A tale of two entropy source validation approaches

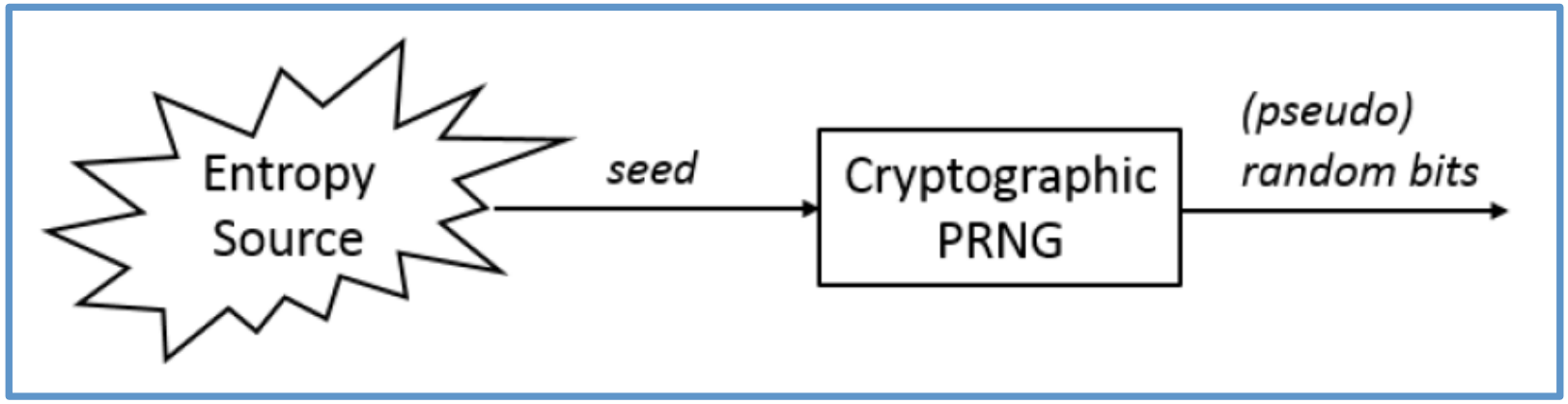
NIST 800 90B vs. BSI AIS 31

Meltem Sonmez Turan

National Institute of Standards and Technology

Random Numbers in Cryptography

- In cryptography, the security relies on the unpredictability of random numbers (e.g., cryptographic key, IV, nonce).
- Cryptographic PRNGs (based on hash functions, block ciphers etc.) use *seed* to generate strong random numbers.
- Entropy sources are based on unpredictable physical events (e.g., thermal noise, oscillator jitter, radioactivity, user interaction, system data, mouse movements)



- Hard to design and validate entropy sources

What is Entropy?

- Measure of randomness or uncertainty
- Different formulations based on the *probability distribution* of the random variables

Min-entropy: $H=-\log(\max p(x))$

Shannon: $H=-\sum\uparrow\blacksquare p(x)\log p(x)$,

Renyi entropy: $H=1/1-\alpha \log\sum\uparrow\blacksquare p(x)\uparrow\alpha$

- Challenging to estimate entropy when the probability distribution and the dependence between variables are unknown.

Real World – Mining your Ps and Qs

- Heninger et al. (2012) performed the largest network survey of TLS and SSH servers.
  - Collected 5.8 million unique certificates from 12.8 million TLS hosts and 6.2 million unique SSH host keys from 10.2 million hosts.
  - Observed that 0.75% of TLS certificates share keys during key generation.
  - Obtained RSA private keys for 0.50% of TLS hosts and 0.03% of SSH hosts, DSA private keys for 1.03% of SSH hosts.
- Why?
  - Using manufacturers-default keys,
  - Low **entropy** during key generation.

Nadia Heninger, Zakir Durumeric, Eric Wustrow, J. Alex Halderman "*Mining Your Ps and Qs: Detection of Widespread Weak Keys in Network Devices*", Proceedings of the 21st USENIX Security Symposium, August 2012.

