


**From:** Jack Vanlightly jack.vanlightly@gmail.com   
**Subject:** Talk submission: Obtaining statistical properties  
**Date:** June 28, 2022 at 10:42 AM  
**To:** tla2022@tlapl.us



# Obtaining Statistical Properties via TLC Simulation

## About me

I currently work as a researcher, analyst and technical advisor at Confluent. I am a distributed systems engineer and have been working with TLA+ for a number of years. Some recent highlights were discovering a bug in the Apache BookKeeper replication protocol with TLA+ and identifying issues of identity reuse in Raft reconfigurations. You can read my writings on distributed systems and messaging systems on my website <https://jack-vanlightly.com>.

## The talk

In this talk I will describe how simulation can be used to obtain statistical properties of algorithms and how we can apply this technique using TLA+ with TLC in simulation mode.

Understanding the statistical properties of an algorithm are often as important as verifying its correctness. These statistics can pertain to both liveness and correctness. An example of liveness statistics are the number of protocol rounds to achieve a steady state, or the time to detect some condition. Examples of correctness are calculating the probability of a data loss event given a correlated failure.

In some cases an algorithm must be discarded or heavily modified because it exhibits some problematic behaviour that cannot be detected through standard formal verification techniques be it correctness or liveness. These kinds of problems may only come to the surface late in the development cycle or worse, while in production.

Some algorithms are simple enough so that the statistical properties can be calculated using statistics, probability and combinatorics theory. Other algorithms may be too complex to realistically calculate. Another factor is the engineer's level and experience of statistics which can lead to either an inability to make those calculations, lead to lack of confidence or worse, relying on incorrect results.

The talk will be broken down into the following sections:

- Why are statistical properties valuable?
- How simulation can be used to obtain statistical properties.

- Specifically how can this technique be applied using TLA+ and TLC.
- Real-world use cases where I have applied this technique while working on distributed data systems as VMware and Splunk.
- The changes we made to TLC to simplify the collection of statistics and other features that improved non-determinism in simulations.
- A summary of the challenges of using TLA+ and TLC for this simulation technique.
- Future work

For a more complete description of this research, you can find a more detailed write-up here: <https://docs.google.com/document/d/1pW-RuxN5Nfw6sNpfO67R2j4dTajS19G9bKCyuJu32Ric/edit?usp=sharing>

Kind regards  
Jack Vanlightly