Feasible Privacy for Lightweight RFID Systems

David Evans
work with Karsten Nohl
University of Virginia

SPAR Seminar
Johns Hopkins University
17 October 2007

UPC Bar Code
EPC Gen 2 RFID

Identities
8-12 digits
(product identity)

Reading
Optical Scanner

Tag Cost
Ink, Paper
($0.00001?)

64-128 bits
(item identity)

Wireless Reader
Circuit, Antenna
($0.05)

www.cs.virginia.edu/evans/talks/spar07

“More-Efficient Mugging”

“Just in case you want to know, she’s got 700 Euro and 8 World Cup tickets…”

From Ari Juels USENIX Security 2004 talk
http://www.usenix.org/events/sec04/tech/slides/juels.htm

Realistic Threats

Profiling/Tracking
Corporate Espionage

Solutions for Paranoids

RFID Shield ($9.99)
Tin Foil

www.cs.virginia.edu/evans/talks/spar07
www.cs.virginia.edu/evans/talks/spar07
www.cs.virginia.edu/evans/talks/spar07
www.cs.virginia.edu/evans/talks/spar07
Basic Hash Protocol

\[ R, H_K(R) \]

- **Key**: \( K \)
- **Nonce**: \( R \)

\[ N \text{ tags} \]
\[ N \text{ hashes} \]

Tree-Hash Protocol

- **David Molnar and David Wagner. CCS 2004.**

\[ \begin{align*}
  k_{1,0} & \quad k_{1,1} \\
  k_{2,0} & \quad k_{2,1} & \quad k_{2,2} & \quad k_{2,3} \\
  T_1 & \quad T_2 & \quad T_3 & \quad T_4
\end{align*} \]

Basic Hash Protocol at each level

Reader computes up to \( b \log_b N \) hashes

Analysis of Tree Protocol

- Attacker wants traces of individuals
- Attacker can easily acquire tags and break their secrets
- Assume no side channels: only protocol layer leaks
- Assume a good cryptographic hash function
  - Second part of the talk is about whether this is reasonable

Shared Secrets

- Each broken tag enables attacker to group intercepted tags using shared secrets

\[ \begin{aligned}
  & \quad \text{Group of } n \text{ tags} \\
  & \quad \text{Stolen secret} \\
  & \quad \text{Broken tag}
\end{aligned} \]

Information theoretic measure of privacy based on the group size

Groups and Leakage

- \( b=1 \), upper bound
- \( b=1 \), real value
- \( b=10 \), upper bound
- \( b=10 \), real value

\[ \text{Information Leakage vs } x \text{ [bits]} \]
Cost Trade-Off

Low-Leakage Tree Protocol
- Avoid small groups
- Leads to two-level tree for systems with billions of tags
- Opposite of originally proposed binary tree

Reader computes up to √N hashes

1B tags ~ 31K hashes

Tree-Hash Protocol Feasible?
- Random Number
  An RN16 drawn from a Tag's RNG... shall not be predictable with a probability greater than 0.025% if the outcomes of prior draws from the RNG, performed under identical conditions, are known. 

  [EPC Class 1 Gen 2 Standard]

  ~12 good bits out of 16

  • Hash function (rest of this talk...)

Implementing Hash Functions

| 10k gates | 4k gates | 2k gates |
| SHA-256 | AES | RFID tag |

Power consumption scales with gates, not "Moore's Law". Reading distance is inverse square-cube of power needed.

Cryptographic Hash Functions
- Pre-image resistance
  - Given H(x) it is hard to find x

  Not sufficient for privacy!

- Second pre-image resistance
  - Given y hard to find x such that H(x) = y

  Not necessary for privacy!

- Collision resistance
  - Hard to find x and y such that H(x) = H(y)

  Hardest

Non-Private Strong Hash

H(x) = G(x) || x

where G is a strong, cryptographic hash function
Private Hash Function

\[ H(R, K) \]

- **R**: (non-secret) nonce
- **K**: key shared with reader
- **Correctness**: given \( H(R, K) \), \( R \), and key set easy to find \( K \)
- **Privacy**: given a set of \( \langle H(R, K), R \rangle \) tuples it is hard to identify two tuples generated by the same key (without knowing key set)

Security Argument: 2-split

\[ X = D(R_1, K_1) \oplus D(R_2, K_2) \]

- \( n = \) total key bits, divided between \( K_1 \) and \( K_2 \)
- Precompute one side
- Try values to find match
- Brute force attack: \( \Theta(2^n) \)
- Meet-in-middle attack: \( \Theta(2^{n/2}) \) space, time

