Bias in the TRNG of the Mifare DesFire EV1, a CC EAL 4+ RFID Card, and what went wrong

Darren Hurley-Smith
Julio Hernandez-Castro

19/1/2017
Outline

Introduction
Motivation
Methodology
Results
Analysis
Conclusion
Future Work
Acknowledgements
Introduction

- We are researching potential vulnerabilities in the Mifare DESFire EV1, and others: EV2, FeliCa, LEGIC.
- The DESFire EV1 is a Common Criteria EAL4+ certified smart card used in transport, Univ. student/staff ID, fare payment and other micro-payments, around the world [1]
- It is also used in loyalty schemes and access control applications
- Transport for London (TfL) issued approximately 8 million DESFire EV1-based cards in the 2015/2016 period [2]
Motivation: Background

- A weak PRNG and cryptographic algorithm (CRYPTO-1) contributed to dismantling the Mifare Classic [3–6]
- Let’s do it again!
- Also, they continue to use security by obscurity, not publicising anything about the TRNG design they use
- Let’s prove again that’s not a good idea!
Motivation: Rationale

- As a cash-value bearing card, the DESFire EV1 has a monetary value to criminals
- The Mifare DESFire EV1 has been successfully emulated [7], and its power characteristics have been analysed in depth [8]
- This card has proven resilient to side-channel attacks (SCA), by implementing hardware countermeasures [9]
- We believed that an in depth evaluation of the DESFire EV1’s TRNG could be interesting
Methodology: Data collection

- 64 MB of data was retrieved from three DESFire EV1 and one EV2 using an ACR122U reader
  - Data collection took an average of 12 days per card to gather 4 million AES-128 encrypted values
  - Each nonce (16-bytes long) was extracted from a different authentication session
- The data was acquired from the protocol used to secure card PICC and Application read/write functions
- The values were decrypted using a default AES-128 key (initialised to zero) before analysis
Methodology: Lab Set-up

Figure 1: Experimental set-up used to collect TRNG data

- Laptop with reader and DESFire EV1 cards
- ACR122U reader with two of three Mifare DESFire EV1 cards

- Toshiba Laptop Specification: i7 processor, 8GB RAM
- Reader: ACR122U (CCID), Scripts: Python 2.7 and Bash
Methodology: Randomness Tests

The collected data was subjected to three randomness test batteries:

- The NIST Statistical Test Suite v2.1.2
- Dieharder
- ENT
### Results: NIST STS 2.1.2

**Table 1:** Mifare DESFire EV1 ENT results for 64MB of TRNG output

<table>
<thead>
<tr>
<th></th>
<th>Passed Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card 1</td>
<td>198/200</td>
</tr>
<tr>
<td>Card 2</td>
<td>200/200</td>
</tr>
<tr>
<td>Card 3</td>
<td>197/200</td>
</tr>
</tbody>
</table>

- All cards passed the NIST STS 2.1.2 battery within acceptable parameters (greater than 193)
- Cards 1 and 3 perform poorly in one of the non-overlapping template tests, but passed all other tests
  - This seems to be statistically insignificant
- Card 2 passes all tests with no weak results
### Table 2: Diehard results for 64MB of TRNG Output

<table>
<thead>
<tr>
<th>Diehard Test</th>
<th>EV1 Card 1</th>
<th>EV1 Card 2</th>
<th>EV1 Card 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t-samples</td>
<td>p-values</td>
<td>p-values</td>
</tr>
<tr>
<td>Birthday Spacings</td>
<td>default</td>
<td>0.18194520</td>
<td>0.61105583</td>
</tr>
<tr>
<td>Overlapping Permutations</td>
<td>125,000</td>
<td>0.38044164</td>
<td>0.58693289</td>
</tr>
<tr>
<td>6x8 Binary Rank</td>
<td>25,000</td>
<td>0.31311490</td>
<td>0.32387215</td>
</tr>
<tr>
<td>Bitstream</td>
<td>default</td>
<td>0.97724174</td>
<td>0.18743536</td>
</tr>
<tr>
<td>Count the 1’s (stream)</td>
<td>default</td>
<td>0.17108396</td>
<td>0.74984724</td>
</tr>
<tr>
<td>Parking Lot</td>
<td>default</td>
<td>0.86481241</td>
<td>0.92578024</td>
</tr>
<tr>
<td>Count the 1’s (byte)</td>
<td>default</td>
<td>0.18078043</td>
<td>0.24200626</td>
</tr>
<tr>
<td>Minimum Distance (2d sphere)</td>
<td>default</td>
<td>0.76328000</td>
<td>0.95091635</td>
</tr>
<tr>
<td>3d sphere (minimum distance)</td>
<td>default</td>
<td>0.23871272</td>
<td>0.20826216</td>
</tr>
<tr>
<td>Squeeze</td>
<td>default</td>
<td>0.62598919</td>
<td>0.08843989</td>
</tr>
<tr>
<td>Runs</td>
<td>default</td>
<td>0.63756836</td>
<td>0.80941394</td>
</tr>
<tr>
<td>Craps</td>
<td>20,000</td>
<td>0.54077256</td>
<td>0.92769962</td>
</tr>
</tbody>
</table>
Table 3: Mifare DESFire EV1 ENT results for 64MB of TRNG output

