mathbf{Adv}^{\overline{\mathrm{srv}}}_{\Pi,\widetilde{\Pi}}(\mathscr{B})$.

Notes on The Proposed Definition

- BPR's DETECT game was meant for negative results, while our $\overline{\mathrm{DETECT}}$ game replaces the decryptability condition.

- Contrary to the DETECT game, in $\overline{\mathrm{DETECT}}$ the detection test $\mathscr{U}$ is universal and can be run by a single user.

- In the proposed security definition, $\mathscr{U}$ is guaranteed to **always** detect a subversion. In the BPR security definition we were only guranteed a **non-zero** success probability of detecting a subversion.

Security of Unique Ciphertext Schemes

- An encryption scheme is said to have unique ciphertexts if for all message sequences and all keys there exists exactly one ciphertext sequence that decrypts to this message sequence.
- Schemes with unique ciphertexts must be deterministic, but not all deterministic schemes have unique ciphertexts.

Theorem

Let $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be a symmetric encryption scheme with unique ciphertexts. Then for every Π there exists a detection test $\mathscr{U}$ such that for all subversions $\widetilde{\Pi}$ and all adversaries $\mathscr{B}$ the following holds

$$\mathbf{Adv}^{\overline{\mathrm{det}}}_{\Pi,\widetilde{\Pi}}(\mathscr{B}, \mathscr{U}) \geq \mathbf{Adv}^{\overline{\mathrm{srv}}}_{\Pi,\widetilde{\Pi}}(\mathscr{B}) .$$

Limitations of The Analysis

- The analysis from [BPR14] and by extensions ours as well, only considers leakage of information through ciphertexts.
- Thus other types of ASAs may be possible based on side information such as timing, power analysis, electromagnetic radiation, etc. These settings are **not** covered by our analysis.
- Arguably, such ASAs may be harder to mount as they need to be targeted attacks.

Summary

- We build on the work of [BPR14] to converge to a better security model for ASAs and re-established their positive results.

- However our analysis highlights that detecting ASAs is more challenging than what was indicated by [BPR14].