When Things Go Wrong
Authenticating Nuclear Warheads with High Confidence

Moritz Kütt, Sebastien Philippe, Boaz Barak, Alexander Glaser, and Robert J. Goldston
Princeton University, Technische Universität Darmstadt
Microsoft Research New England, Princeton Plasma Physics Laboratory
55th Annual Meeting, Institute of Nuclear Materials Management
Georgia, Atlanta, July 2014
Inspection Systems for Nuclear Warhead Verification Have Been Under Development Since the 1990s

Attribute Approach
Confirming selected characteristics of an object in classified form (for example, the presence/mass of plutonium)

Template Approach
Comparing the radiation signature from the inspected item with a reference item (“golden warhead”) of the same type

Information Barrier
Technologies and procedures that prevent the release of sensitive nuclear information (generally needed for both approaches)
What To Do When “Red Lights” Come On?

Fissile Material Transparency Technology Demonstration (FMTTD), Los Alamos, August 2000

What To Do When “Red Lights” Come On?

Fissile Material Transparency Technology Demonstration (FMTTD), Los Alamos, August 2000


---

Secure Mode

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weapon component</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Large oxide sample on its side</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
Methodology
(applicable to both attribute and template approaches)
Assume basic test (single measurement) is “not strong enough”
(i.e., too many valid items can fail, too many invalid items can pass)

We repeatedly apply basic test until we declare item “good” or “bad”
Terminology and Definitions

Assume basic test (single measurement) is “not strong enough”
(i.e., too many valid items can fail, too many invalid items can pass)

We repeatedly apply basic test until we declare item “good” or “bad”

4 Key Parameters

\[ \alpha : \text{probability that valid item passes basic test (e.g. 0.95)} \]
\[ \beta : \text{probability that invalid item passes basic test (0.05–0.10, but could be higher)} \]
\[ \alpha^* : \text{probability that valid item is declared “good” (host wants high, e.g. 0.99)} \]
\[ \beta^* : \text{probability that invalid item is declared “good” (inspector wants low, e.g. 0.01)} \]
Re-Testing Protocol

Assume we have observed a sequence of \( n \) basic tests \( X_1, X_2, \ldots, X_n \) (Bernoulli trials or “biased coin flips”) that includes \( P \) “passes” and \( F \) “fails”

We define two hypotheses that we will test

A. Item is valid and passes each basic test with probability \( \alpha \)
B. Item is invalid and fails each basic test with probability \( (1 - \beta) \)

For any given \( n \), we can check the following inequalities to accept one of the hypotheses:

\[
\text{A. } \sum_{i=F}^{n} \binom{n}{i} (1 - \alpha)^i \alpha^{n-i} \leq 1 - \alpha^* \\
\text{B. } \sum_{i=P}^{n} \binom{n}{i} \beta^i (1 - \beta)^{n-i} \leq \beta^*
\]

(both can’t be true at the same time if \( \beta < \alpha \))

We want to find the smallest \( n_{opt} \) such that one of the inequalities is true for any combination of passed tests \( (k_p) \) and failed tests \( (k_f) \) with \( k_p + k_f = n_{opt} \)
Scorecards
Making Scorecards

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)

In this particular realization, after \(n = 3\) tests, the item seems to meet our criteria.
Making Scorecards

Default parameter set: \( (\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01) \)
Making Scorecards

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)

For \(n = 3\) tests, there remains a possibility for an “inconclusive” outcome; method requires further testing (i.e., higher \(n\))
Making Scorecards

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)

For \(n = 3\) tests, there remains a possibility for an “inconclusive” outcome; method requires further testing (i.e., higher \(n\))
Making Scorecards

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)

For \(n = 3\) tests, there remains a possibility for an “inconclusive” outcome; method requires further testing (i.e., higher \(n\))
Making Scorecards

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)

For \(n = 3\) tests, there remains a possibility for an “inconclusive” outcome; method requires further testing (i.e., higher \(n\))

For this particular choice of parameters, \(n = 2 + 3 + 1 = 6\) tests will always lead to a “good” or “bad” position.
“Pick Your Values ... Make Your Scorecard”

Stopping criterion (allowed fraction of missed invalids)

- $\beta^* = 0.001$
- $\beta^* = 0.010$
- $\beta^* = 0.100$

Similarity of invalid items (fraction of invalids passing basic test)

- $\beta = 0.01$
- $\beta = 0.05$

Number of passed basic tests

Number of failed basic tests
Simulated Results
Honest Host

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)
Honest Host

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)

<table>
<thead>
<tr>
<th>Untested item</th>
<th>Item looking “good”</th>
<th>Inconclusive</th>
<th>Item looking “bad”</th>
<th>Item declared “bad”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Honest Host

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)
Honest Host

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)
Honest Host

Default parameter set: $(\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)$
Dishonest Host

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)
**Dishonest Host**

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)
# Lookup Table for Inspector

(who has to assume dishonest host and worries about invalids declared “good”)

<table>
<thead>
<tr>
<th>Probability that invalid item passes basic test ($\beta$)</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.40</th>
<th>0.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of tests per item (99% detection probability for invalid items)</td>
<td>2.1</td>
<td>3.2</td>
<td>4.2</td>
<td>6.3</td>
<td>8.4</td>
<td>13.7</td>
<td>17.9</td>
</tr>
</tbody>
</table>

Similar lookup table(s) can be constructed for the host who worries about valid items being declared “bad”
Zero-knowledge Template Approach

(Glaser, Barak, Goldston, Nature, 510, June 2014)
Zero-knowledge Template Approach
(with preloaded non-electronic detectors, simplified)

1. Preloading detector pairs (host only)

2. Inspecting item together with reference item
   (inspector chooses which preloaded detectors are used with which item)

In the basic one-on-one approach, reference item needs to be present for every inspection
(Can we authenticate N items without comparing them to the template each and every time?)
Zero-knowledge Template Approach
(with preloaded non-electronic detectors, simplified)

1. Preloading detector pairs (host only)

2. Inspecting item together with reference item
   (inspector chooses which preloaded detectors are used with which item)

In the basic one-on-one approach, reference item needs to be present for every inspection
(Can we authenticate N items without comparing them to the template each and every time?)
Zero-Knowledge Template Approach

Default parameter set: $(\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)$

The template $T$ is one of the inspected items and is tested like all others.
Zero-Knowledge Template Approach

Default parameter set: \((\alpha = 0.95, \beta = 0.05, \alpha^* = 0.999, \beta^* = 0.01)\)

The template \(T\) is one of the inspected items and is tested like all others.

Default parameter set:

1. \(\alpha = 0.95\)
2. \(\beta = 0.05\)
3. \(\alpha^* = 0.999\)
4. \(\beta^* = 0.01\)
Conclusion
Summary and Conclusion

**Proposed Testing Protocol**

Protocol can be used to control the probabilities of inspection failures for successful nuclear warhead authentication

Protocol allows the host and the inspector to adjust testing parameters such that both achieve their desired performance requirements with a minimum inspection effort

Protocol can be used for both attribute-type and template-type measurements

**Findings**

Critical parameter is the fraction of invalid items passing a basic test (single measurement)

This parameter depends on diversion scenario, cannot be “defined” independently up to 20–30% may be acceptable; above this range, number of repetitions increases sharply

Future work needs to determine tradeoffs between repeated tests and longer individual inspections