Real World – Coppersmith in the wild

- Factoring RSA keys from certified smart cards: Coppersmith in the wild
  - Efficiently factor 184 distinct RSA keys out of more than two million 1024-bit keys downloaded from Taiwan's national "Citizen Digital Certificate" database.
  - Low entropy (e.g., hardware RNG stuck in a short cycle with period 3 001 001 001 …)

Daniel J. Bernstein, Yun-An Chang, Chen-Mou Cheng, Li-Ping Chou, Nadia Heninger, Tanja Lange, Nicko van Someren: Factoring RSA Keys from Certified Smart Cards: Coppersmith in the Wild. ASIACRYPT (2) 2013: 341-360

Guidelines for Constructing Random Numbers

NIST

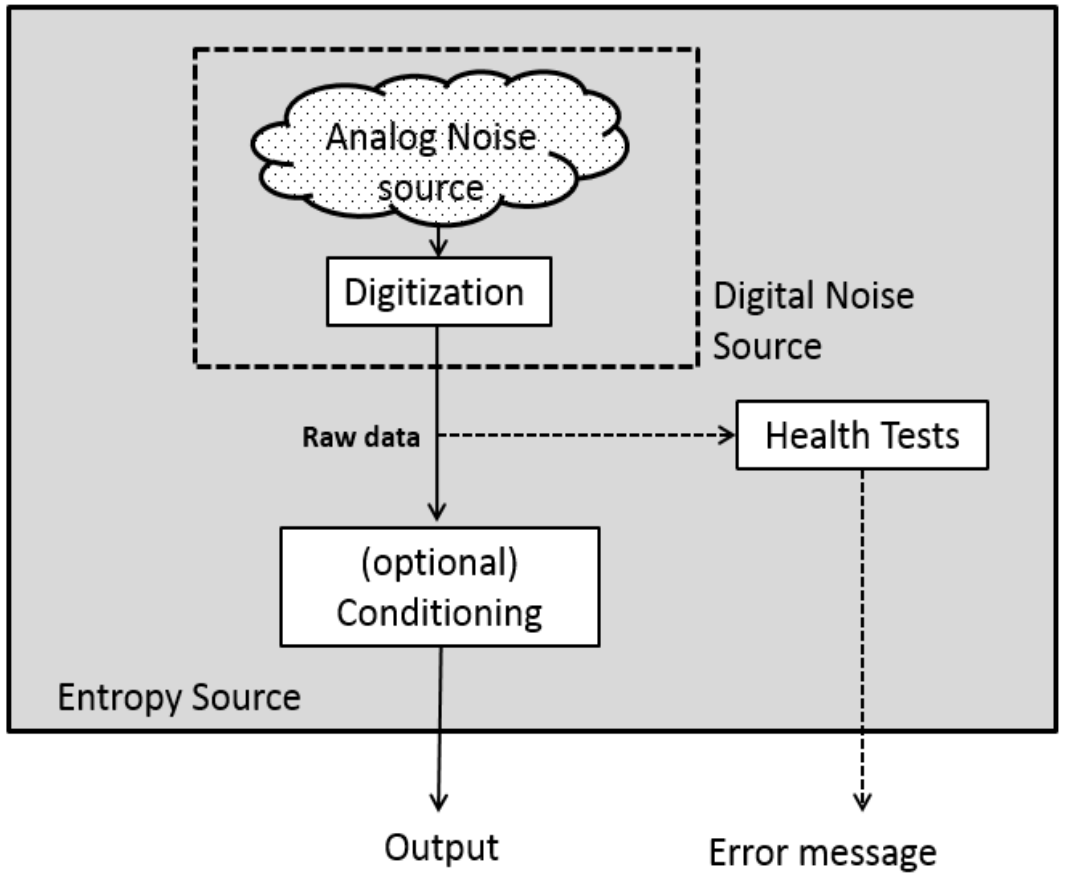
- SP 800-90A Recommendation for Random Number Generation Using Deterministic Random Bit Generators
- SP 800-90B Recommendation for the Entropy Sources Used for Random Bit Generation
- SP 800-90C Recommendation for Random Bit Generator Constructions

BSI

- AIS 20: Functionality Classes and Evaluation Methodology for Deterministic Random Number Generators
- AIS 31: Functionality Classes and Evaluation for Physical Random Number Generators

NIST’s approach

NIST's Entropy Source Model



- Noise Source extracts randomness from a physical phenomena (black box approach).
- Health Tests detect deviations from the intended behavior of the entropy and the noise source, during operation.
- Conditioning may increase the statistical quality of noise source outputs or entropy rate.

