

Overlapping Experiment Infrastructure

@nid90, @sriharisriraman

@nilenso

Sections

- About the paper, and the problem statement
- Basic entities and terms in Experimentation
- Domains, Layers and Experiments
- Diversion types, Conditions
- Launch Layers
- Logic flow for a request
- An alternate approach

About the paper

- 2010
- KDD'10, July 25–28, 2010, Washington, DC, USA
- <http://research.google.com/pubs/pub36500.html>
- Used in google since ~2007
- Previous works (papers) do not cover scaling an experiment infrastructure and the overall experimentation environment to support running more experiments more quickly and more robustly

Why

- Traffic is precious; Traffic is not infinite (even for Google)
- Not 100s or 1000s of sessions; probably millions
- Difficult to get statistically significant results in a reasonable timeframe
- Multiple, Multivariate, Fast, Ramp-ups

Terminology

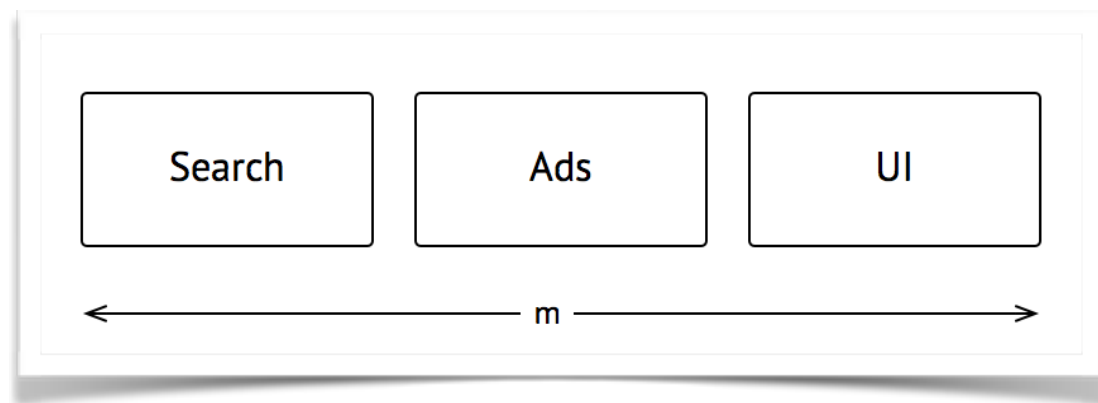
Basic entities and terms

- Treatment / Bucket
- Controlled experiments
- A/B testing
- Website Testing
- MultiVariable Testing
- Multifactorial testing

Parameters

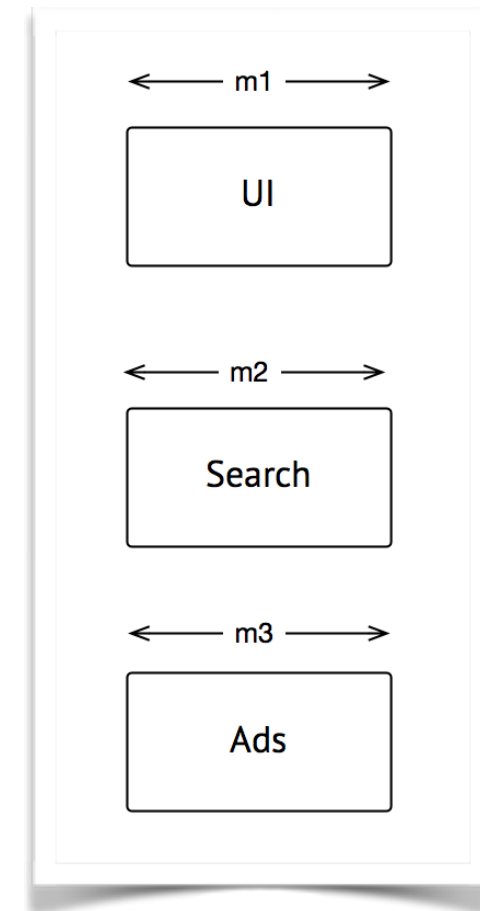
- Parameters are variables with a set of possible values
- Examples:
 - search algorithm
 - ad background color
 - autocomplete on/off
 - number of results in search page

Two extremes



Single Layer

Request belongs to at most one experiment



Multi Factorial

Request belongs to N simultaneous experiments

The cookie-mod-bucket “binaries” flow

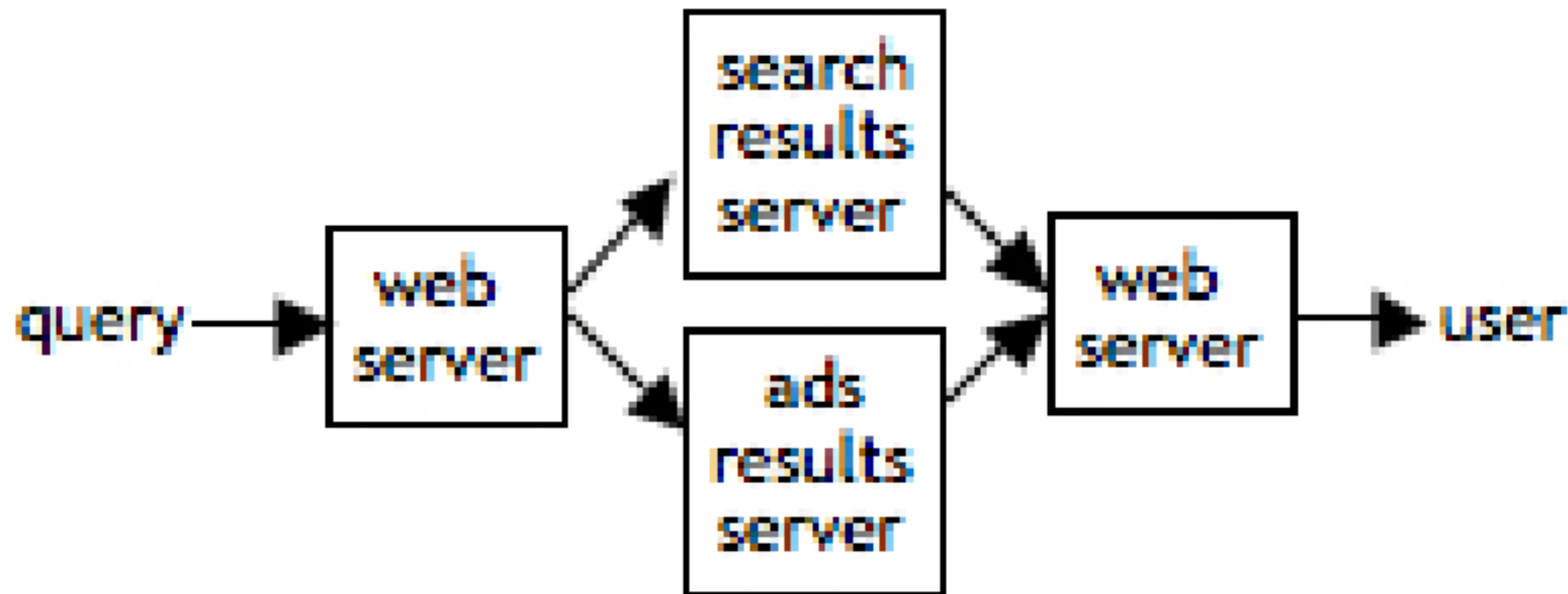
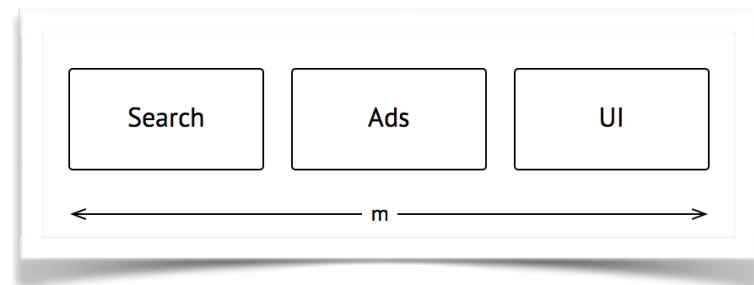


