Contribution ID: **220**   Type: **Oral**

An Analysis of Reproducibility and Non-Determinism in HEP Software and ROOT Data

Thursday, 13 October 2016 14:15 (15 minutes)

Reproducibility is an essential component of the scientific process.
It is often necessary to check whether multiple runs of the same software
produce the same result. This may be done to validate whether a new machine
produces correct results on old software, whether new software produces
correct results on an old machine, or to compare the equality of two different approaches to the same objective.

Unfortunately, many technical issues in computing make it surprisingly hard to get the same result twice. Non-determinism in both codes and data can arise unexpectedly from the use of concurrency, random number sources, real-time clocks, I/O operations, and other sources. Some of these issues are merely cosmetic, like displaying the current time and machine for diagnostic purposes. Some of these issues are much more subtle, such as race conditions between threads changing the order of records in a file. As a result, one cannot simply compare objects at the binary level: different bits might reflect fundamentally different algorithms, or the might reflect accidents of the runtime environment.

In this talk, we will present an analysis of non-determinism and reproducibility issues in standard HEP software and the ROOT data format.
We will present an evaluation of simulation and analysis codes to reveal sources of non-determinism, so that this property
can be eliminated (where possible) and understood (where necessary).
Where non-determinism is unavoidable, we present tool (root-diff) for
the detailed comparison of output files to evaluate differences between structured data, so as to distinguish between files that differ in fundamental data, as opposed to structure, ordering, or timestamps. We conclude with recommendations for best practices in software development, program analysis, and operations.

Tertiary Keyword (Optional)

Secondary Keyword (Optional)

Accounting and information

Primary Keyword (Mandatory)

Preservation of analysis and data

Primary author: Prof. THAIN, Douglas (University of Notre Dame)

Co-authors: Mr ZHENG, Charles (University of Notre Dame); Mr IVIE, Peter (University of Notre Dame)

Presenter: Prof. THAIN, Douglas (University of Notre Dame)

Session Classification: Track 8: Security, Policy and Outreach

Track Classification: Track 8: Security, Policy and Outreach