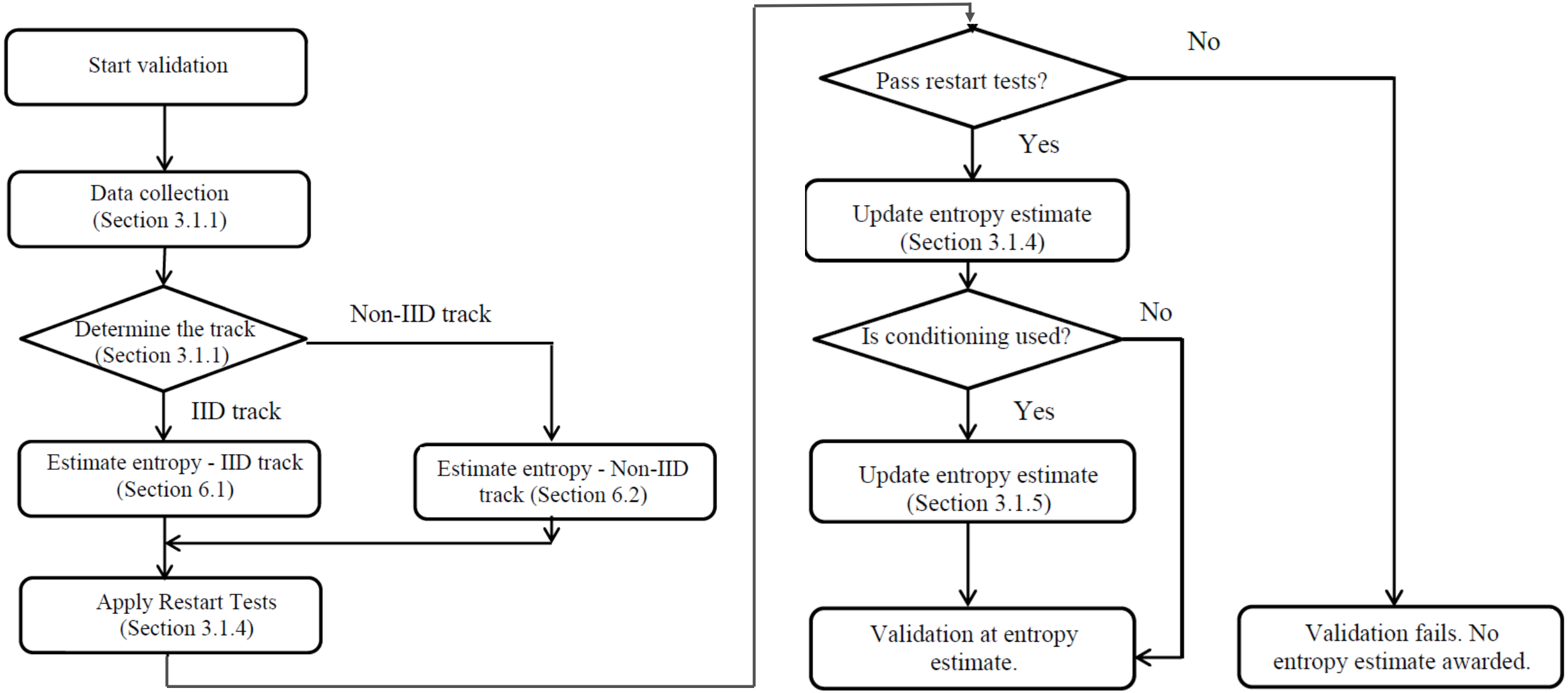
NIST SP 800-90B: Basic Requirements

- Documentation on the entire design of the entropy source, noise source, including interactions between components, parameter selections
- Justification for why the source can be relied upon to produce bits with entropy
- Requirements on the noise source, conditioning component, health tests
- Requirements on data collection
- Range of operating conditions
- Entropy estimate from the submitter
- etc.