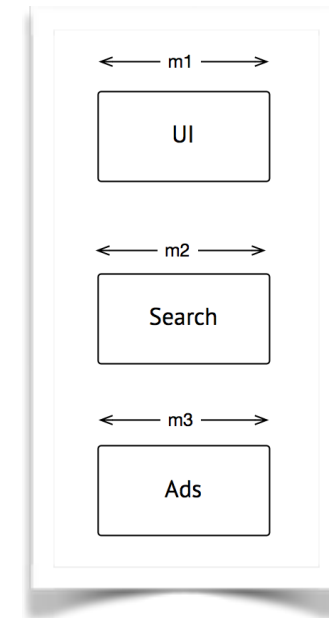
Figure 1: A sample flow of a query through multiple binaries. Information (and time) flows from left to right.

Two extremes



Traffic is split, meaning we need more time to attain statistical significance

Causes downstream binary starvation



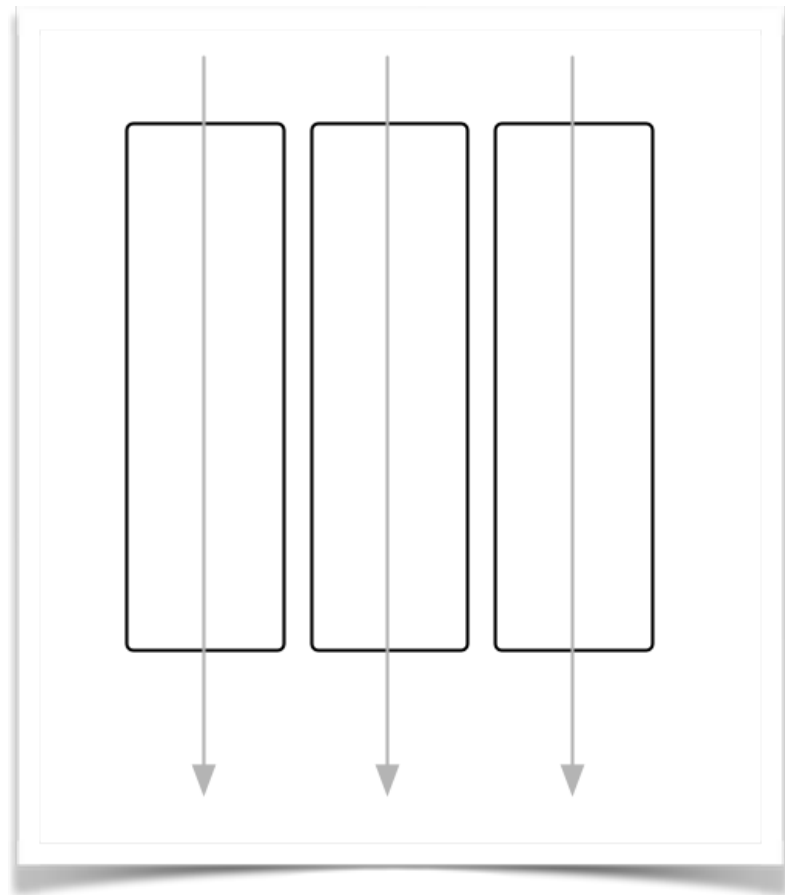
Treatments might conflict - blue text on blue background

Parameters change often, difficult to change one experiment without affecting another

