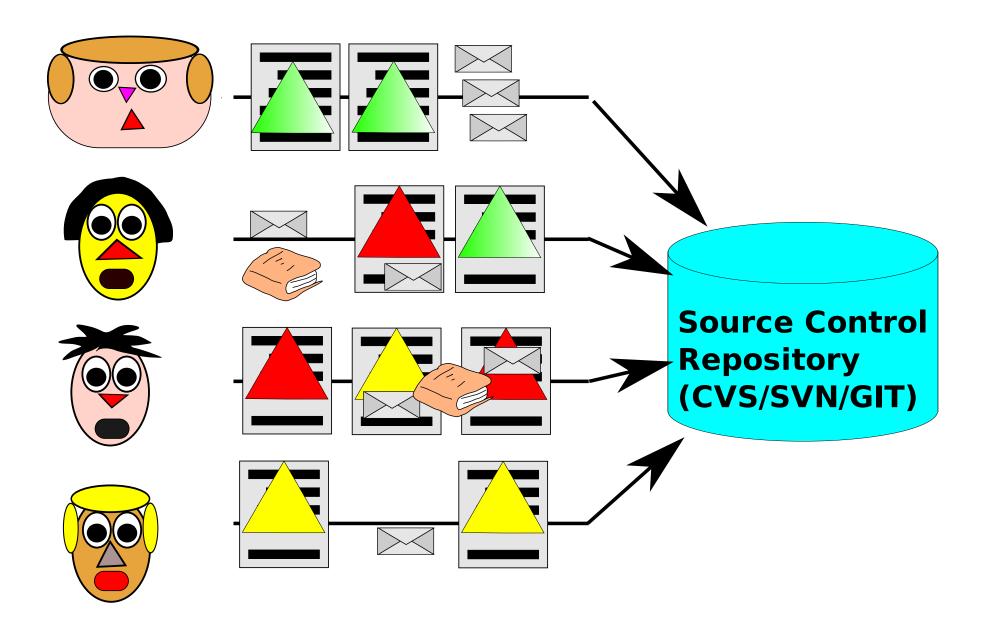
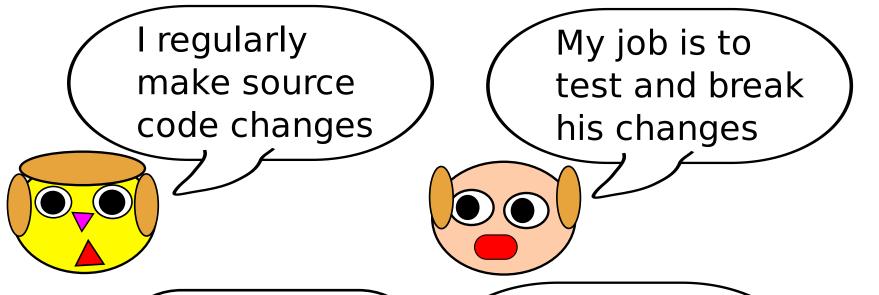
Mining Recurrent Activities: Fourier Analysis of Change Events

Abram Hindle, Michael Godfrey, Ric Holt University Of Waterloo, Waterloo, Canada

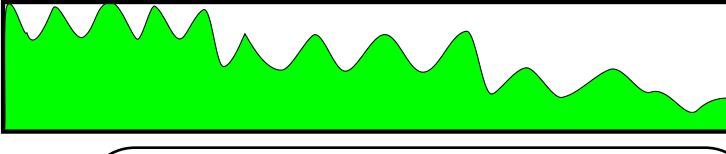
Developers create artifacts



Software development has recurrent behaviour



How do we discover recurrent behaviour? With Time-series analysis A signal, multiple years in length



You must choose a period to use

ASSUME A PERIOD

<u>A vear</u>

time-series analysis!

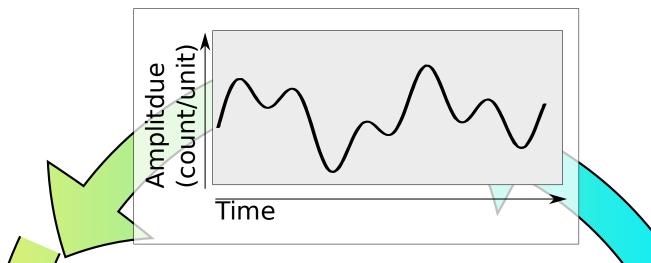
<u>A week</u>

Developers create various kinds of artifacts when they make changes:

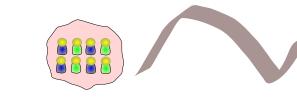
source code, test suites, bug reports, documentation, mailing list messages, etc



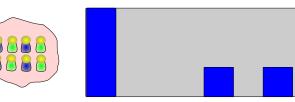
We can convert from a time/amplitude view to...



How can we apply the Fourier transform to software change and software related data?



