National Institute of Standards and Technology
U.S. Department of Commerce

INFORMATION TECHNOLOGY LABORATORY

The Testing of Entropy Sources for Cryptography

Sarah Scheffler

Computer Security Division
Security Testing, Validation, & Management Group
Dr. Allen Roginsky

7 August, 2014

Outline

1. Background, definitions, and standards
2. Random-looking numbers with zero entropy
3. The collision test

Cryptography is essential

- Everyone needs cryptography: personal, business, national, international
- Most cryptography heavily relies on random number generators (RNG)
- These RNGs get non-deterministic bits from an *entropy source*
- NIST has a draft publication for ensuring sufficient randomness of entropy sources

Entropy

Definition: Entropy

$$H(X) = \text{ the “uncertainty” in a random variable } X$$

Examples:

- A fair coin toss has an entropy of 1 bit
- A set of eight fair coin tosses has an entropy of 8 bits
- What about a weighted coin toss?
  - *Worst-case* uncertainty

Min-entropy

Definition: Min-entropy

If an adversary extracted *as much information* from this variable as possible, how much uncertainty remains?

$$H_\infty(X) = -\log p_{max}$$

- Min-entropy is the largest real number m such that all events occur with probability no greater than 2^{-m}
- A fair coin toss: $H_\infty(X) = -\log \frac{1}{2} = \mathbf{1\ bit}$
- A $\frac{3}{4}$-weighted coin toss: $H_\infty(X) = -\log \frac{3}{4} \approx \mathbf{0.415\ bits}$

NIST DRAFT Special Publication 800-90B

Recommendation for the Entropy Sources Used for Random Bit Generation

Elaine Barker

John Kelsey

Computer Security Division

Information Technology Laboratory

800-90B Tests

The 800-90B statistical tests are split into two categories:

- IID tests
  - Shuffle the data to ensure IID
  - Calculate the min-entropy directly using $H_\infty(X) = -\log p_{max}$
- Non-IID tests
  - **Collision test**
  - Partial collection test
  - Markov test (map 6 bits)
  - Compression test
  - Frequency test

Definition: IID

A dataset that is IID (Independent and Identically Distributed) means that each sample is independent from each other sample, and that each sample uses the same distribution.

Background takeaways

- Min-entropy is the "worst-case" measure of the uncertainty of a variable
- For the purposes of this talk, entropy and min-entropy are interchangable
- IID means "independent and identically distributed."
- NIST has statistical tests (800-90B) to evaluate entropy sources, sorted into IID and non-IID tests.
- My job: Evaluate the tests in 800-90B

Random does not mean secure

Question

If an entropy source passes the statistical tests in 800-90B, is it secure?

Proof by counterexample

Problem

We want a sequence of numbers that *looks* random, but we know each next outcome with 100% certainty.

Proof by counterexample

Problem

We want a sequence of numbers that *looks* random, but we know each next outcome with 100% certainty.

Soultion

Test digits of irrational numbers like π.

Constants chosen

- π
- $\zeta(3)$
- $\log 2$

The entropy of π


```
Read in file ../../../general-bbp/pi/piBin/piBin_8, 1250000 bytes long.
Dataset: 1250000 8-bit symbols.
Output symbol values: min = 0, max = 255
        collision test not run:
        mu_bar = 20.7708, max valid value for this test and model = 20.7261
- Collision test *not valid* for this data set.
- Partial collection test : p(max) = 0.00609112, min-entropy = 7.35908
~~- Markov test (map 6 bits): p(max) = 1.05961e-223, min-entropy = 5.78677~~
- Compression test        : p(max) = 0.00686646, min-entropy = 7.18622
- Frequency test          : p(max) = 0.0041008, min-entropy = 7.58853

***************
** min-entropy = ~~5.78677~~ 7.18622
***************
```

```
Read in file ../../../general-bbp/pi/picopy_8_8, 1250000 bytes long.
Dataset: 1250000 8-bit symbols.
Output symbol values: min = 0, max = 255

Compression Test:        Result = passed         Over/under Test:        Result = passed
Excursion Test:          Result = passed         Directional runs Test:  Result = passed
Covariance Test:         Result = passed         Collision Test:         Result = passed
** Passed iid shuffle tests
Chi square independence score = 65169.2, degrees of freedom = 65280, cut-off = 66402.2
** Passed chi-square independence test
Chi square stability    score = 2358.01, degrees of freedom = 2313 cut-off = 2528.88
** Passed chi-square stability test

min-entropy for ../../../general-bbp/pi/picopy_8_8 = 7.88435
```

