Reproducible Research for OMNeT++ Based on Python and Pweave

Kyeong Soo (Joseph) Kim

Department of Electrical and Electronic Engineering

Xi'an Jiaotong-Liverpool University

07 September 2017

Outline

- Reproducible Research
- Python and Pweave
- Reproducible Research for OMNeT++
- Example: OMNeT++ FIFO Simulation

Reproducible Research

Reproducible Research

- Reproducible research is a key to any scientific method and ensures repeating an experiment and the results of its analysis in any place with any person.
- A study can be truly reproducible when it satisfies at least the following three criteria:
 - All experimental methods are fully reported.
 - All data and files used for the analysis are (publicly) available.
 - The process of analyzing raw data is well reported and preserved.
- Reproducible research is to ensure
 - Same data + Same script = Same results

Why Do We Need Reproducible Research: Two Examples

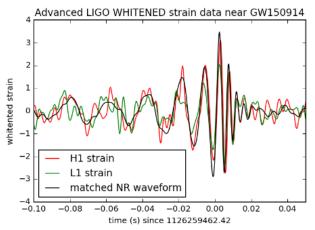
- LIGO Gravitational Wave Detection
- Schön scandal Molecular Computing

LIGO - Gravitational Wave Detection

- The <u>Laser Interferometer</u>
 Gravitational-Wave Observatory
 (<u>LIGO</u>) is a large-scale physics
 experiment and observatory to
 detect cosmic gravitational
 waves.
 - The detection of gravitational wave was reported in *Physical Review Letters* in Feb. 2016, together with <u>ipython notebook</u> with analysis code and data.



```
In [9]: # We need to suppress the high frequencies with some bandpassing:
        bb, ab = butter(4, [20.*2./fs, 300.*2./fs], btype='band')
        strain_H1_whitenbp = filtfilt(bb, ab, strain_H1_whiten)
        strain L1 whitenbp = filtfilt(bb, ab, strain L1 whiten)
        NR H1 whitenbp = filtfilt(bb, ab, NR H1 whiten)
        # plot the data after whitening:
        # first, shift L1 by 7 ms, and invert. See the GW150914 detection paper
        strain L1 shift = -np.roll(strain L1 whitenbp,int(0.007*fs))
        plt.figure()
        plt.plot(time-tevent, strain_H1_whitenbp, 'r', label='H1 strain')
        plt.plot(time-tevent, strain_L1_shift, 'g', label='L1 strain')
        plt.plot(NRtime+0.002,NR H1 whitenbp,'k',label='matched NR waveform')
        plt.xlim([-0.1,0.05])
        plt.ylim([-4,4])
        plt.xlabel('time (s) since '+str(tevent))
        plt.ylabel('whitented strain')
        plt.legend(loc='lower left')
        plt.title('Advanced LIGO WHITENED strain data near GW150914')
        plt.savefig('GW150914_strain_whitened.png')
```

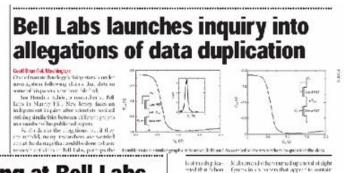


Schön Scandal - Molecular Computing

- No records found for his groundbreaking experimental results, including lab notebook, experimental samples and data, hard disk drives.
- During the investigation, he kept repeating

"I clearly observed them in the

Lab but ..."



Misconduct finding at Bell Labs shakes physics community

Physicists are coming to turns this week with me of the most and done admitted fruids verial and at one of the age of a rest-known. industrial laboratories and published in tagcouncil - including this one see editoric

date are div., page 427, Anisd pendentee-mittee-rangel with extending work to no by let Front the Schore at 3dl Laboratories in Murray Hill, New



": found a "prepondence of evidence" that he iddfied or febric med data in 16 of the 24. ren anomand. The renearth in causion was a direct case of a not death at i. locked at. rechang Expendence of papers.
"Handrik Schon showed reckess diere

gard for the lancing of data in the value system is science' sold the report of the committee, server was crained by Marchin. Services on effectively engineer of Stanford University in California Bellians - percofselecement rications corporation Locate of Scionthe coming before the finding controllered, intitle after its for feets over with physicists asking not only what sould have been dimeno presentitio irand, battako what will account of the balls nat Schöm archered sincy arriving at Bell-Tom Germany in 1908.

School dui. od somali g"mistaka "Js. se d

Schor

the walks of red middle enhanced account to.

single molecules and induce superconduc-In a statement accompanying the several thicks in carbon lenders as were time schin, and his findings advanced the

Bell Labs inquiry spreads to superconductors

months, better

tion at Bell, Johnstonias and Some Expanded. Four eithe Execution Crossp. a Wishington. It is linear, will not a research to say competito include these important papers on superconduct vity published in Name.

sid that he had support data, the researchers stylln one case, suretainment he tangraph look de trival over the mental sales says Charles , along a contest at

The coding most clawesign or, which is being conducted by a panel chaired by Make in Beakley in Sun ford Union sity at California, now slarmed physicians downsof where have been engaged in aforts to replicas. In a perconductivity work.