Domains, Layers, and Experiments

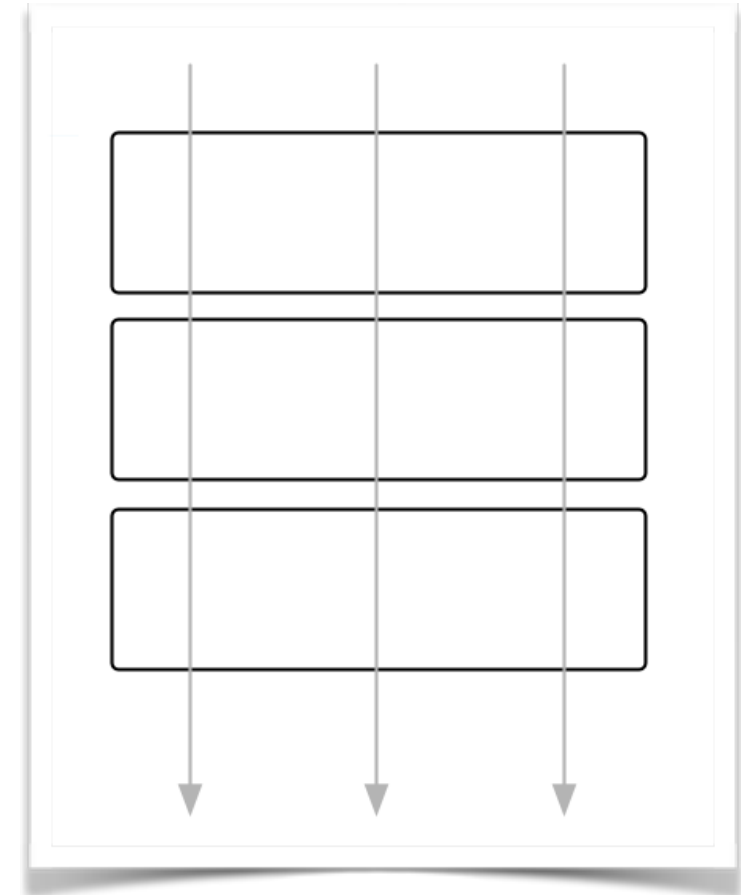
The building blocks

Domains



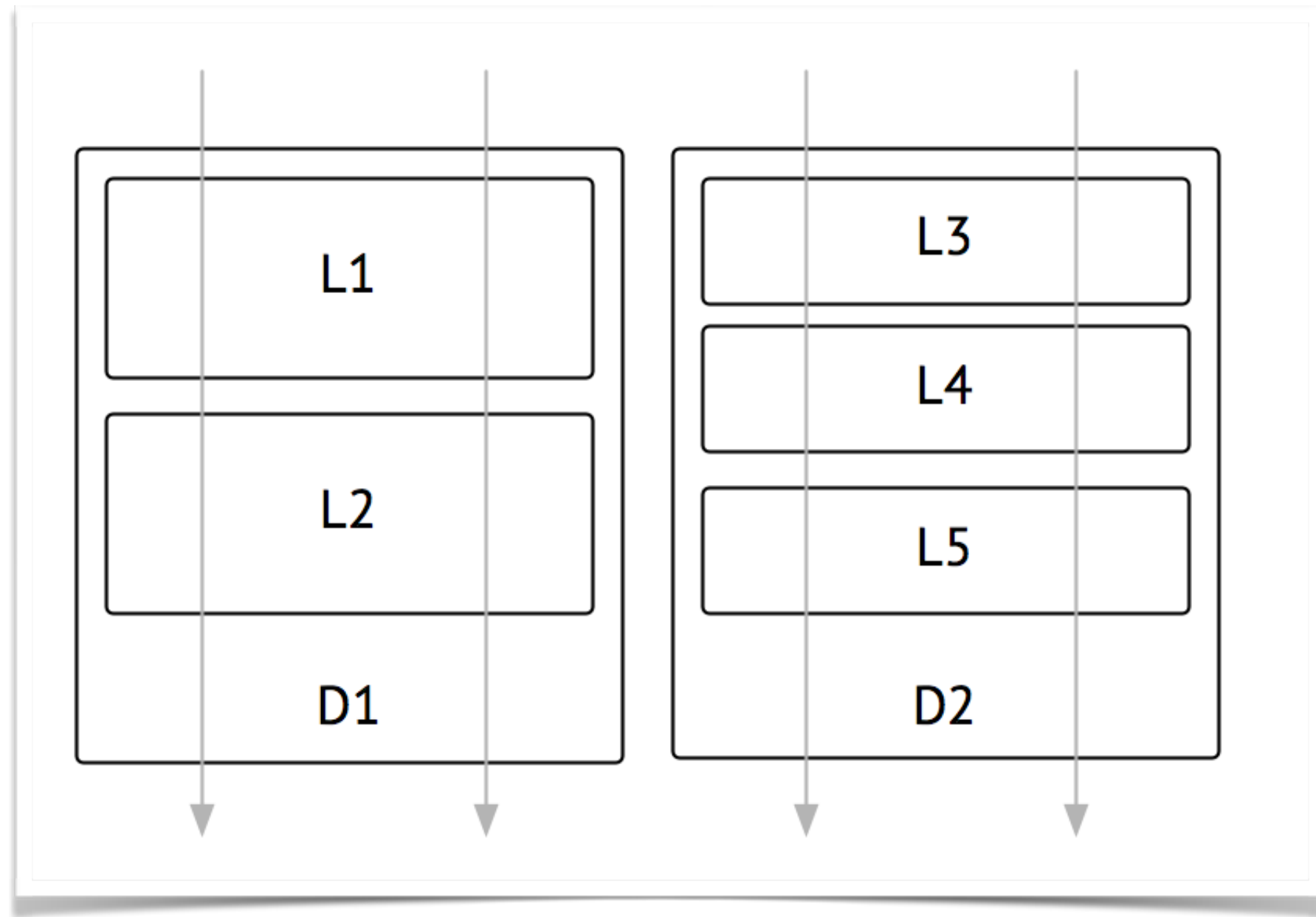
Models splitting of traffic

Layers

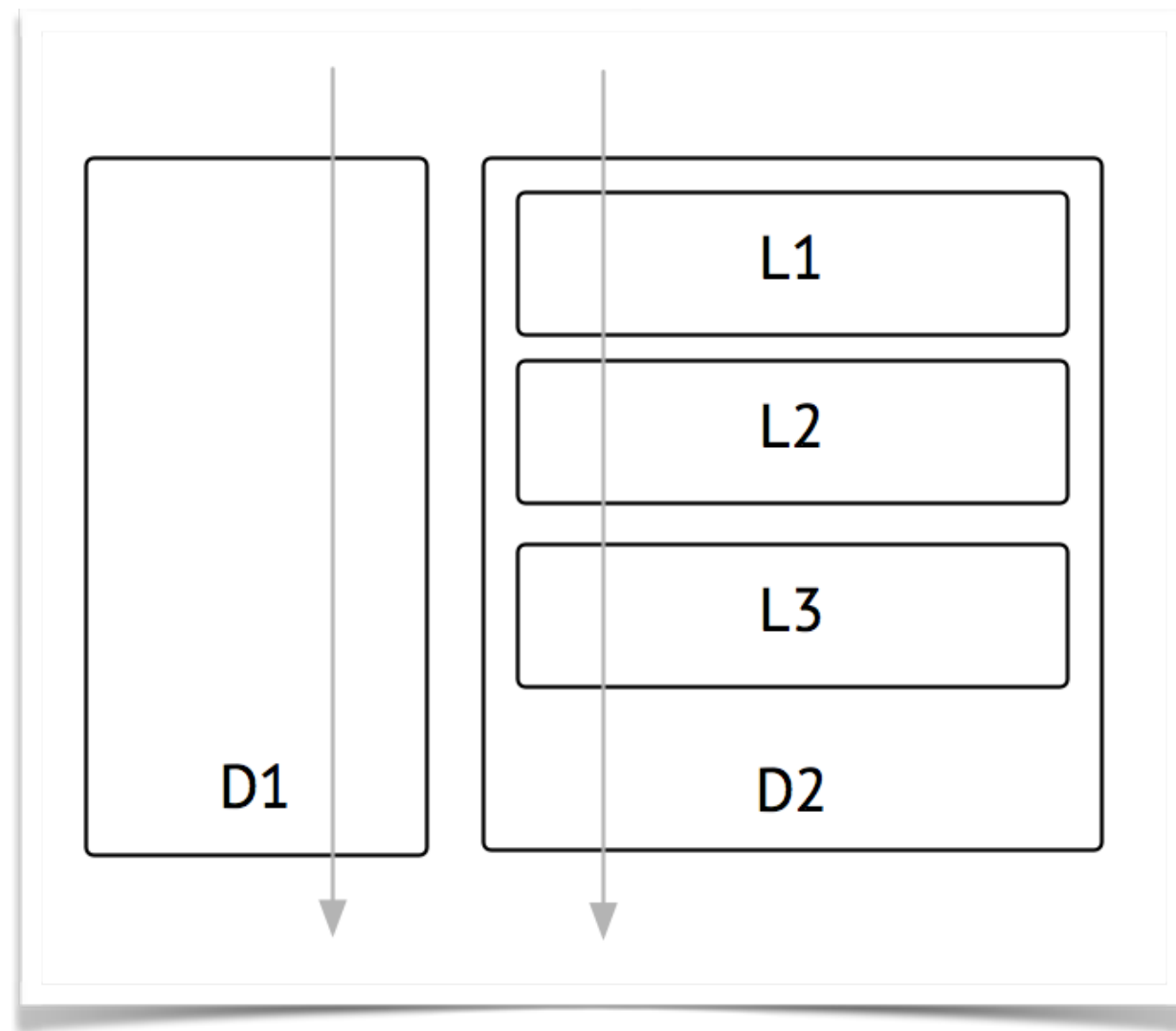


Models grouping of parameters

Domains contain Layers