The entropy of $\zeta(3)$

Read in file ../../../zeta3-bbp/zeta3Bin/zeta3Bin_8, 1250000 bytes long.
Dataset: 1250000 8-bit symbols.
Output symbol values: min = 0, max = 255
- Collision test : p(max) = 0.00997925, min-entropy = 6.64685
- Partial collection test : p(max) = 0.00644398, min-entropy = 7.27783
- ~~Markov test (map 6 bits): p(max) = 2.24652e-223, min-entropy = 5.7783~~
- Compression test : p(max) = 0.00830078, min-entropy = 6.91254
- Frequency test : p(max) = 0.0040824, min-entropy = 7.59365

****
** min-entropy = ~~5.7783~~ 6.64685
****

Read in file ../../../zeta3-bbp/zeta3Bin_8_8, 1250000 bytes long.
Dataset: 1250000 8-bit symbols.
Output symbol values: min = 0, max = 255

Compression Test: Result = passed
Excursion Test: Result = passed
Covariance Test: Result = passed
Over/Under Test: Result = passed
Directional runs Test: Result = passed
Collision Test: Result = passed
** Passed iid shuffle tests
Chi square independence score = 65424.1, degrees of freedom = 65280, cut-off = 66402.2
** Passed chi-square independence test
Chi square stability score = 2263.29, degrees of freedom = 2313 cut-off = 2528.88
** Passed chi-square stability test

****
min-entropy for ../../../zeta3-bbp/zeta3Bin_8_8 = 7.89074
****

The entropy of log 2

Read in file ../../../general-bbp/log2/log2Bin, 1250000 bytes long.
Dataset: 1250000 8-bit symbols.
Output symbol values: min = 0, max = 255
- Collision test : p(max) = 0.00619507, min-entropy = 7.33466
- Partial collection test : p(max) = 0.00695324, min-entropy = 7.1681
~~- Markov test (map 6 bits): p(max) = 9.35661e-224, min-entropy = 5.78817~~
- Compression test : p(max) = 0.00756836, min-entropy = 7.0458
- Frequency test : p(max) = 0.0041352, min-entropy = 7.57901

****
** min-entropy = ~~5.78817~~ 7.0458
****

Read in file ../../../general-bbp/log2/log2Bin, 1250000 bytes long.
Dataset: 1250000 8-bit symbols.
Output symbol values: min = 0, max = 255

Compression Test:	Result = passed	Over/Under Test:	Result = passed
Excursion Test:	Result = passed	Directional Runs Test:	Result = passed
Covariance Test:	Result = passed	Collision Test:	Result = passed

** Passed iid shuffle tests
Chi square independence score = 64754.7, degrees of freedom = 65280, cut-off = 66402.2
** Passed chi-square independence test
Chi square stability score = 2186.42, degrees of freedom = 2313 cut-off = 2528.88
** Passed chi-square stability test

****
min-entropy for ../../../general-bbp/log2/log2Bin = 7.87249
****

What have we learned?

What have we learned?

- A constant is an entropy source with *zero* entropy.

What have we learned?

- A constant is an entropy source with *zero* entropy.
- But, these constants pass all our tests for entropy.

What have we learned?

- A constant is an entropy source with *zero* entropy.
- But, these constants pass all our tests for entropy.
- Conceptually, there is *no way* to tell the difference between random and nonrandom entropy sources using only the output.

What have we learned?

- A constant is an entropy source with *zero* entropy.
- But, these constants pass all our tests for entropy.
- Conceptually, there is *no way* to tell the difference between random and nonrandom entropy sources using only the output.
- **There must be a qualitative as well as quantitative assessment of entropy sources.**

Back to the pi result...


```
Read in file ../../../general-bbp/pi/piBin/piBin_8, 1250000 bytes long.
Dataset: 1250000 8-bit symbols.
Output symbol values: min = 0, max = 255
        collision test not run:
        mu_bar = 20.7708, max valid value for this test and model = 20.7261
- Collision test *not valid* for this data set.
- Partial collection test : p(max) = 0.00609112, min-entropy = 7.35908
- Markov test (map 6 bits): p(max) = 1.05961e-223, min-entropy = 5.78677
- Compression test        : p(max) = 0.00686646, min-entropy = 7.18622
- Frequency test          : p(max) = 0.0041008, min-entropy = 7.58853

**************
** min-entropy = 5.78677
**************
```

The Collision Test

The Collision Test

Measures the min-entropy by observing time between "collisions" (repeated outputs)

