

# Investigating Network Testbed Usage

Jelena Mirkovic, Hao Shi, Alefiya Hussain and Ted Faber  
USC/ISI  
{sunshine, haoshi, hussain, faber}@isi.edu

April 13, 2011

## Abstract

## 1 Introduction

These are key questions that we're seeking to answer:

1. Do testbeds help people conduct useful research?
2. What aspects of testbeds hinder their wider use?
3. Can testbed use/management policies be improved and how?

Most of the findings reported here relate to questions 1 and 3. Unfortunately testbeds today do not collect enough data nor do they collect the right data to answer the above questions conclusively so we are often forced to draw bold conclusions based on our interpretation of the existing, limited data.

We observe that there are three primary types of experimentation patterns on testbeds today: (a) **Hypothesis Validation** where the experimenter rigorously explores the parameter space to validate a particular hypothesis. (b) **Deployment Study** where the experimenter installs new technology to study its impact and/or to test it out (c) **Exploration** where the experimenter takes unknown technology and immerses it into the testbed to study it further.

In the subsequent sections, we have classified all three cases of experimentation patterns as *research* experiments. Additionally, we have *class* experiments on DETER, that are instantiated due to use of DETER in graduate and undergraduate security courses across 22 universities world-wide. Testbed is also used for internal testing – we exclude all data of such experiments and users from our statistics.

## 2 Related Work

## 3 Network Testbeds

## 4 User Diversity

Testbed use evolves after a few years to outsider-dominant

Class use grows easily

Most users are in US, as expected

## 5 Testbed Usage Patterns

**Hypothesis:** People use testbeds in multiple visits. Some experiments are small because they are exploring the space

## 5.1 Experiment and Project measures

clustering TODO

### 5.1.1 Experiment Duration

*Experiment duration* is defined as the time lapse between when an experiment is allocated testbed resources to when the experiment releases the assigned resources. Figure 1 shows the *cummulative distribution function (cdf)* of duration for our two experiment categories, research and class experiments. Since each allocation is plotted as an independent event, the same experiment (identified by its name) can generate multiple points in Figure 1 if it allocated and released resources multiple times in the course of its lifetime. Research experiment duration is heavy-tailed with 26% of experiments lasting less than 15 minutes, 51% lasting less than 1.5 h and 90% lasting less than a day. But a few experiments last more than a year. Class experiment duration is also heavy tailed but longer experiments dominate more: only less than 7% last less than 15 minutes, 35% last less than 1.5 h and 96% last less than a day. Longest class experiments last for a few weeks. We attribute this longer duration of class experiments when compared with research experiments to the fact that class experiments are usually well specified in advance by the class instructor. Class users can thus just allocate resources for the experiment and do useful work, while research users may need several trial resource allocations while they test out their set up and discover the combination that best works for their research purpose.

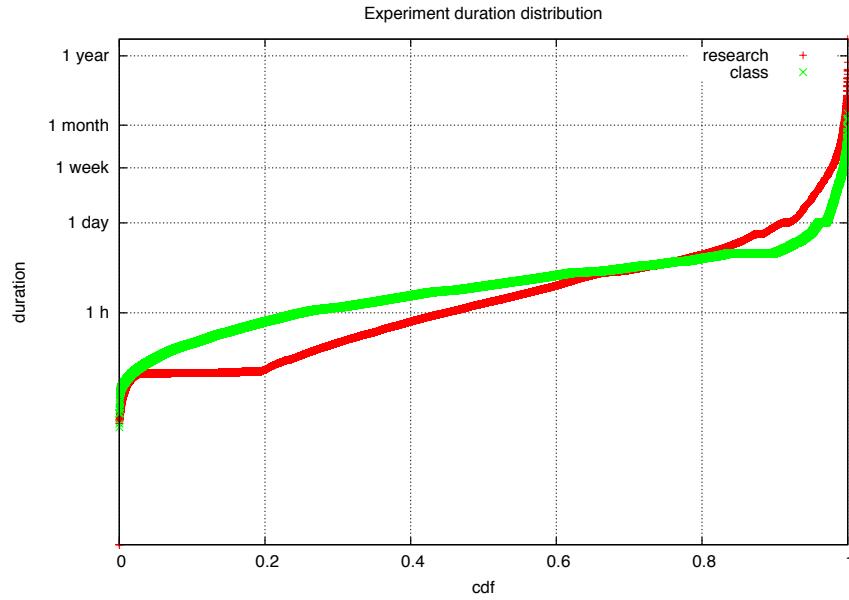


Figure 1: Experiment duration distribution

We also sought to correlate experiment duration and size, which is shown in figures 2 and ???. As it can be seen, there is no obvious correlation.