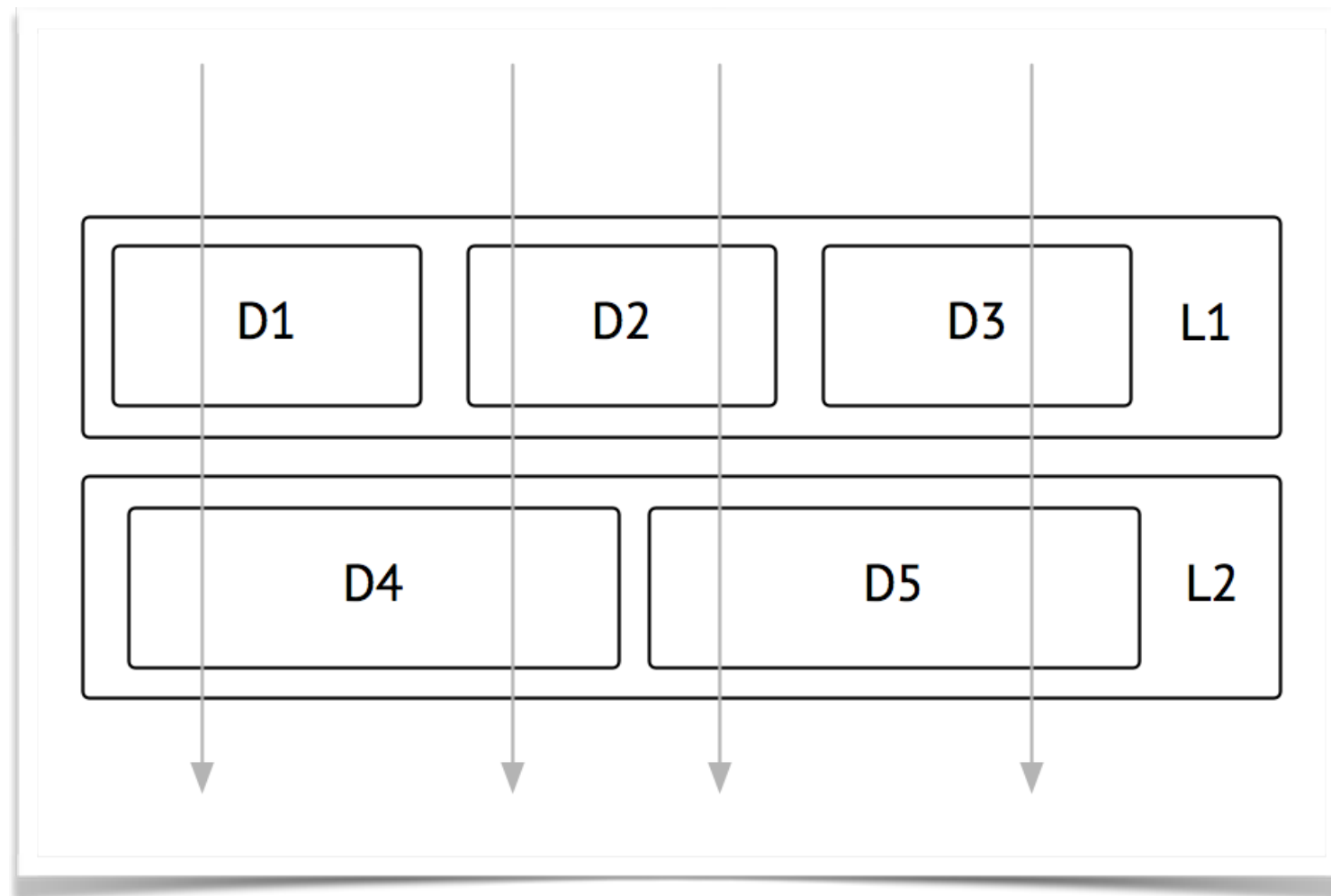


Domains contain Layers



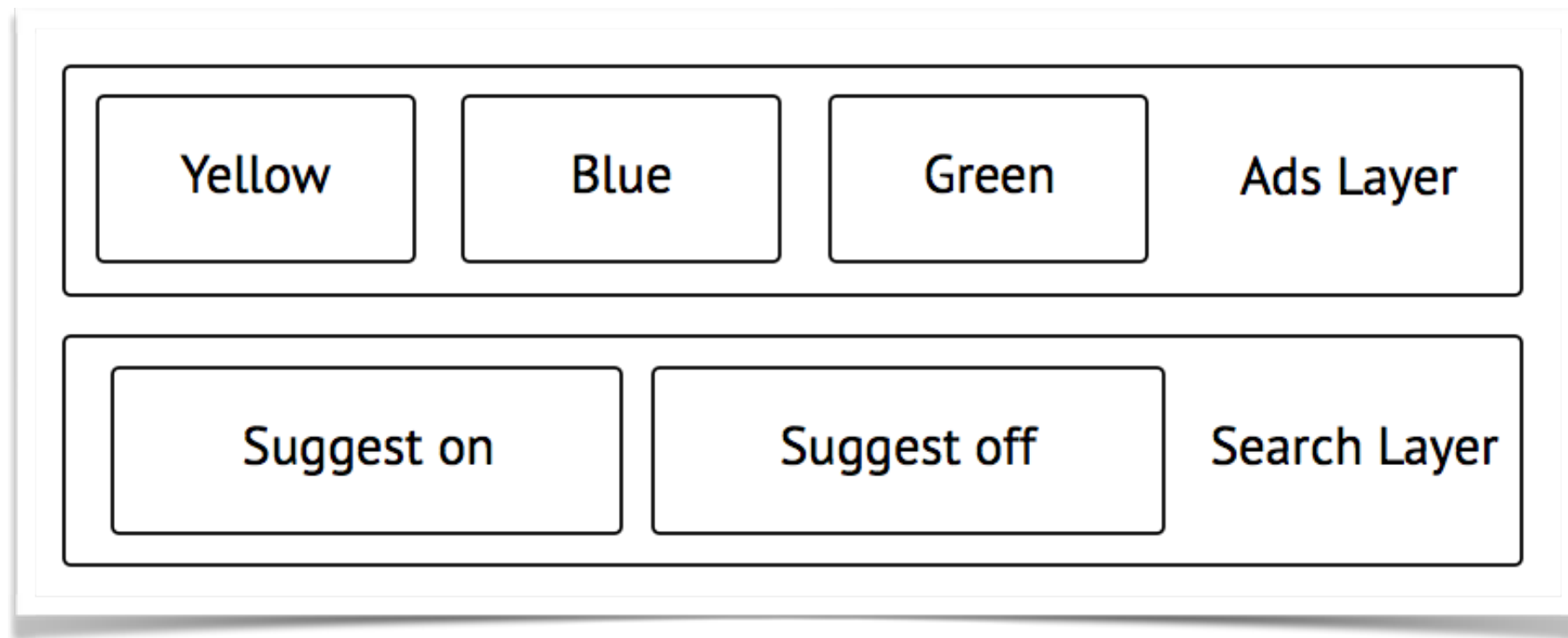
We can change a lot of parameters in D1
In D2, we change parameters more clinically,
by grouping them in layers

Layers contain Domains