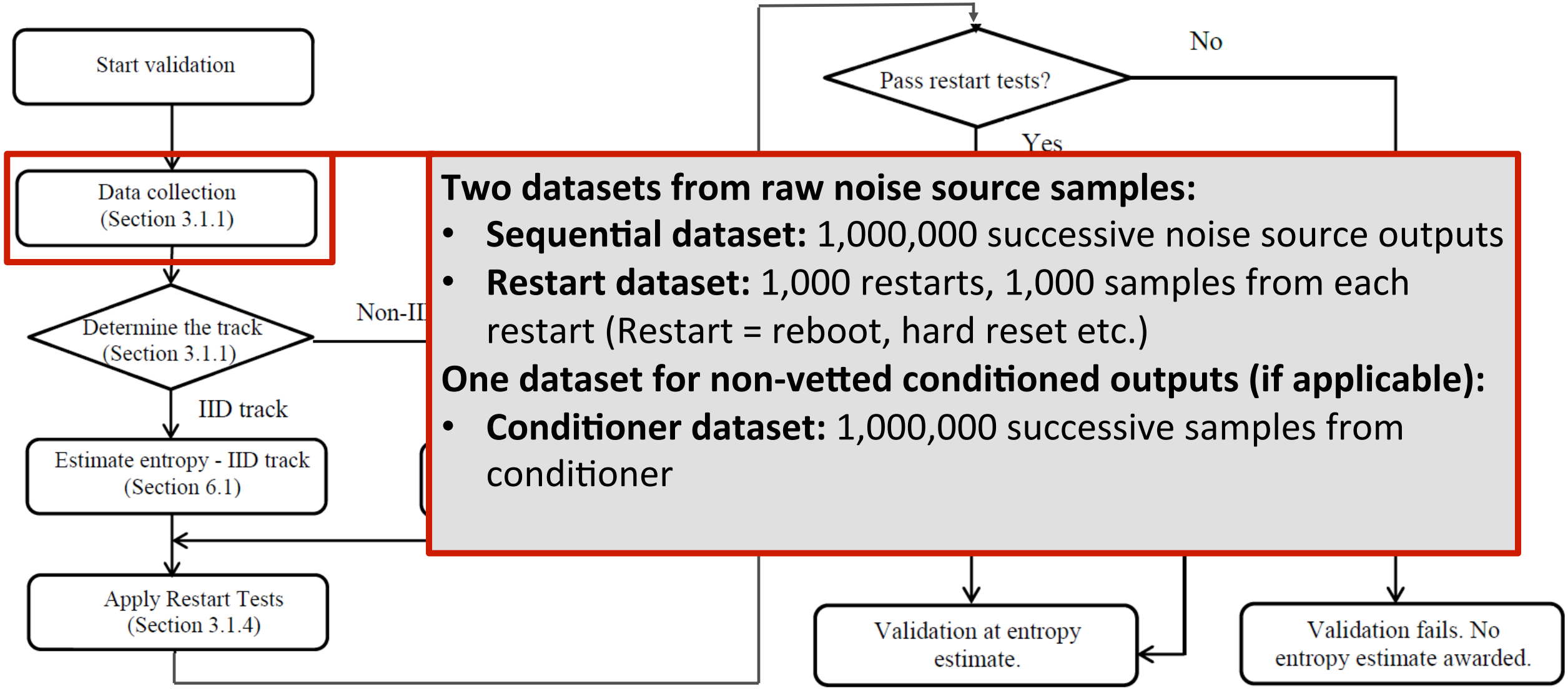
NIST Validation

- In order to comply with Federal Information Processing Standards 140-2, designers/vendors first prepares a submission package with all the required documentation.
- Labs checks the documentation, and generate data from the source, and estimate entropy using the 90B estimators.
- For validation purposes, estimation process cannot be too complex, due to time and cost constraints, and the process cannot require a lot of expertise specific to a noise source.
- Any two validation labs must get the same result for the same source.

Validating the Entropy Source



Validating the Entropy Source – Data Collection



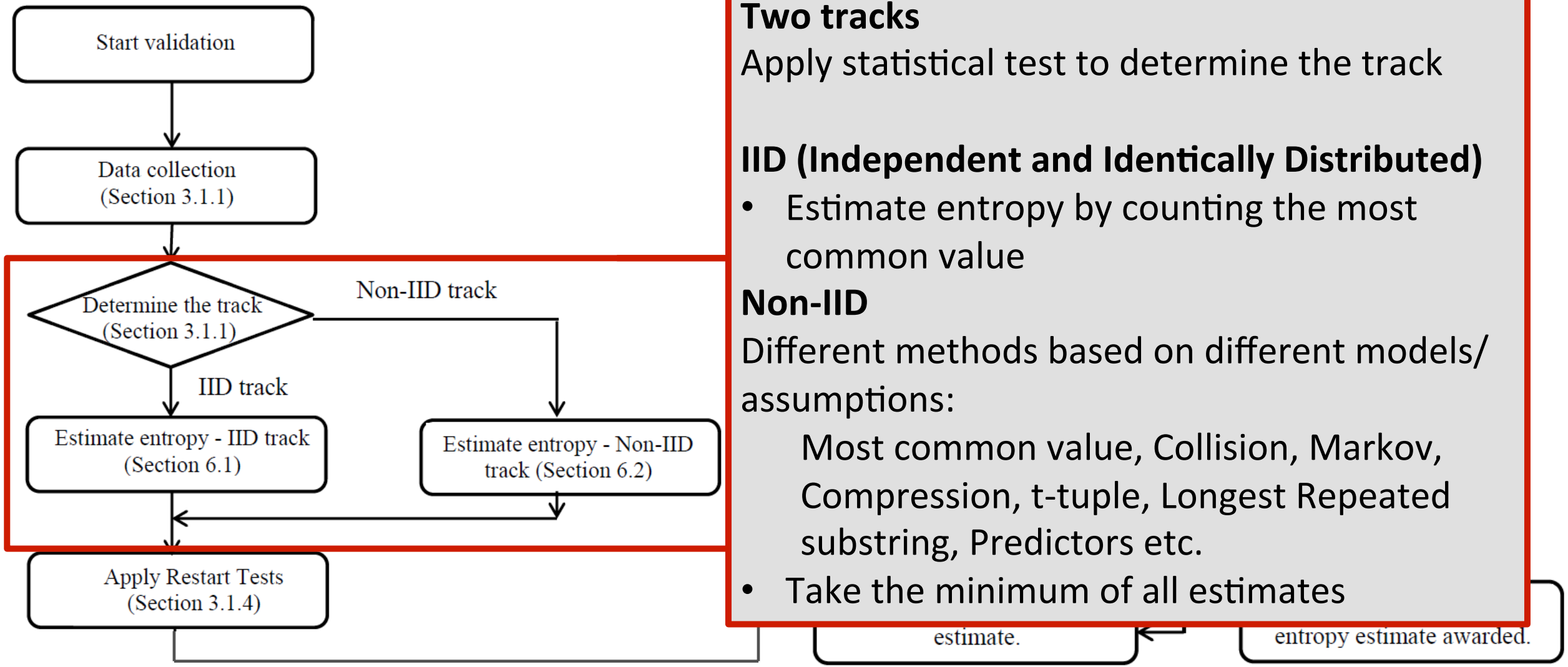
Two datasets from raw noise source samples:

- **Sequential dataset:** 1,000,000 successive noise source outputs
- **Restart dataset:** 1,000 restarts, 1,000 samples from each restart (Restart = reboot, hard reset etc.)

One dataset for non-vetted conditioned outputs (if applicable):

- **Conditioner dataset:** 1,000,000 successive samples from conditioner

Validating the Entropy Source – Determining the Track



Two tracks

Apply statistical test to determine the track

IID (Independent and Identically Distributed)

- Estimate entropy by counting the most common value

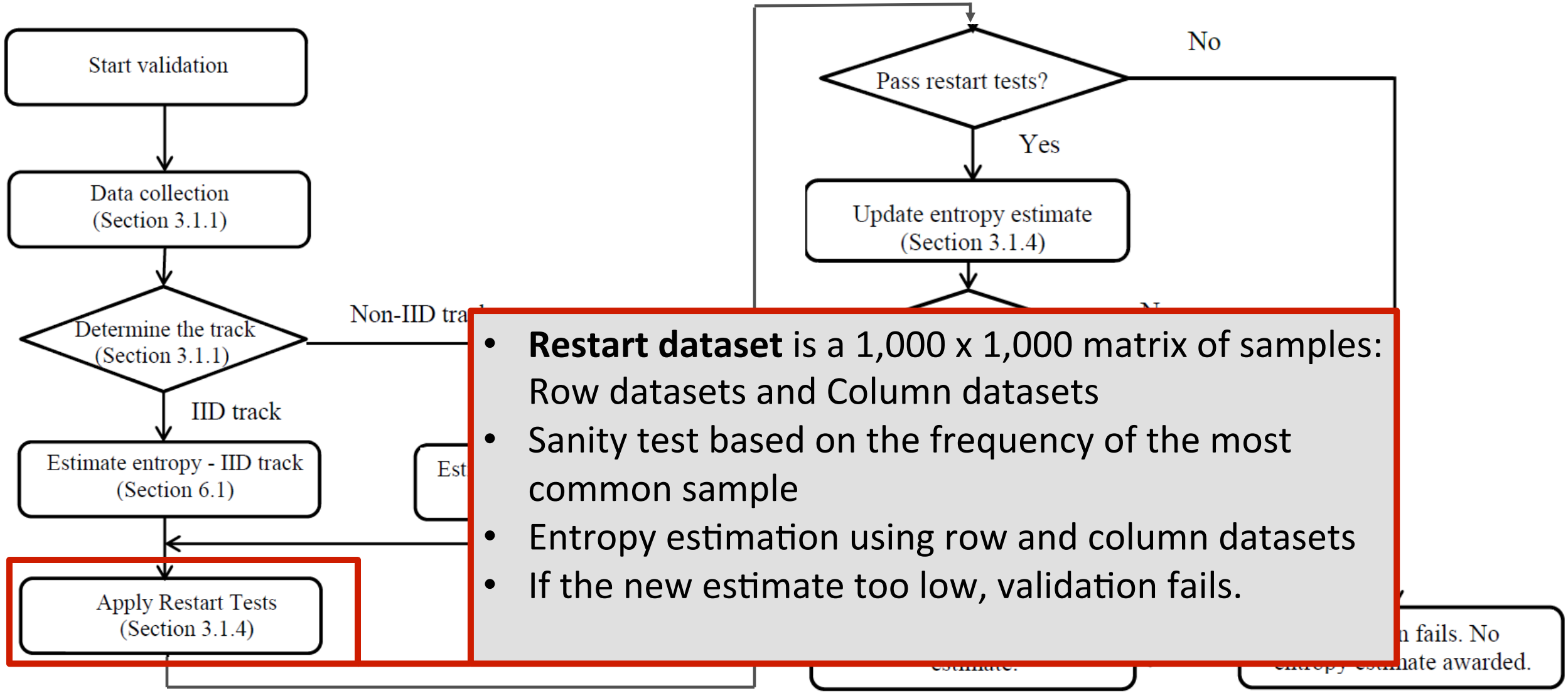