### 5.1.2 Experiment Lifetime

*Experiment Lifetime* is defined as time lapse between the first experiment creation event to the last experiment deallocation event. Thus each unique experiment is only represented by one point in Figure 4. Research

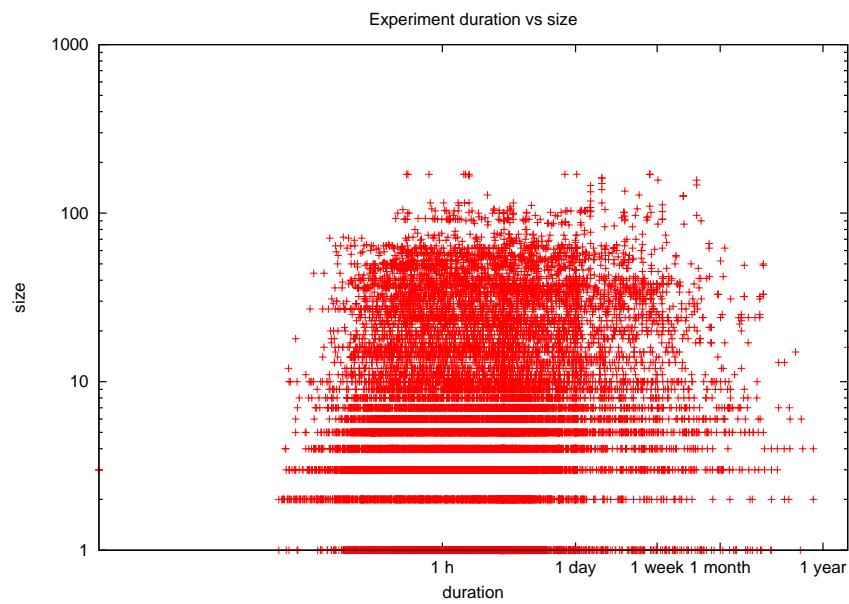


Figure 2: Research: Experiment duration vs size

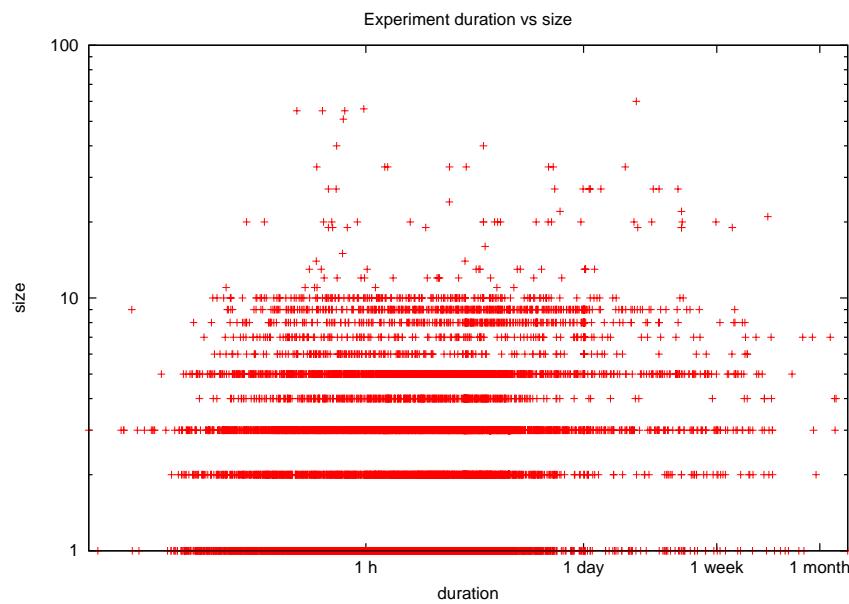


Figure 3: Class: Experiment duration vs size

experiment life is heavy-tailed with almost 51% of experiments lasting less than 10 minutes. Conversely, only 1.4% of class experiments last less than 10 minutes. To verify our hypothesis that short research experiment life is due to users trying to determine the best setting for their purpose we examined the percentage of short experiments that are followed by longer experiments in the same project. If we define "short" as lasting 10 minutes or less 3,835 out of 3,846 (or 99%) short research experiments are followed by a longer experiment in the same project. Similarly, we investigated how many short experiments are preceded by a long experiment, hypothesizing that this is due to the user perfecting their scripts and automating the experiment so that it can run under 10 minutes. This time 3,839 out of 3,846 (or 99%) short research experiments were preceded by a longer experiment. We conclude that