Abstract Design

\[ H(R, K) = D(R_1, K_1) \oplus \cdots \oplus D(R_n, K_n) \]

where

- \( R = R_1 \parallel \cdots \parallel R_n \)
- \( K = K_1 \parallel \cdots \parallel K_n \)
- independent nonce/key shares

\( D(r, k) \) is a "Distortion Function" with:
- Even output distribution
- Black-box function with poly-time reversing oracle that outputs set of \( k \)'s producing a given output

Concrete Abstract Design

- **3-split**: \( D(R_1, K_1) \oplus D(R_2, K_2) \oplus D(R_3, K_3) \)
- **Implementable Distortion Function**
  - Even output distribution
  - Black-box function with reverse oracle
  - Implementable function such that attacker cannot find correlations: no easier way to break than by finding the intermediate values

CRC

- Cyclic Redundancy Check
- Already required on EPC tags
- Designed [Peterson, 1961] to be easy to implement in hardware, error-checking code (no crypto goals)

\[ CRC_g(X) = \text{remainder of polynomial division } X \text{ by } g \text{ in } GF(2) \]
Attempted CRC Privacy Protocol


Tag $\rightarrow$ Reader : $R, CRC(ID \parallel R) \oplus K$

Fixed (standard) generator polynomial $K$ changes when updated by legitimate reader

CRC Properties

$\exists \mathbf{x} \in X_g = \text{set of values evenly divided by } g$

$CRC_g(x) = (x) \oplus \mathbf{x}$

$A_1 = CRC_g(K_1 \parallel R_1) = (K_1 \parallel R_1) \oplus x_1$

$A_2 = CRC_g(K_2 \parallel R_2) = (K_2 \parallel R_2) \oplus x_2$

$x_1, x_2 \in X_g$

Private Hash Function

$D(R_1, K_1) \oplus D(R_2, K_2) \oplus D(R_3, K_3)$

$D(r, k) = CRC_{k_2 \oplus k_1}(k_b)$

Distortion Function Required Properties

- Confusion: changing one input bit flips each output bit with probability $\frac{1}{2}$
- Diffusion: changing one generator bit flips each output bit with probability $\frac{1}{2}$
- Even distribution: all outputs are equally likely
- Complexity: hard to correlate better than black box

Proof Sketches

- Confusion and Diffusion
  - Requires: Hamming weight of generator is $\frac{1}{2}$ length
  - Proof: Follow bit probabilities through CRC
- Even Distribution
  - CRC provides even outputs over $[0,g-1]$
    - But not over all output bits
  - To get approximately even distribution: use only $i$ low-order output bits, and combine outputs (second is reversed)

Attacks on Complexity

- Most known crypto attacks don't apply
- No chosen plaintext makes differential/linear cryptanalysis infeasible
  - Recall assumption: if attacker has physical access they can just extract key
- Statistical Attacks (e.g., distinguishing attacks) fail because output is evenly distributed and no state is kept
Algebraic Attacks

• Create and solve system of equations for bits
• Successfully break many stream ciphers (and some block ciphers)
• Even partial knowledge of single key bit can weaken privacy
• No general defense exists

3-bit CRC Complexity

After 5 shifts:
\[
H_{1,5} = k_1 g_1 + k_2 g_2 + k_3 \\
H_{2,5} = k_1 g_1 g_2 + k_2 g_2 + k_4 \\
H_{3,5} = k_1 g_1 g_2 g_3 + k_2 g_3 + k_5
\]

Distortion Complexity

Implementation

• CRC with fixed generator already included on tags (required by EPC Class 1 standard)
• Extend to support variable generator: 130 gates (355 total GE)
  – Smallest known AES: 3400 gates
• Reader: simple implementation can do 10x (AES) - 40x (SHA-256) as many hashes as alternatives

Summary

• Cheap RFIDs are expensive bar codes, not little computers
  – Can’t do division, encryption, cryptographic hashing, etc.
• Privacy does not require strong crypto hashing
  – Very simple, inexpensive functions may be sufficient for privacy
“We cannot even answer the most basic questions because we don’t know enough about you. That is the most important aspect of Google’s expansion.”

Eric Schmidt (Google’s CEO)
May 2007

For more information:
evans@cs.virginia.edu
http://www.cs.virginia.edu/evans

Karsten Nohl and David Evans. Private Hash Functions: Lightweight Protection for RFID Systems. (In submission, request by email)

Karsten Nohl and David Evans. Optimizing Secret Trees for Privacy. (In submission, request by email)