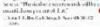
The expersion is the or single planter. at Del Less in Marray E. J. New Je cruighageener, Luorat Ischnologies, lost areast add a began ailebit r he subaction lest week set graph, right), as product radio to extra of a mail definor-

Alle ... Langue aborne and one it hased market and job firm. "Many or themtive in the sergirum across congesion a ducing forces are fireign expressed againment contact the future of bank

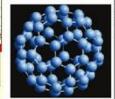
withbackupicy." Devict Larber, a processor of telecommu-Philadelphia enda berno securencia dile. Labe, also sees tought inner ahead for the ab-

research them. The continue to be expende sismions at the U. county ... Parato raniad to westing the autopassmion tells polegies to make our company outsets: craws W. argust come to direct as de you a certited that Beac arch, he core, But

in IRI, till district betweenth.



Sawata Diesa spokamina and Bell Laka



N.C. School agree being salind into proving

Is a bell tolling for Bell Labs?

It would be wise of Bell abs to help others reproduce the iscientists' results

Paul Grant

Durk dixade quiddy bogan to gather over the exceptional finding of superconductivity at 1.7 & reported bet you by menonic to conand collaboratory at Hell Laboratories in Marray Hill, New Jeney, Shortly after publication of the paper, two salue by a seport are work of my reputation as a soptical observer of



thorough encoration of all and during the back ground pressure of the system is in the same 1011 same states states vapour is the major component. The presents of a unfinition hydrocide in plasma depocted films of ale. mins not been discurrented (. M. Schwiczer erslapel Pine Lee 74, 200-202 (1999) and is a source of some of the concern over the variability Schöninger; ein sample steld und

commentary

Python and Pweave

R/Sweave to Python/Pweave

- Until recently, R was the language of choice for statistical processing and data analysis.
 - Still, R has the largest code base for a wide variety of statistical and graphical techniques.
- Like *ipython* (now *jupyter*), R provides a nice tool called *Sweave* (now replaced by *knitr*) to weave documentation and the results of the execution of R code chunks into one source file for integrated documentation.
- Python one of the most popular languages in scientific computing, including artificial intelligence & machine learning — recently takes over R in statistical processing and data analysis as well.
 - Thanks to <u>pandas</u> implementing DataFrame object similar to R and <u>Pweave</u>,
 python can replace R for most statistical and data analysis tasks, while retaining
 its many advantages over R (i.e., fully-featured programming language with easy
 syntax and higher speed).

```
### customize
.old <- theme set(theme bw())</pre>
.pt size <- 3.5
### generate summary plots for reference architecture with N=1
.rf N1.data <- paste(.rf N1.wd, paste(.rf N1.base, "data", sep="."), sep="/")
.df <- read.csv(.rf N1.data, header=TRUE)</pre>
## .df <- .df[order(.df$N, .df$n, .df$dr, .df$br, .df$repetition), ] # order data frame
.df <- sort_df(.df, vars=c("N", "n", "dr", "br", "repetition")) # sort data frame</pre>
.rf_N1.df <- ddply(.df, c(.(n), .(dr)), function(df) {return(GetMeansAndCiWidths(df))})</pre>
.rf_N1.plots <- list()</pre>
for (.i in 1:7) {
     .df <- subset(.rf_N1.df, select = c(1, 2, (.i*2+1):((.i+1)*2)))
    names(.df)[3:4] <- c("mean", "ci.width")</pre>
    .limits <- aes(ymin = mean - ci.width, ymax = mean +ci.width)</pre>
     .p <- ggplot(data=.df, aes(group=dr, colour=factor(dr), x=n, y=mean)) + geom line() + scal</pre>
.width)))
    .p <- .p + xlab("Number of Users per ONU (n)") + ylab(.labels.measure[.i])</pre>
    ## .p <- .p + geom point(aes(group=dr, colour=factor(dr), x=n, y=mean), size=.pt size)
    .p <- .p + geom_point(aes(group=dr, shape=factor(dr), x=n, y=mean), size=.pt_size) + scale</pre>
    .p <- .p + geom_errorbar(.limits, width=0.1) + scale_colour_discrete("Line Rate\n[Gb/s]")</pre>
    .rf_N1.plots[[.i]] <- .p
```

Snippets of R Source Code and Sweave File for LaTeX

Pweave source file ("*.Plw")

 Mix of documentation (e.g., LaTeX) and code Chunks (e.g., Python)

Weaving (pweave)

Tangling (ptangle)