This allows for further division of traffic and enables allotting an experiment only the portion of traffic that is relevant to it

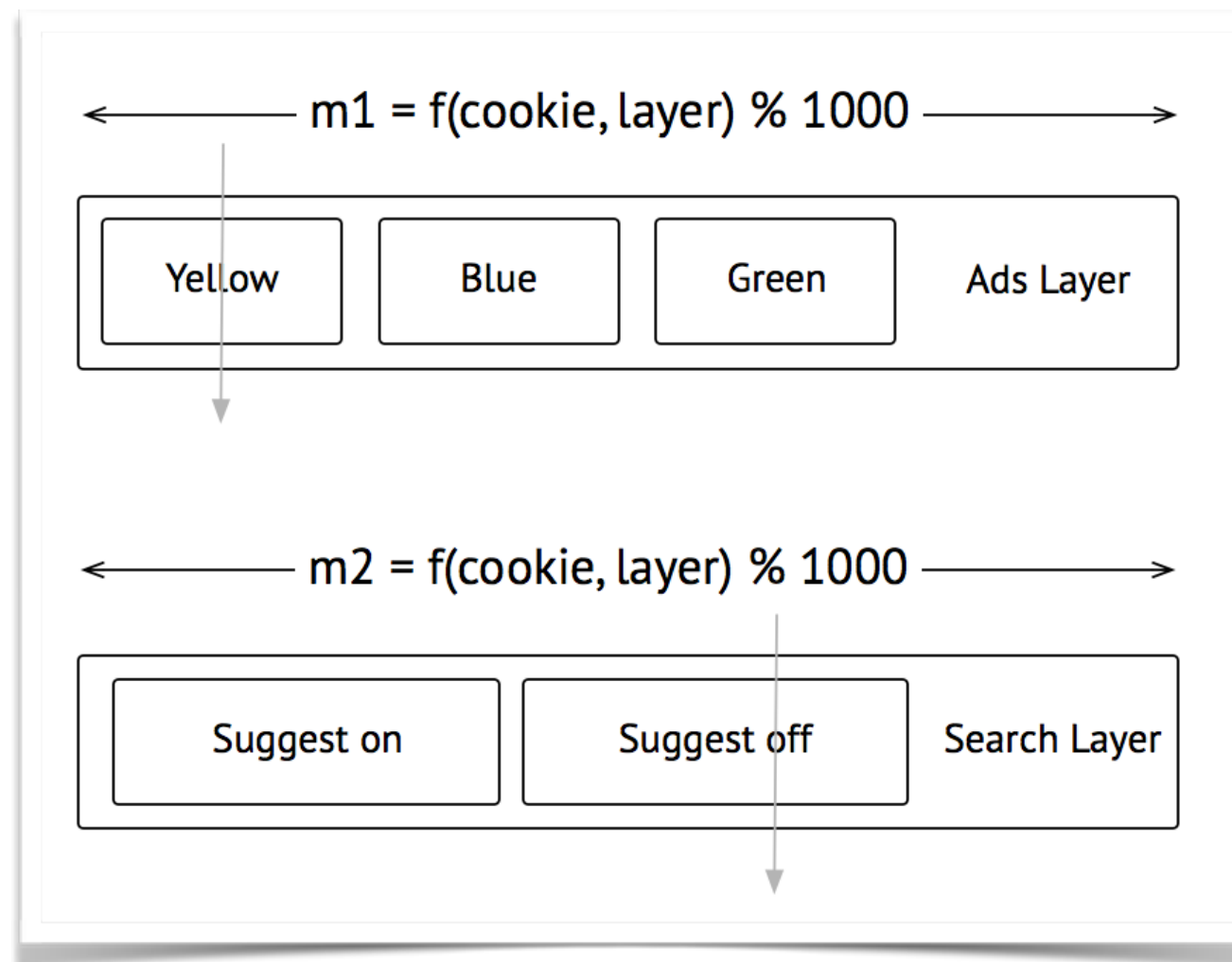
Layers contain Experiments



“Experiment” in google’s terminology is what a “Treatment” in statistical literature. Traffic is split between treatments.