Non-IID

Different methods based on different models/ assumptions:

Most common value, Collision, Markov, Compression, t-tuple, Longest Repeated substring, Predictors etc.

- Take the minimum of all estimates

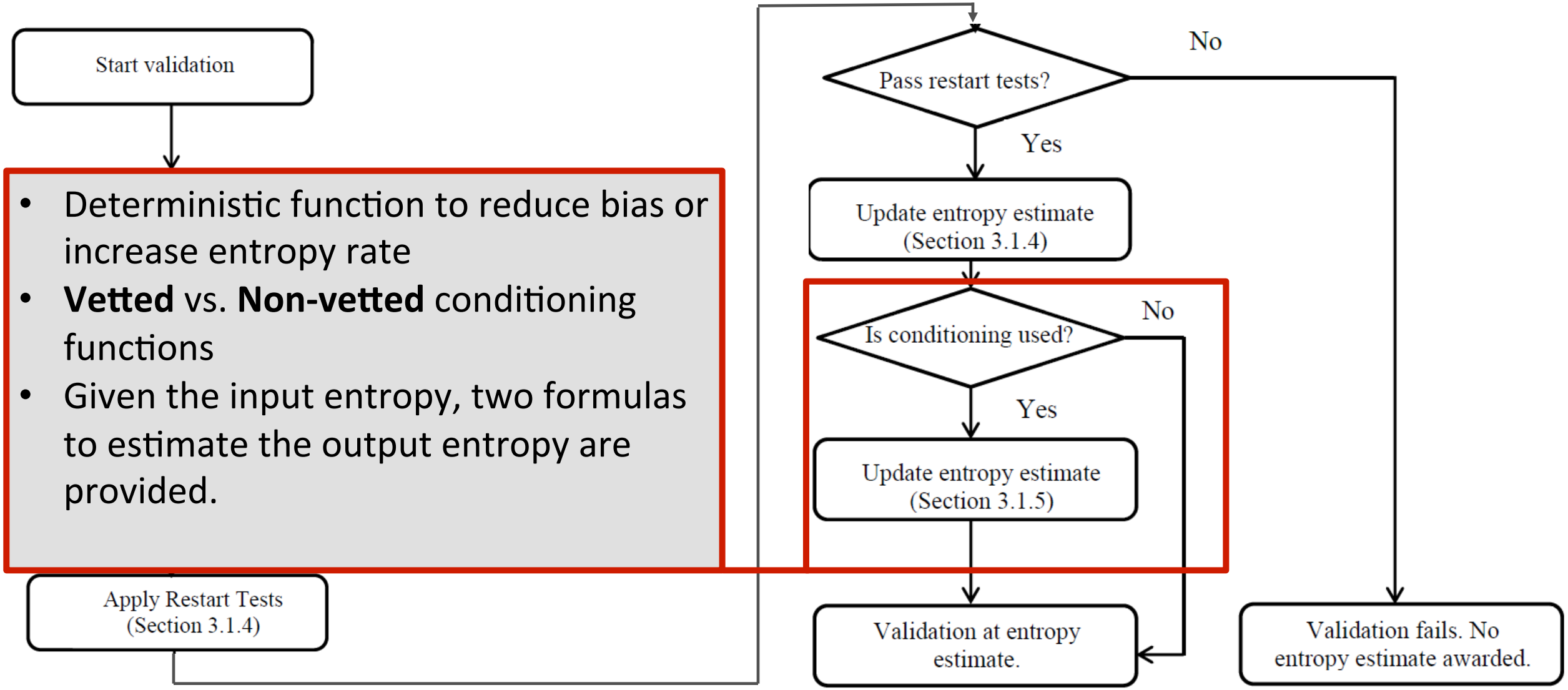
Validating the Entropy Source – Restart Tests



- **Restart dataset** is a 1,000 x 1,000 matrix of samples: Row datasets and Column datasets
- Sanity test based on the frequency of the most common sample
- Entropy estimation using row and column datasets
- If the new estimate too low, validation fails.