short research experiments occur often in the middle of experimentation stream when users want to either investigate new setup or they have sufficiently automated their experiments that they can finish quickly.

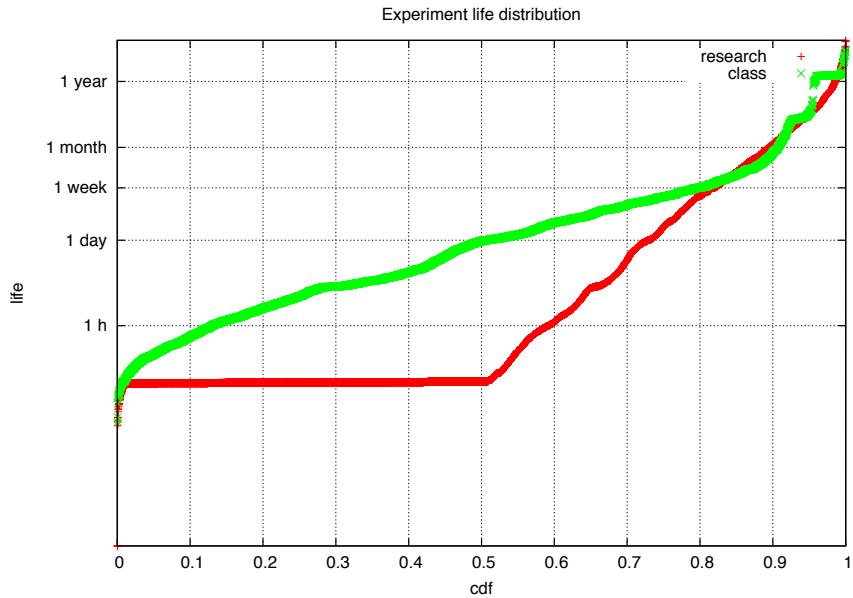


Figure 4: Experiment life distribution

To validate this theory that short experiments are somehow related to longer, more useful experiments we investigated the minimum elapsed time between a short experiment and a longer one either following it or preceding it. Figure 5 shows the *cdf* of these time intervals. About half of the short experiments are performed within an hour of a longer experiment, and 96% are performed within a day of a longer experiment. This validates our theory.

### 5.1.3 Project Size

Figure 6 shows the project size in number of unique experiments (one experiment can be allocated resources multiple times but it still accounts for only one data point) for class and research projects. We notice that most projects have a small number of experiments: 56% of research projects have less than 10 experiments and 23% of class projects have less than 10 experiments. Again this distribution is heavy-tailed with a few projects generating hundreds of experiments. Distributions of the number of allocations per project (Figure 7 exhibit are similarly heavy-tailed.