- Data set: BACBCBACCAABAABCBC
- Marked collisions: BACB CBAC CAA BAA BCB C
- Collision difference times: 4 4 3 3 3
- Collision statistic: avg$(4, 4, 3, 3, 3)$

Collision Statistic

- From a data set $\{X_s\}$, define a sequence of collision times $\{T_i\}$, where $T_0 = 0$, and

$$T_i = \min\{j > T_{i-1} : \exists m \in (T_i - 1, j)\,\text{such that}\, X_j = X_m\}$$

Collision Statistic

$$S_k = \text{average of differences of collision times}$$

$$\begin{aligned} S_k &= \frac{1}{k}\sum_{i=1}^{k}(T_i - T_{i-1}) \\ &= \frac{1}{k}\left((T_1 - T_0) + (T_2 - T_1) + \cdots + (T_k - T_{k-1})\right) \\ &= \boxed{\frac{T_k}{k}} \end{aligned}$$

Expected Value of S_k

Collision Statistic

$$S_k = \frac{T_k}{k}$$

- Calculate expected value for a probability distribution $\mathbf{p}$:

$$\begin{aligned}
\mathbb{E}_{\mathbf{p}}[S_k] &= \mathbb{E}_{\mathbf{p}}[T_k/k] \\
&= \frac{1}{k}\mathbb{E}_{\mathbf{p}}[T_k] \\
&= \mathbb{E}_{\mathbf{p}}[T_1] \\
&= \sum_{i=1}^{n+1} P_i
\end{aligned}$$

(P_i = probability that no collisions have occurred after i outputs)

Expected Value of S_k

Collision Statistic

$$S_k = \frac{T_k}{k}$$

- Calculate expected value for a probability distribution **p**:

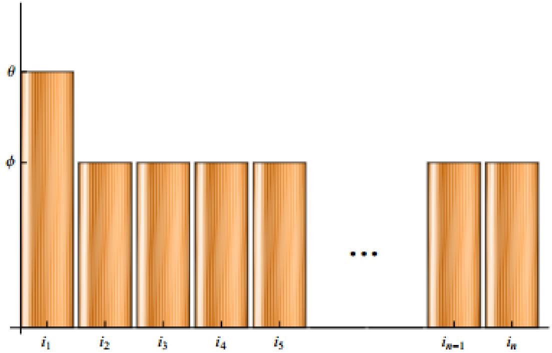
$$\begin{aligned}
\mathbb{E}_{\mathbf{p}}[S_k] &= \mathbb{E}_{\mathbf{p}}[T_k/k] \\
&= \frac{1}{k}\mathbb{E}_{\mathbf{p}}[T_k] \\
&\boxed{= \mathbb{E}_{\mathbf{p}}[T_1]} \text{ (requires IID)} \\
&= \sum_{i=1}^{n+1} P_i
\end{aligned}$$

(P_i = probability that no collisions have occurred after i outputs)

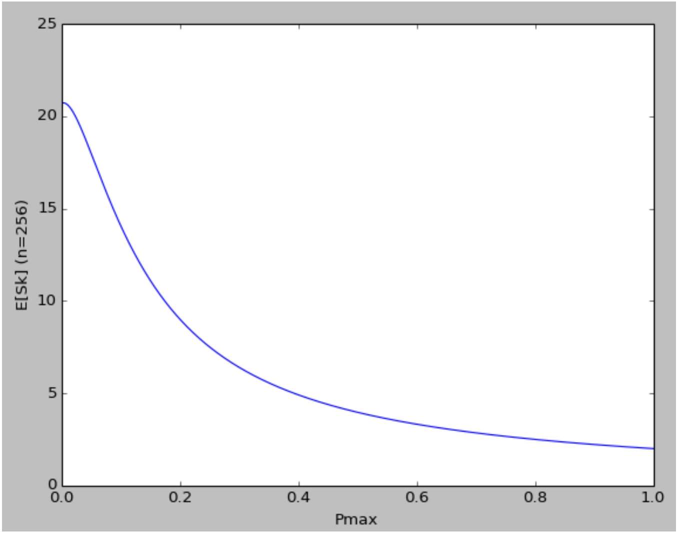
Near-uniform distributions

Near-uniform distribution family

$$\mathbf{p}_\theta[Z = i] = \begin{cases} \theta & i = i_1, \\ \frac{1-\theta}{n-1} & \text{otherwise.} \end{cases}$$