Validating the Entropy Source

- Deterministic function to reduce bias or increase entropy rate
- **Vetted** vs. **Non-vetted** conditioning functions
- Given the input entropy, two formulas to estimate the output entropy are provided.



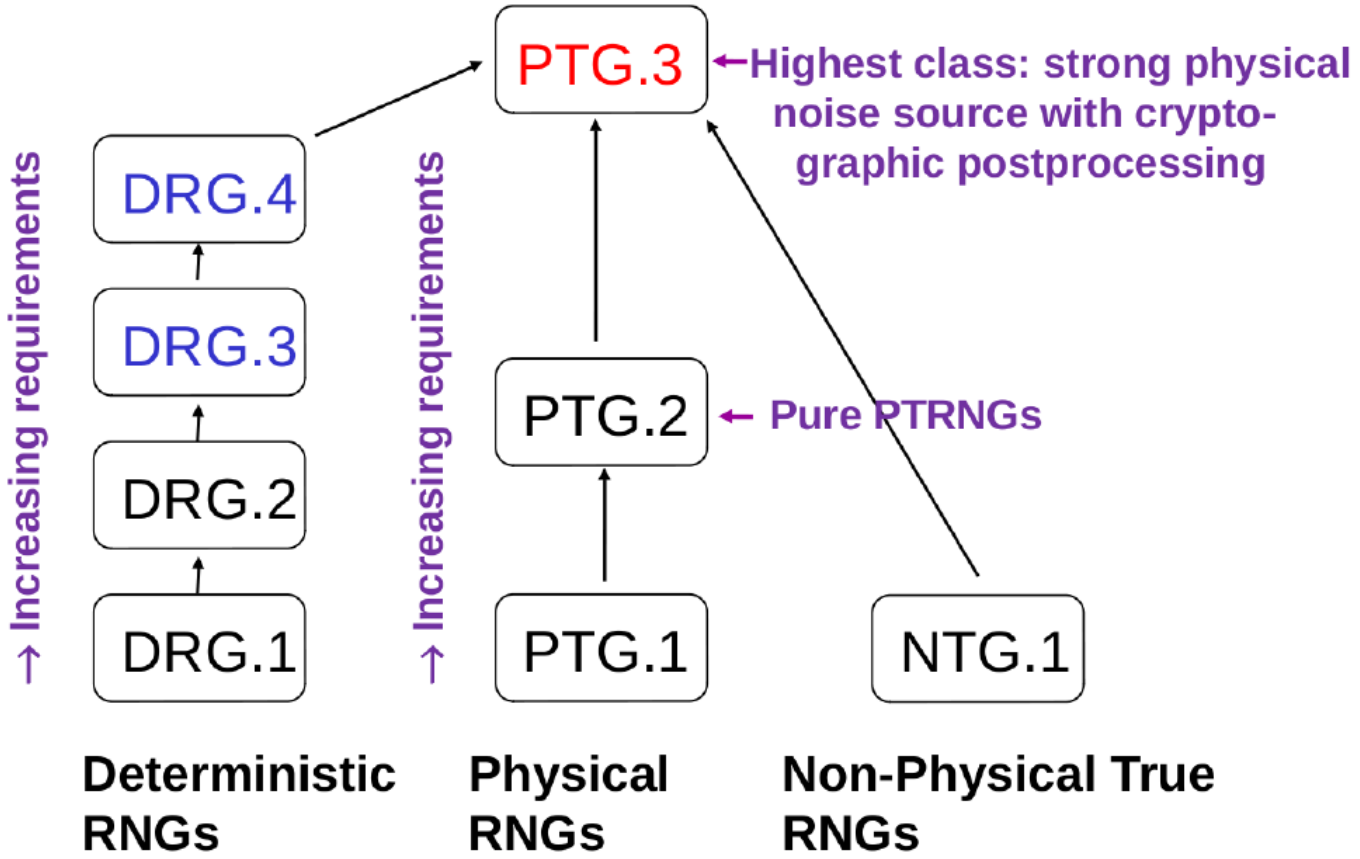
BSI’s approach

Functionality Classes

- DRG.1: forward secrecy
- DRG.2:+backward secrecy
- DRG.3: +enhanced backward secrecy
- DRG.4: +enhanced forward secrecy

