Strong Randomness from Common Hardware
Chris Federici (cfederic@stevens.edu), Professor Antonio Nicolosi (nicolosi@cs.stevens.edu)

Introduction
- In order to properly encrypt data for transport over the internet, strong random numbers are required
- Server-side results cannot be trusted
- A client-side solution is required, but how to get sufficiently random numbers from unknown hardware?
- Assumptions: User has a physical keyboard and a Javascript enabled browser
- Is randomness from typing enough?
- We decided to utilize delay in typing as a source of randomness

Experiment
- First, a body of random text (approx. 80 character minimum) is generated by a Markov chain (often nonsensical).
- The user is tasked to type in this text.
- The collection of delays between presses are used as the seed for the Pseudo-Random Number Generator (PRNG).
- All of this happens client-side via Javascript, which lets the user inspect the exact method of seed generation to verify its integrity.
- Snag in the setup: How much data from the user is required to meet the randomness requirements?
- New mission: What kind of random variable is down-down typing delay?
- Gather and analyse delays, look at different views
- Does the type of word matter? Which hand types it? Etc. Gather data and find out!
- Things to look at:
  - Left hand or right hand? Take “left hand” keys to be “azwxsedcrfv”, or to the left of “tgb” on a QWERTY keyboard, and let “right hand” keys be “yhnjumikolp”, ignoring case.
  - Depends on what words and combinations of letters are given

Results
- Down-down delay, or the timing between pressing down on two different keys, is log-normal
- That means if the random variable X is log normal, then let Y = ln(X), Y is a normally distributed random variable
- Looking at different “hand” transitions gives to fuzzy of an idea.
- As a whole, indeed is log-normal
- Entropy of a log-normal distribution:
  \[ \frac{1}{2} + \frac{1}{2} \ln(2\pi \sigma^2) + \mu \]

Conclusion
- An online survey to collect more data from a larger sample population is needed
- While the actual delays change from person to person, they maintained a log-normal distribution, but need more data.
- Some variation, where did it come from? Perhaps a Gaussian mixture?
- Need to determine what kind of random variable to determine how much text to generate
- Very possible and easy to implement