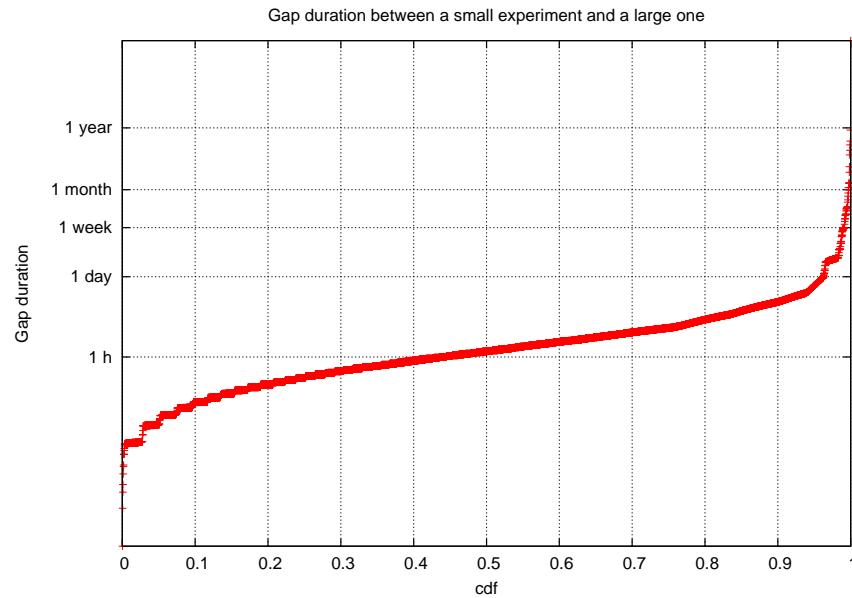


Figure 5: Gap duration between a small experiment and a large one

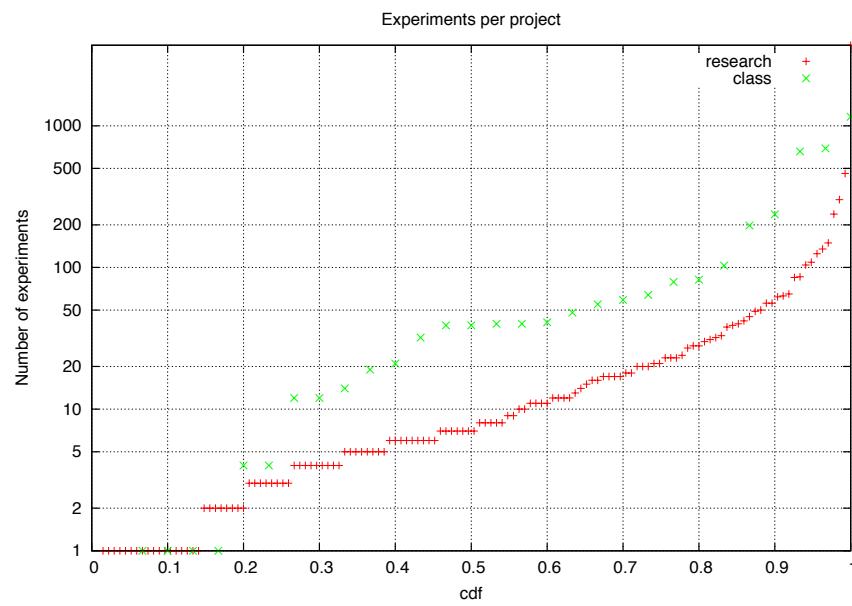


Figure 6: Number of experiments per project

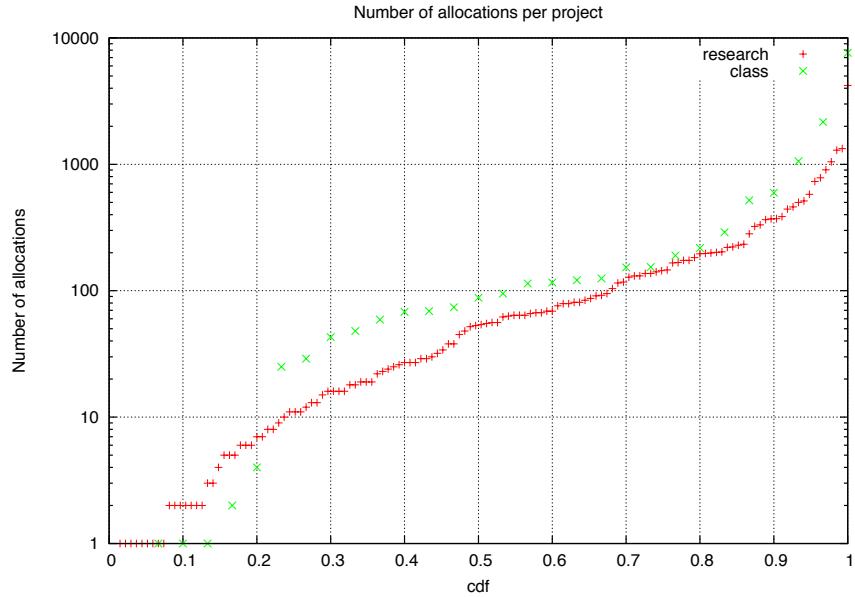


Figure 7: Number of allocations per project

#### 5.1.4 Project Lifetime

Figure 8 shows the distribution of project lifetime, from the first to the last event (experiment creation, resource allocation, resource release, modification). We notice that the smallest lifetime is 46 days, with more than half of the research projects being active for more than 2.5 years and one third of class projects being active for more than a year. Coupled with project activity data such as number of experiments and number of resource allocation, and with experiment activity data such as duration this shows that people use testbeds in multiple short visits spread over a long time period.

Additionally, we found that a significant number of projects are created but never used, that is, not a single experiment is created within these projects. The percentages are: 24% for DETER [?], 24% for Emulab [?] and 11% for Schooner-WAIL testbeds [?]. We contacted the PIs of these unused DETER projects to understand their reasons for not using the testbed and the responses we received can be broadly classified as:

1. Found that simulation or live deployment are better fit with my research,
2. Couldn't find sponsors or students for the project
3. Expected there was some specific software or hardware in the testbed, which proved wrong, and
4. Didn't really need to create experiments since the goal was simply to learn how to make a testbed on our own