Diversions and Conditions

Diversion



`f` needs to be pure for stickiness reasons
i.e, Given the same cookie and layer-id, it will result in the
same behavior every time

Diversion types

Random | roll a die

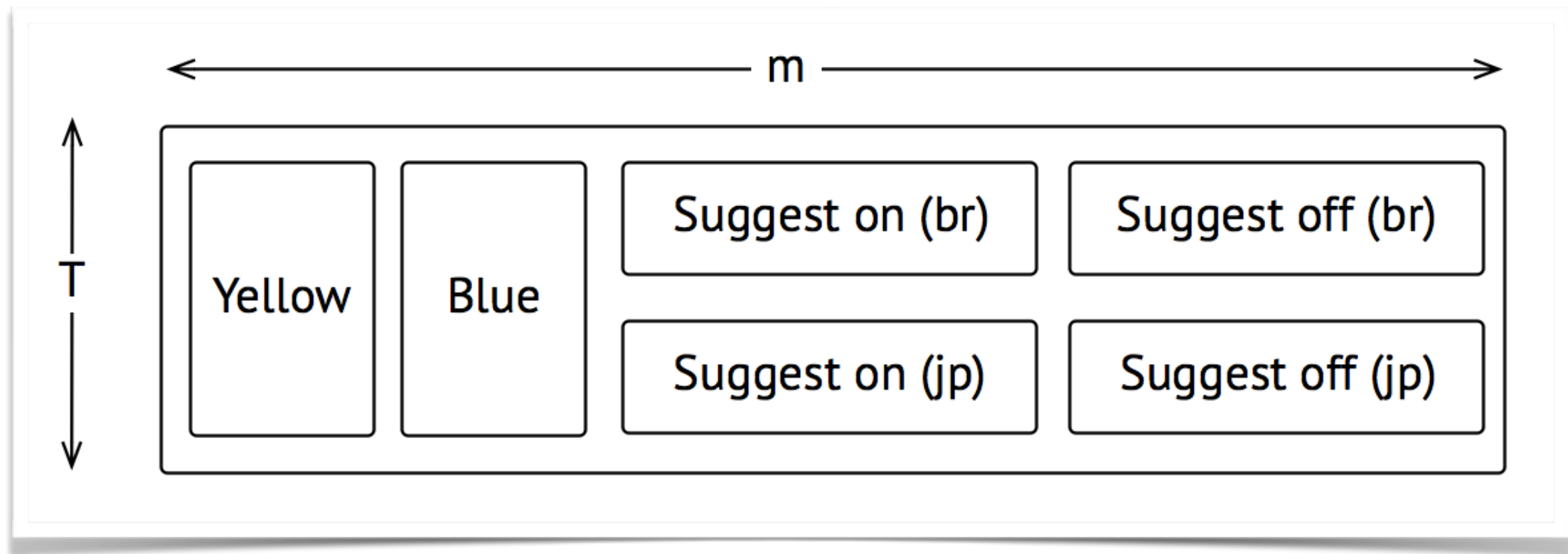
Cookie mod | cookie is random \Rightarrow cookie % 1000 is random

Cookie day | cookie + day % 1000

User ID | higher level of stickiness

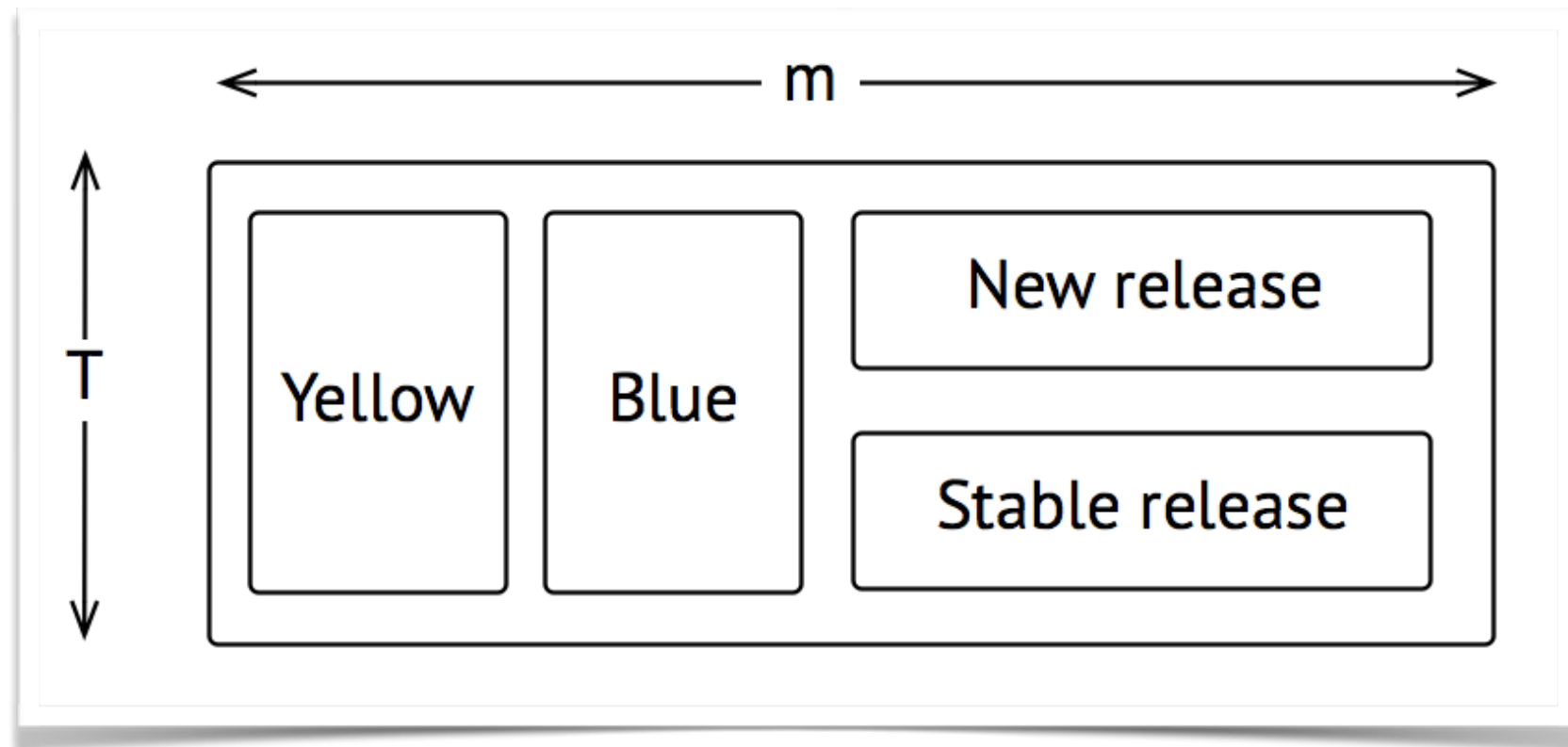
User-ID > Cookie mod > Cookie day > Random