We have a signal



We apply the Fourier transform.

Convert discrete events into signals

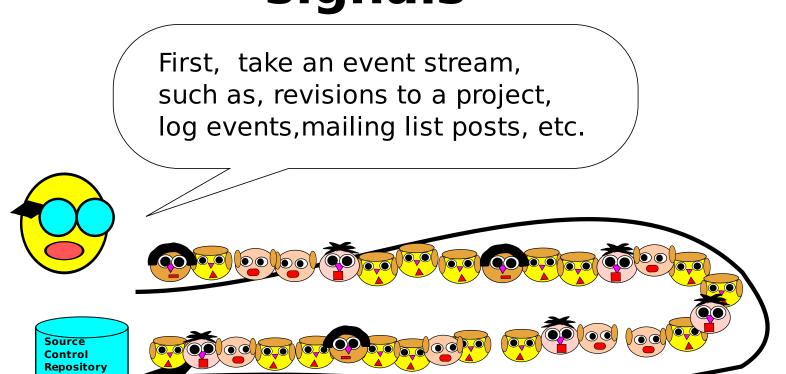
But which is correct or useful?

What period should I

by? If only I had a tool

analyze this signal

to tell me what do!



I hold regular am master of this current inspection iteration. meetings <u>A day</u> Within an iteration, there are recognizable repeating patterns. Time Iterations themselves are also a recurrent and repeating behaviour.

Recompose the signal from sinewaves in the signal down into sinewaves rime Break the signal down into sinewaves ... a frequency/magnitude view which

shows us dominating frequencies and periods

So what? Does it work?

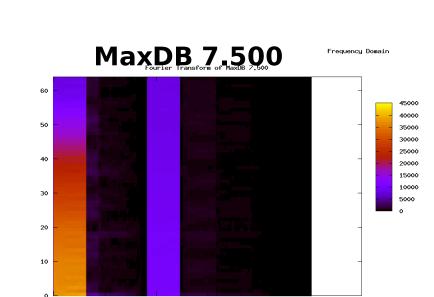
We can find periodicities of software Here's the Fourier transform of MaxDB 7.500

MaxDB 7.500 - Fourier Transform of revisions per day (16384 bins) Amplitude of Frequency

18000

Spectrograms of software change over time

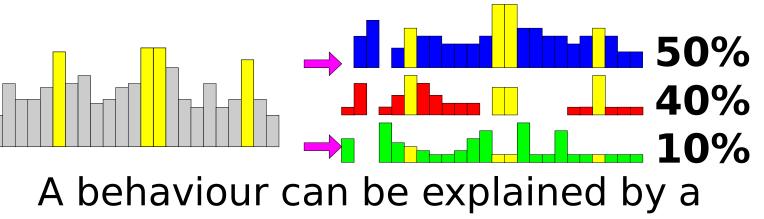
Spectrograms are Fourier transforms of short periods, shown side by side. The x-axis is time, y-axis frequency and color



lime —

Future Work

Who's responsible for this behaviour? Semantic Slicing



combination of slices.

Investigation of each frequency bin shows that 2 authors are individually responsible for 2 bins. Their frequencies imply their periodicities.