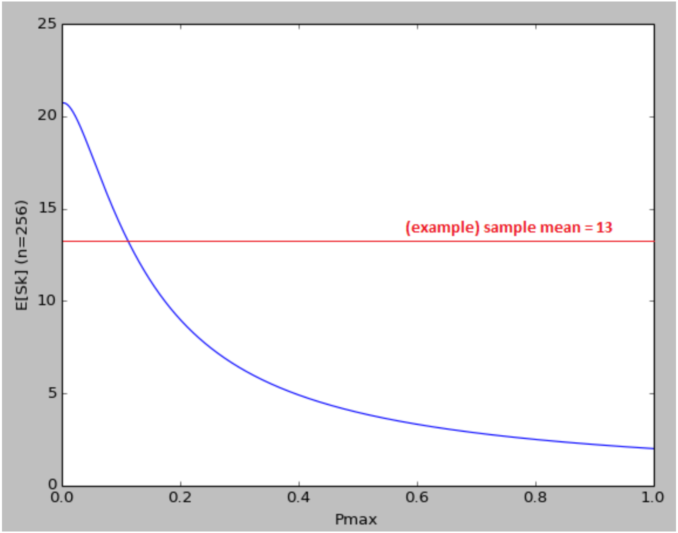


Steps of the Collision Test



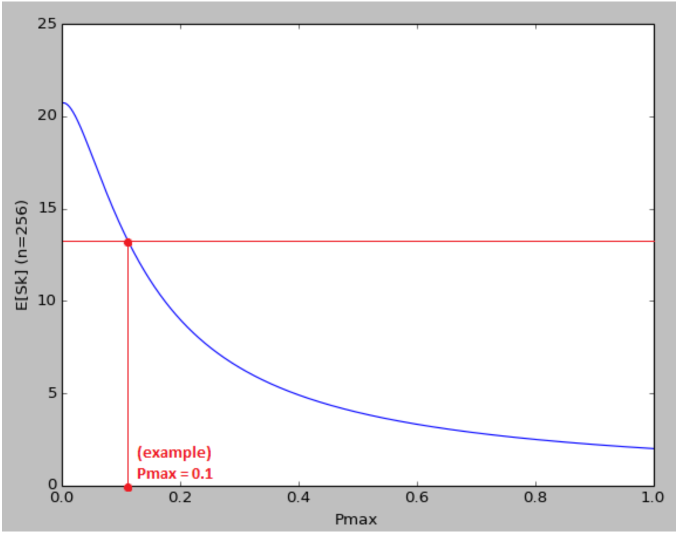
1. Get a family of near-uniform expected values

Steps of the Collision Test



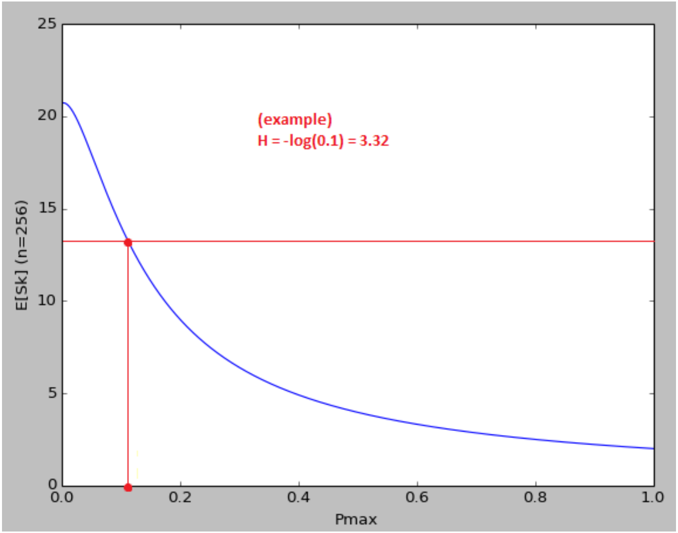
2. Compare sample mean to near-uniform expected values