<table>
<thead>
<tr>
<th></th>
<th>EV1 Card 1</th>
<th>EV1 Card 2</th>
<th>EV1 Card 3</th>
<th>Optimal/Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>7.9999969</td>
<td>7.999989</td>
<td>7.999972</td>
<td>8</td>
</tr>
<tr>
<td>Optimal Compress.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chi-square</td>
<td>2709.10</td>
<td>973.07</td>
<td>2470.32</td>
<td>255</td>
</tr>
<tr>
<td>Arith. Mean</td>
<td>127.492921</td>
<td>127.500582</td>
<td>127.5006</td>
<td>127.5</td>
</tr>
<tr>
<td>Monte Carlo (\pi) est.</td>
<td>3.14167</td>
<td>3.142019</td>
<td>3.141909196</td>
<td>3.14159</td>
</tr>
<tr>
<td>S. Correlation</td>
<td>0.000008</td>
<td>0.000045</td>
<td>0.000093</td>
<td>0.0</td>
</tr>
</tbody>
</table>

- Both cards demonstrate very poor performance on the chi-square test
- This indicates that there is a strong bias in the distribution of byte values throughout both data samples
### Table 4: Mifare DESFire EV1 ENT results for 1MB of TRNG output

<table>
<thead>
<tr>
<th>ENT</th>
<th>EV1 Card 1</th>
<th>EV1 Card 2</th>
<th>EV1 Card 3</th>
<th>Optimal/Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>7.999780</td>
<td>7.999820</td>
<td>7.999786</td>
<td>8</td>
</tr>
<tr>
<td>Optimal Compress.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>chi-square</td>
<td>305.47 (1.65%)</td>
<td>249 (59.41%)</td>
<td>297.03 (3.62%)</td>
<td>255</td>
</tr>
<tr>
<td>Arith. Mean</td>
<td>127.6015</td>
<td>127.5626</td>
<td>127.5082</td>
<td>127.50</td>
</tr>
<tr>
<td>Monte Carlo $\pi$ est.</td>
<td>3.13620558</td>
<td>3.140892564</td>
<td>3.140388562</td>
<td>3.14159</td>
</tr>
<tr>
<td>S. Correlation</td>
<td>-0.000068</td>
<td>0.001339</td>
<td>-0.001751</td>
<td>0.0</td>
</tr>
</tbody>
</table>

- All cards perform better on this test with smaller samples and have p-values greater than 0.01 for the chi-square test
- This indicates that the bias may be missed if tests are not performed on a large enough sample of TRNG data (greater than 1 MB in the case of ENT)
### Table 5: Mifare DESFire EV1 ENT results for 5KB of TRNG output

<table>
<thead>
<tr>
<th></th>
<th>EV1 Card 1</th>
<th>EV1 Card 2</th>
<th>EV1 Card 3</th>
<th>Optimal/Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>7.999635</td>
<td>7.999640</td>
<td>7.999641</td>
<td>8</td>
</tr>
<tr>
<td>Optimal Compress.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>chi-square</td>
<td>253.55 (51.3%)</td>
<td>249.26 (58.96%)</td>
<td>249.03 (59.36%)</td>
<td>255</td>
</tr>
<tr>
<td>Arith. Mean</td>
<td>127.6015</td>
<td>127.6324</td>
<td>127.4534</td>
<td>127.5</td>
</tr>
<tr>
<td>Monte Carlo $\pi$ est.</td>
<td>3.13744549</td>
<td>3.145452582</td>
<td>3.140388562</td>
<td>3.14159</td>
</tr>
<tr>
<td>S. Correlation</td>
<td>-0.000579</td>
<td>0.001990</td>
<td>-0.001204</td>
<td>0.0</td>
</tr>
</tbody>
</table>