Conditions



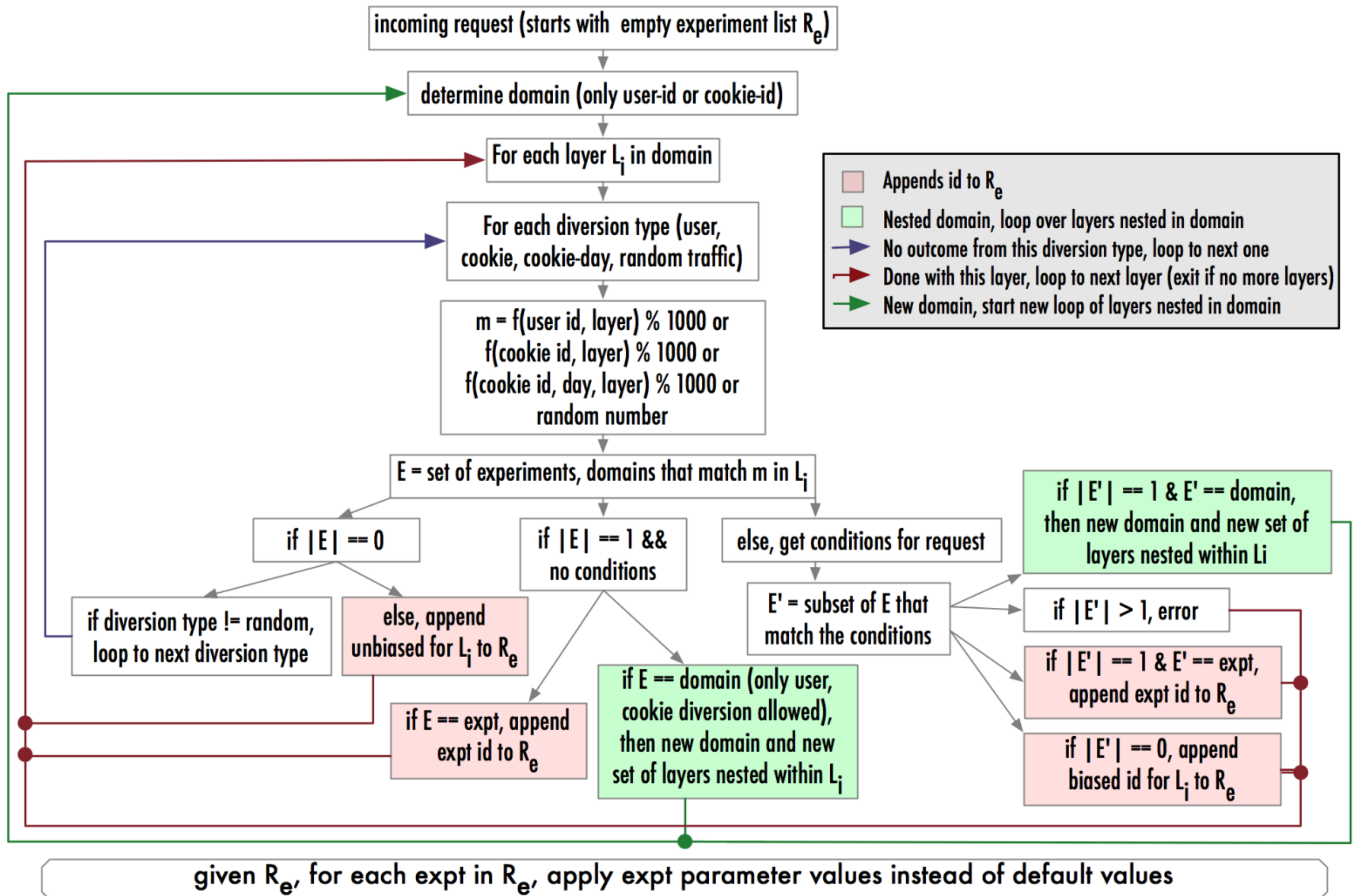
Domain based allocation after diversion

Conditions

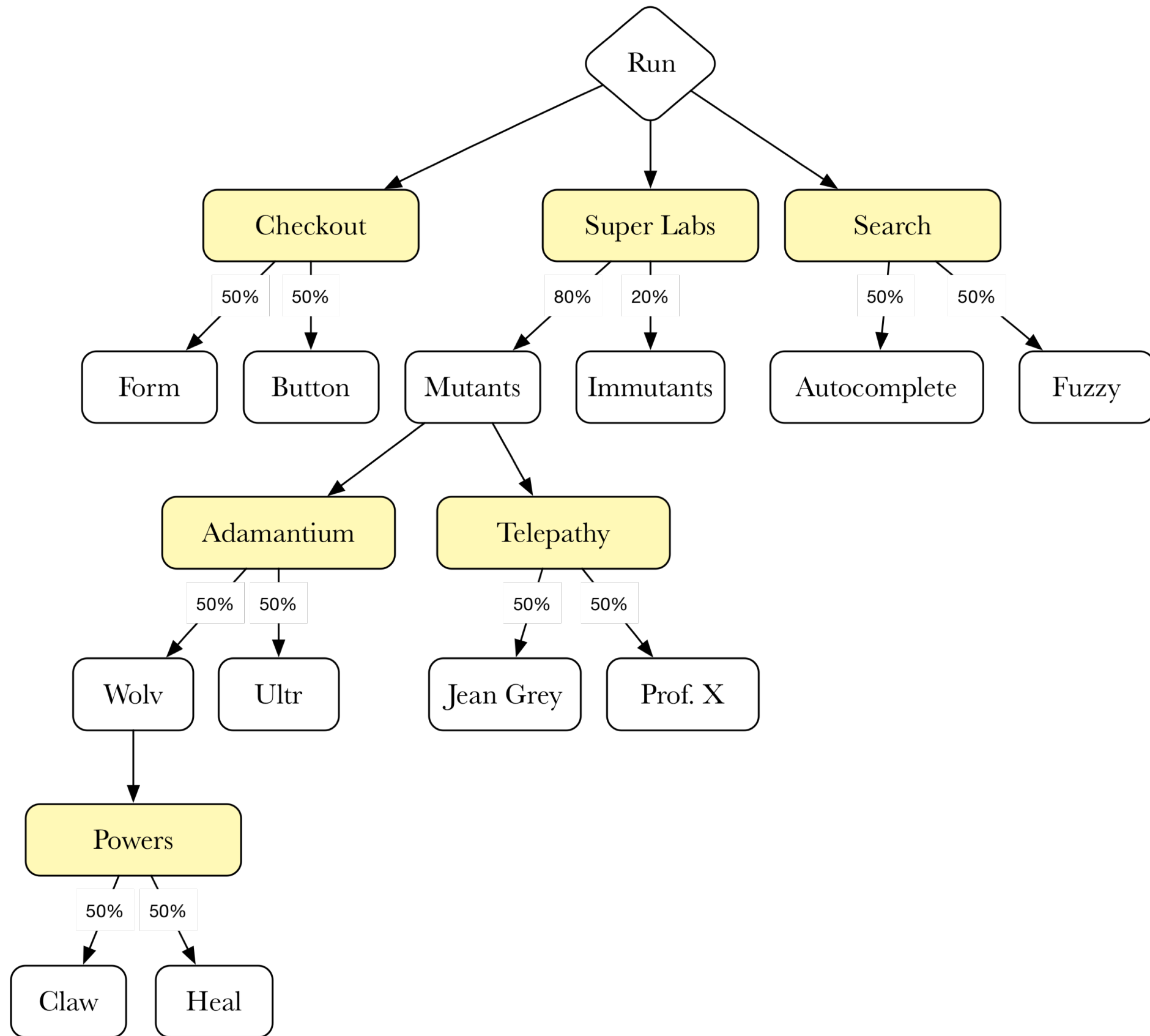


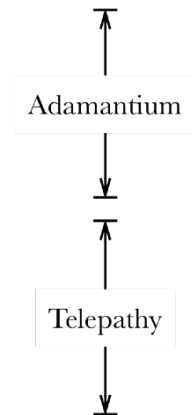
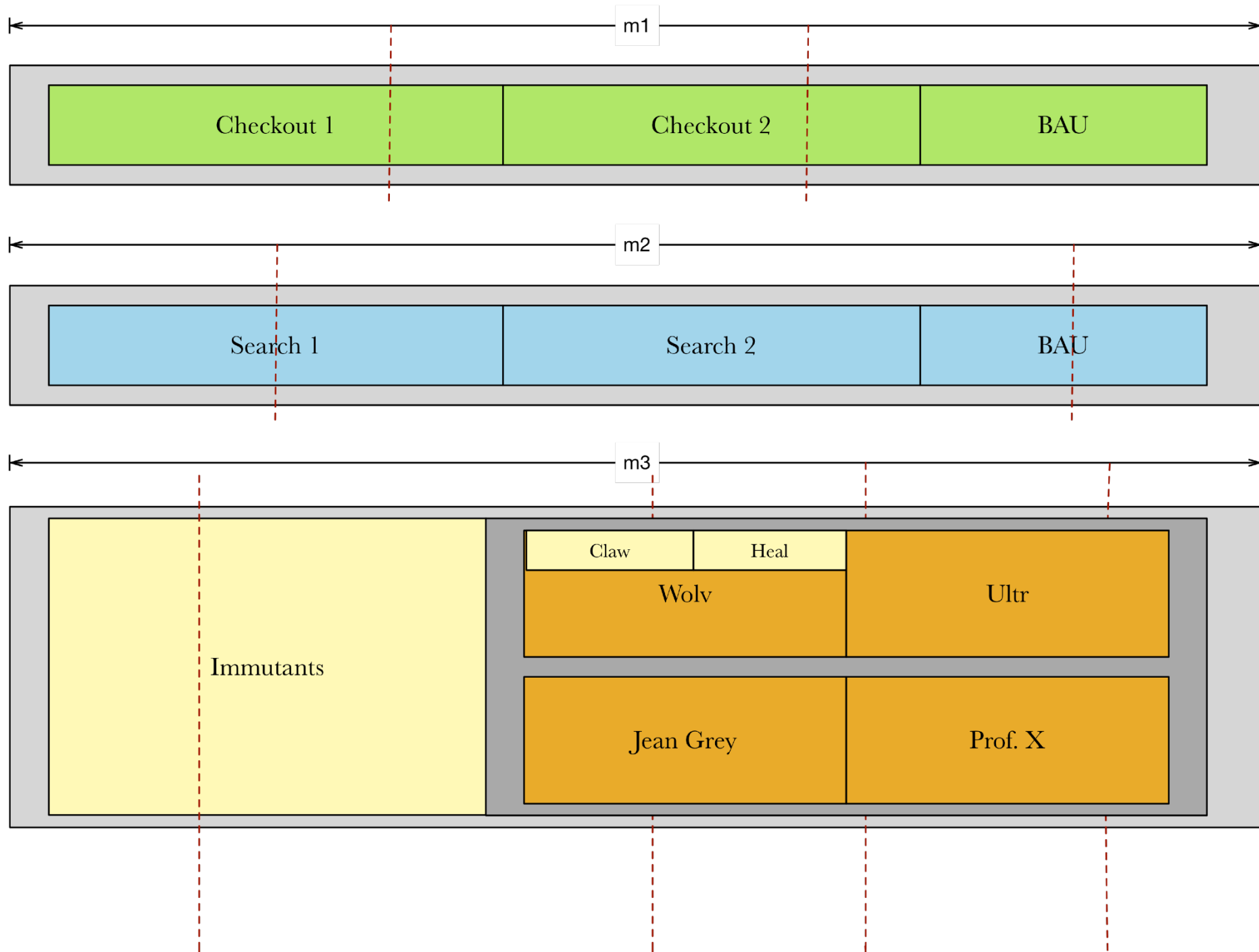
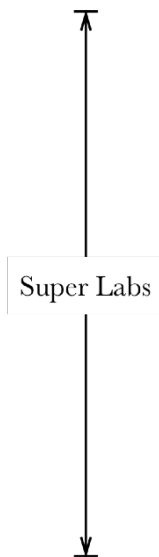
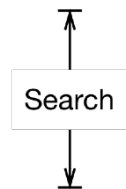
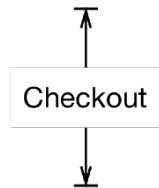
Test (“Canary”) new code with live traffic, and ramp up to a full release based on datacenters or machines

Logic flow



Alternate Approach





Other important aspects

- Education – it is easy to design an experiment wrong
- Coverage – when users are actually part of the treatment
- Scheduling – scheduling within the nested blocks
- Confidence Level, Confidence Intervals, A/A tests
- Reporting, Accuracy
- Testing

Great Material

- Overlapping Experiment Infrastructure: More, Better, Faster Experimentation (Google)
- A/B testing @ Internet Scale (LinkedIn, Bing, Google)
- Controlled experiments on the web: survey and practical guide
- D. Cox and N. Reid. The theory of the design of experiments, 2000
- Netflix Experimentation Platform
- Online Experimentation at Microsoft
- Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO (Microsoft)