Steps of the Collision Test



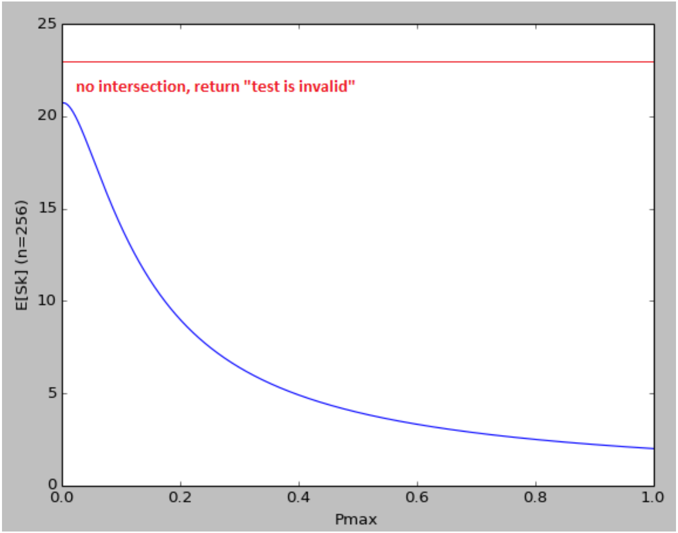
3. Get the p_{max} that corresponds to this expected value

Steps of the Collision Test



4. Calculate min-entropy directly

Why is the test invalid?



- There is no intersection, so the test cannot calculate the min-entropy

Why is this happening?

Collision Statistic

$$S_k = \frac{T_k}{k}$$

- Calculate expected value for a probability distribution $\mathbf{p}$:

$$\begin{aligned}
\mathbb{E}_{\mathbf{p}}[S_k] &= \mathbb{E}_{\mathbf{p}}[T_k/k] \\
&= \frac{1}{k}\mathbb{E}_{\mathbf{p}}[T_k] \\
&\boxed{= \mathbb{E}_{\mathbf{p}}[T_1]} \quad \text{(requires IID)} \\
&= \sum_{i=1}^{n+1} P_i
\end{aligned}$$

(P_i = probability that no collisions have occurred after i outputs)

Conclusions

Conclusions

- Even when the Collision Test "succeeds," it may be overestimating the expected value, which would cause a false increase in the entropy measurement.

Conclusions

- Even when the Collision Test “succeeds,” it may be overestimating the expected value, which would cause a false increase in the entropy measurement.
- **The Collision Test is being run on non-IID data, but it requires its data to be IID.**

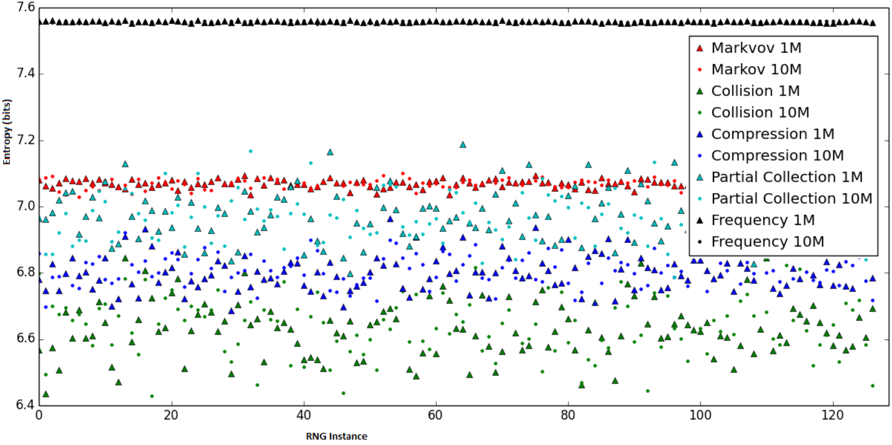
Conclusions

- Even when the Collision Test "succeeds," it may be overestimating the expected value, which would cause a false increase in the entropy measurement.
- **The Collision Test is being run on non-IID data, but it requires its data to be IID.**
- The Partial Collection Test and the Compression Test also assume IID, but are not in the IID section of 800-90B.

Conclusions

- Even when the Collision Test "succeeds," it may be overestimating the expected value, which would cause a false increase in the entropy measurement.
- **The Collision Test is being run on non-IID data, but it requires its data to be IID.**
- The Partial Collection Test and the Compression Test also assume IID, but are not in the IID section of 800-90B.
- But when the data are IID, why not just measure the min-entropy directly?

"Non-IID" tests in 800-90B



Acknowledgements

- My advisor, Dr. Allen Roginsky, for giving me this project, for providing help and guidance, and for having really cool conversations
- Dr. Tim Hall, for being the source of many insights, for showing me the Python code, and for a lot of support
- Fellow SURF intern Bobby Staples, for helping me figure out the 90B's, and sending me that scatterplot
- The SURF directors, especially Dr. David Griffith, for organizing many events and helping me through some logistics
- Dr. Isabel Beichl, for providing some neat opportunities and conversations
- The Crypto Cave, for being awesome and fun friends
- Anyone not on this list