- At 5KB, all tests pass the chi-square test with p-values greater than 0.01 and within acceptable bounds of the expected value.
- Deviations start to suggest themselves (but not totally clear) in the chi-square results for samples larger than 7.5KB.
Analysis: Bias

(a) Mifare DESFire EV1 mean bias

(b) Random Data

Figure 2: Bias of two 64MB datasets

- Only (a) shows a clear non-random trend
- Repetitive pattern, clear cycles, almost no values close to zero...
Analysis: Fourier Analysis of the Bias

Figure 3: Fourier series for the biases from three 64MB TRNG samples

- All cards demonstrate a regular period of 32 biased values
- Exactly half of the possible byte values occur more frequently
Initial Findings

- Previously, evaluators have used AIS-31 (CC)
- and researchers employed NIST, Dieharder, chi-square, and other uniformity tests\textsuperscript{1}
- but these are not always sufficient to find the bias in the EV1 TRNG
- Chi-square tests performed in the literature focus on bit-level analysis, but the bias is only apparent at the byte level in this TRNG

\textsuperscript{1}Private communication.
Further Findings

- We experimentally found that bits 4 and 5 of every byte seemed to frequently take the same value.
- We built a model with different probabilities of this occurring, and found heuristically that the one with a probability of
  \[ p(b_4 = b_5) = \frac{1}{2} - \epsilon \text{ with } \epsilon \sim 0.0004325 \]
  performed the best in approximating the bias of the three EV1 cards.
- After these analysis, we can explain up to \( R^2 \approx 0.8121 \) (average of 77.96\%) of the observed bias for EV1.
Further Findings

Table 6: Coefficient of determination ($R^2$)

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card 1</td>
<td>0.7981</td>
<td>0.7836</td>
</tr>
<tr>
<td>Card 2</td>
<td>0.7288</td>
<td>0.7094</td>
</tr>
<tr>
<td>Card 3</td>
<td>0.8121</td>
<td>0.7987</td>
</tr>
<tr>
<td>urandom</td>
<td>0.0479</td>
<td>-0.0201</td>
</tr>
</tbody>
</table>
Analysis: New Randomness Test: The Bitmask Test

Basic idea is to apply linear cryptanalysis to the raw random data $r$

- We look for $\max_{m \in \{1..255\}} |\sum m \cdot r - r/2|$
- Maximum bias for $m=24$ (00011000), consistently
Analysis: Bit-Mask Test Results

Figure 5: Graphs showing the bias for all masks applied over 64MB samples

- This trend holds across all three EV1 64MB data sets
Analysis: Byte-level Analysis

Figure 6: Visualisation of the bias of bit-adjacency for all bytes (64M)

- (a) shows the same visualisation as (b) over random data
- (b) shows that bits 4 and 5 have a predisposition to sharing values
Conclusion

- We have conducted a study of the Mifare DESFire EV1’s ‘true’ random number generator
  - Clear & consistent biases have been found in the data
- We have responsibly disclosed our findings to NXP
  - They have responded, confirming our findings
  - They have a team ”looking into the root cause”
- No practical attacks have been identified at this point
  - But hopefully future attacks will build on these findings
- We have observed that some of the best known tests do not detect this flaw, PRNG/TRNG evaluation is tricky!
- In particular, we have shown some of the limitations of the current CC certification process and would recommend major changes on it
  - And we have worked with average case, not worst case scenarios!
Future Work

- Continue to collect data from more Mifare DESFire EV1 (check manufacturing issues) and EV2 cards
- Expand the tested cards to include other Common Criteria EAL 4/5+ RFID smart cards:
  - Felica
  - Legic
- Testing collected data with other test batteries (e.g. TestU01)
- Testing under variable environmental conditions (extreme temperatures, etc.)
- Developing hardware model that explains the observed bias (help wanted!)
- Start an ambitious project to analyse current randomness test suites and study their test’s independence and sensibility to come up with a new one to which only new independent test will be added.
Acknowledgements

▶ This work was funded by InnovateUK as part of the authenticatedSelf project, under reference number 102050.

▶ We would like to thank ECOST - Cryptacus Action for their valuable and insightful discussion of this work.

▶ We would also like to thank NXP Semiconductors Ltd. for their timely and professional communication following the responsible disclosure of our findings.
References


Questions?