We are still investigating root causes of this phenomenon but the fact that we observe these trends across very different testbeds points to the fact that testbeds need better experimentation tools to eliminate many cases under (1) and (3) categories. Having an ability to retire and resurrect projects would help better accounting by eliminating projects in (2) category and having a special project type for people that seek to build testbeds would eliminate projects in (4) category.

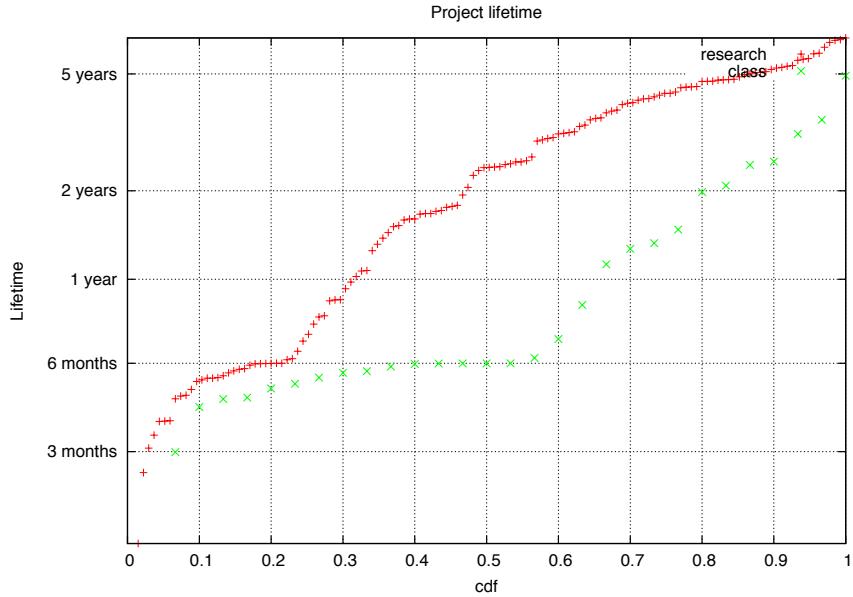


Figure 8: Project lifetime

## 5.2 User activity patterns

User measures (TODO)

### User Activity Patterns

We define an "active" user as a user that has manipulated (created, allocated, released, modified) an experiment within a project that he belongs to. Figure 9 plots the percentage of active users in a project against the number of project members for research and class projects. We notice that for small projects (<10 members for research projects and < 50 members for classes) percentage of active users varies widely. The lowest percentage of active research users is 20% while it is 3% for class users. For large projects, however, a large percentage of users is active. This effect is counterintuitive and requires further investigation.

Additionally, we found some number of users never manipulate (create, allocate or release resources, modify) an experiment . The percentages are: 10% for DETER and 15% for Schooner-WAIL. Closer investigation shows three causes of such behavior:

1. Users tend to open duplicate accounts if they forgot their password and can't retrieve it the regular way or if they change institution affiliation
2. PIs tend to create projects but do not manipulate experiments – their students/employees do
3. Students in class projects may work with an experiment already set up by the instructor or TA

Cause (1) is preventable with better account management from testbed ops. Cause (2) can be easily identified and corresponding user records taken out of the statistics. Testbeds need better accounting to detect behavior due to cause (3).