- PTG.1: internal tests to detect failures
- PTG.2: +stochastic model and statistical tests
- PTG.3: +cryptographic post-processing

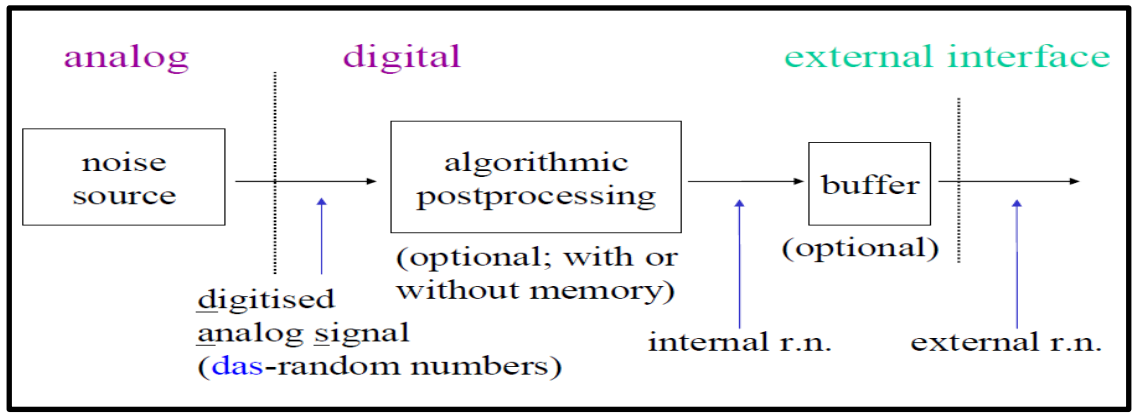
- NTG.1: Non-physical RNG with entropy estimation



Model - BSI

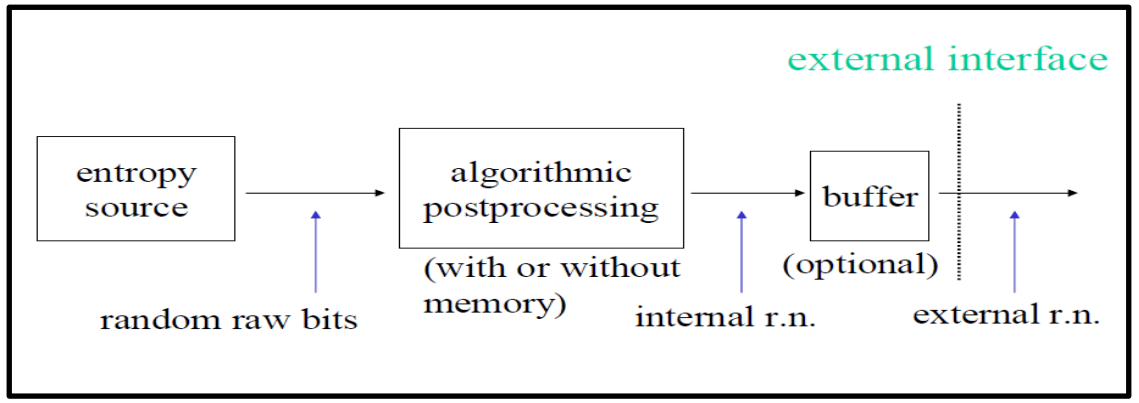
Physical TRNG

- Noise source is dedicated hardware.
- Behaves similar for all copies of the RNG.
- Accurate modelling is suitable.



Non-Physical TRNG

- Exploits system data/human interaction
- Different behavior depending on the platform
- Does not allow accurate modelling.



Evaluation of Physical TRNG

- Good understanding of the PTRNG based on a stochastic model
  - A stochastic model specifies a family of probability distributions that contains the true distribution of the internal random numbers.
  - Hard to explicitly specify the distribution, but a family of distributions that contain the true distribution is sufficient.
  - The stochastic model is verified by measurements/experiment.
  - To estimate entropy the parameters are estimated first, then an entropy estimate is computed.
  - Grey box approach to estimate entropy.
- Online health tests tailored to the characteristics of the design.
- Requires a lot of expertise on the analysis.

Conclusion

- Differences in the BSI's and NIST's validation process, in terms of
  - definitions, requirements, modeling and the evaluation process etc.

- Aligning the standards as much as possible is desired.
  - NIST can emphasize the use of stochastic models.
  - BSI can utilize statistical entropy estimation methods to verify their results.

Thanks!

meltem.turan@nist.gov