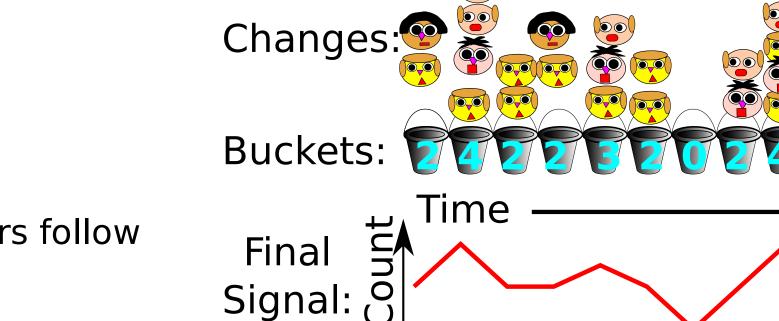
We discover that: * blue commits many

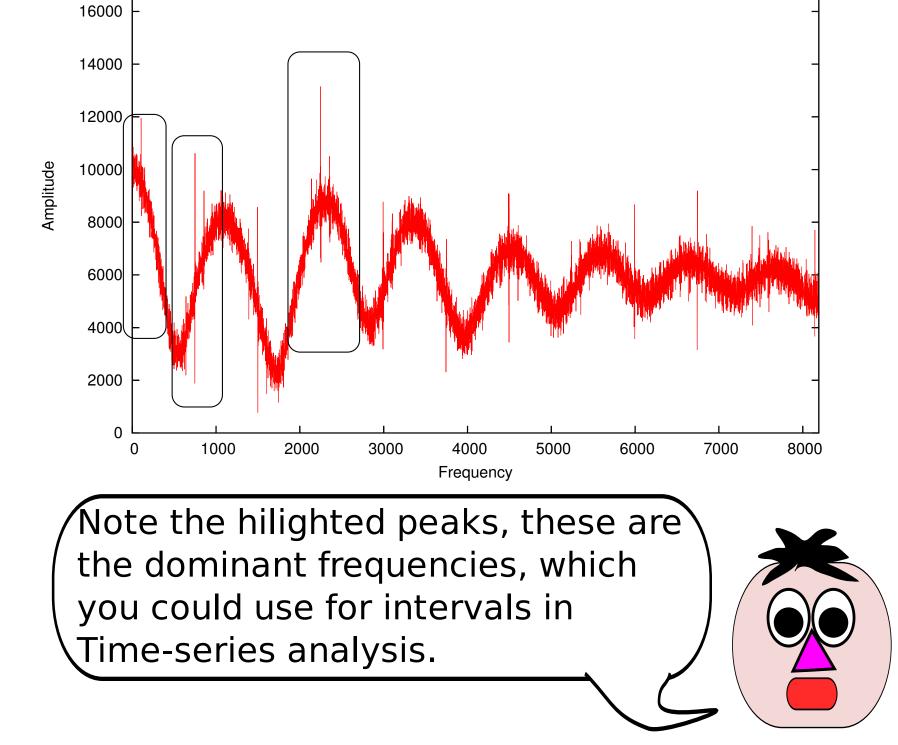
- changes
- * green tests blue's commits
- * The other developers follow their own schedule.

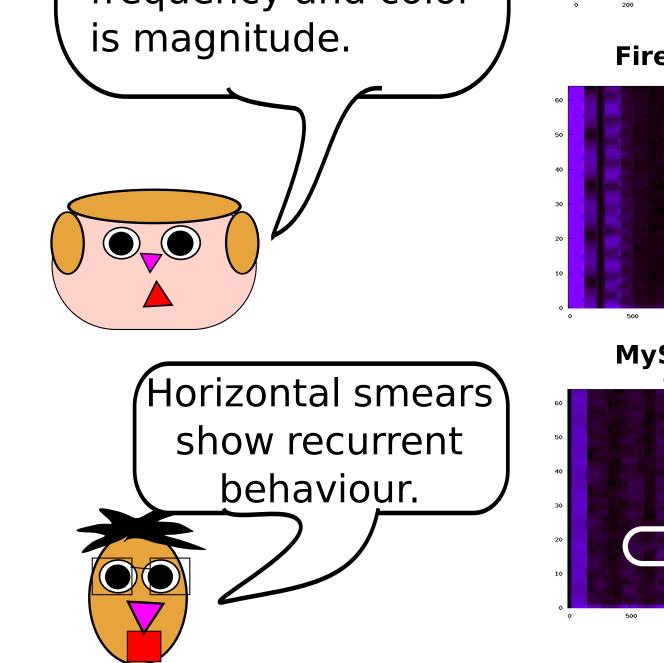
Then we take these events and aggregate them. In this example we put the changes into buckets of a day. Then we get a signal from

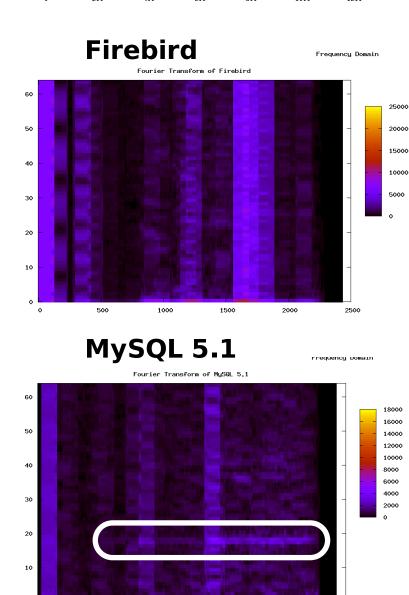
these buckets which we can analysis via Time-series Analysis or the Fourier Transform.



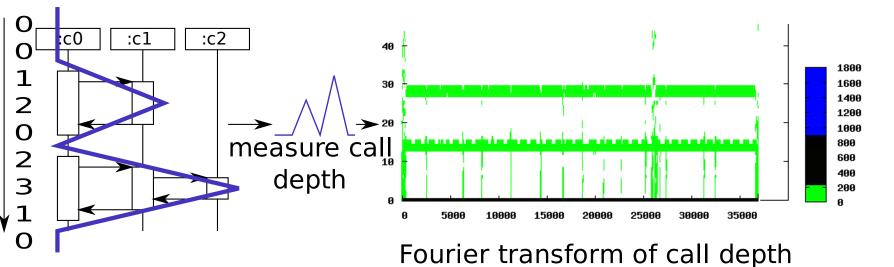








Other uses of the Fourier Transform: Dynamic Analysis via call depth



Partition development time via Self Similarity