Document 1

Result of the Execution of Code Chunk 1

• • •

Document N

Result of the Execution of Code Chunk N

A documentation source file

e.g., "*.tex" for LaTeX

Code Chunk 1

• • •

Code Chunk N

A program source file for separate execution and debugging

• e.g., "*.py" for Python

Weaving Example: Automatic Table Generation

The following Python code chunk can automatically generate a long table over multiple pages from a pandas dataframe¹:

```
<<echo=False,results='raw'>> =
import numpy as np
df = fifo_df.filter(regex="^(?!(r|R)un).*$")  # exclude columns starting with run/Run
print(df.to_latex(longtable=True))
@
```

Weaving & LaTeXing



The following Python code chunk can automatically generate a long table over multiple pages from a pandas dataframe¹:

```
<<echo=False,results='raw'>> =
import numpy as np
df = fifo_df.filter(regex="^(?!(r|R)un).*$")  # exclude columns starting with run/Run
print(df.to_latex(longtable=True))
a
```

	File	Module	Name	Unnamed: 19
0	Fifo1-st=0.01-#0.sca	_runattrs_	st	0.010000
1	Fifo1-st=0.01-#0.sca	FifoNet.fifo	queueingTime:mean	0.000262
2	Fifo1-st=0.01-#0.sca	FifoNet.fifo	queueingTime:max	0.031311
3	Fifo1-st=0.01-#0.sca	FifoNet.fifo	busy:timeavg	0.049941
4	Fifo1-st=0.01-#0.sca	FifoNet.fifo	qlen:timeavg	0.001308
5	Fifo1-st=0.01-#0.sca	FifoNet.fifo	qlen:max	4.000000
6	Fifo1-st=0.01-#0.sca	FifoNet.sink	lifetime:mean	0.010262

¹Note that a space is inserted between '»' and '=' to prevent Pweave from weaving the code; it seems that there is no way to escape Pweave chunk code markers.

Continued on next page

Reproducible Research for OMNeT++

How to Deal with Simulation Input Files

- Include them the document.
 - OK for small simulations
- Use a snapshot of the whole configurations.
 - e.g., git commit hashes

```
Author: Kyeong Soo (Joseph) Kim <kyeongsoo.kim@gmail.com>
                                                                                         Mon Feb 27 08:59:31 2017 +0000
// This file is part of an OMNeT++/OMNEST simulation example.
                                                                                    Add ini file.
// Copyright (C) 1992-2015 Andras Varga
// This file is distributed WITHOUT ANY WARRANTY. See the file
                                                                                 ommit f1e7f6ad0265068d906efd02026e774076c00297
   'license' for details on this and other legal matters.
                                                                               Author: Kyeong Soo (Joseph) Kim <kyeongsoo.kim@gmail.com>
                                                                                         Mon Feb 27 08:56:07 2017 +0000
// Simple queueing network: generator + FIFO + sink.
                                                                                    Remove README.rst; only the markdown version of README
network FifoNet
   submodules:
                                                                                commit 8765336f9e2f5543fea8c4f37a0cf894da7f4c8e
      gen: Source [
                                                                               Author: Kyeong Soo (Joseph) Kim <kyeongsoo.kim@gmail.com>
            @display("p=89,100");
                                                                                         Sun Oct 2 17:32:02 2016 +0000
      fifo: Fifo (
         parameters:
            @display("p-209,100");
                                                                                    Change simulation time.
      sink: Sink (
         parameters:
            @display(*p=329,100*);
      gen.out --> fifo.in;
      fifo.out --> sink.in:
```

nmit 857ae37cd233914fd7271584afc4be10bcf75a61

Listing 1: 'FifoNet.ned' for FIFO sample model.

How to Guarantee Match Between Input Files and Output Data

- Online generation of results
 - Include simulation execution code within a document
 - Refer to the provided sample Pweave file.
 - OK for smaller simulations, but not for larger simulations.
- Use a snapshot of the whole configurations and data
 - e.g., git commit hashes
 - Version controlling output data together with source code and input configuration files, however, may greatly increase the size of a repository.

How to Present and Analyze Output Data

- Unstacking of stacked DataFrame
 - Use *pivot* function (see the example shown here).
- Aggregated processing of measurement data over independent variables
 - Use *pivot_table* function.
 - Useful for the calculation of mean and confidence intervals over multiple iterations.
- Online calculation of confidence intervals
 - Confidence intervals (CIs) can be calculated by assigning a custom function for CI to aggfunc parameter of pivot table function.
 - Now pandas support error bars in its own plot functions.

```
In [1]: df
Out[1]:
         date variable
                           value
  2000-01-03
                     A 0.469112
  2000-01-04
                     A -0.282863
 2000-01-05
                     A -1.509059
  2000-01-03
                     B -1.135632
4 2000-01-04
                       1.212112
  2000-01-05
                     B -0.173215
  2000-01-03
                        0.119209
   2000-01-04
                     C - 1.044236
  2000-01-05
                     C -0.861849
  2000-01-03
                     D -2.104569
10 2000-01-04
                     D -0.494929
11 2000-01-05
                     D 1.071804
```

Example: OMNeT++ FIFO Simulation