We now look deeper into these statistics. Figure 10 shows the number of unique experiments manipulated by a user. One unique experiment counts only once even if the user performed multiple runs of that

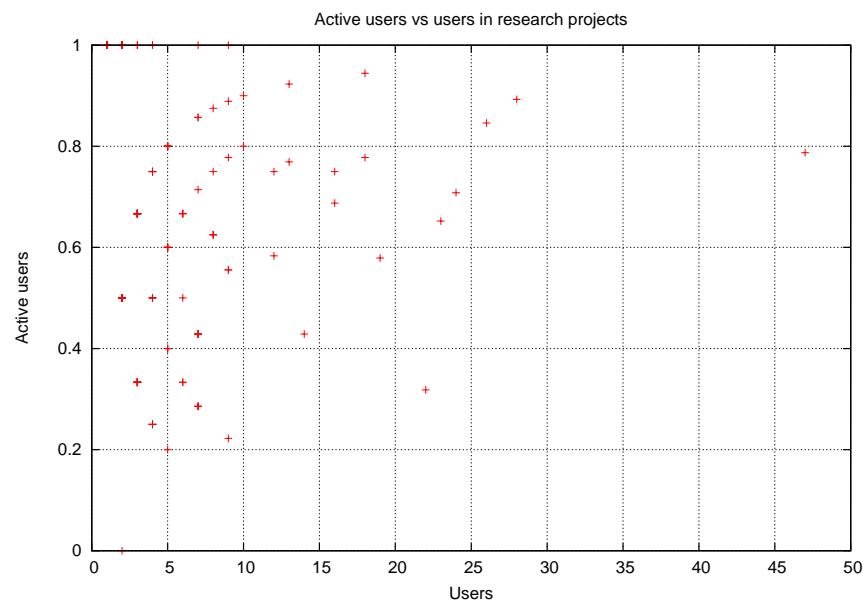
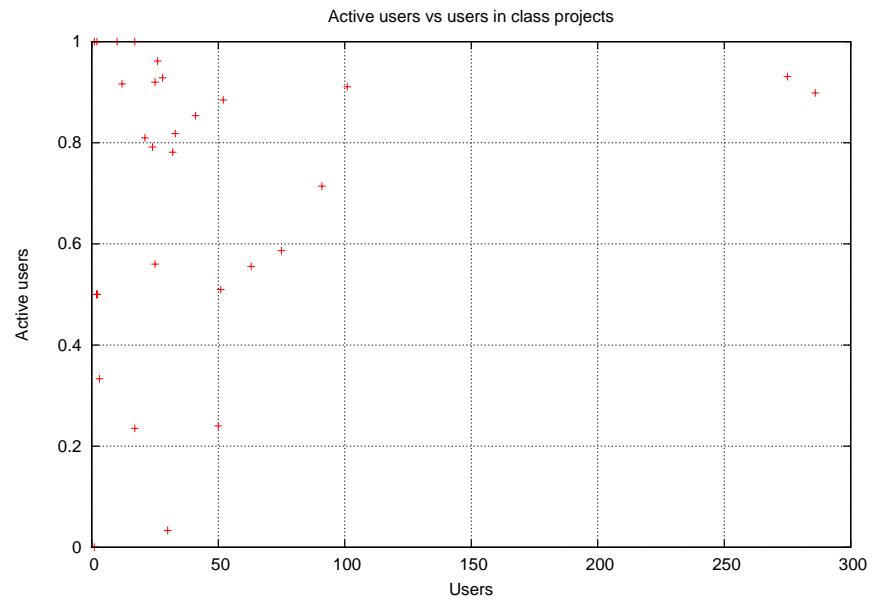


Figure 9: Percentage of active users among all members in class and research projects

experiment. Around half of research and class users manipulate less than five experiments. Around 90% of research users manipulate 20 experiments or less, and 97% manipulate 50 or less. Majority of class users (95%) manipulate 10 experiments or less.

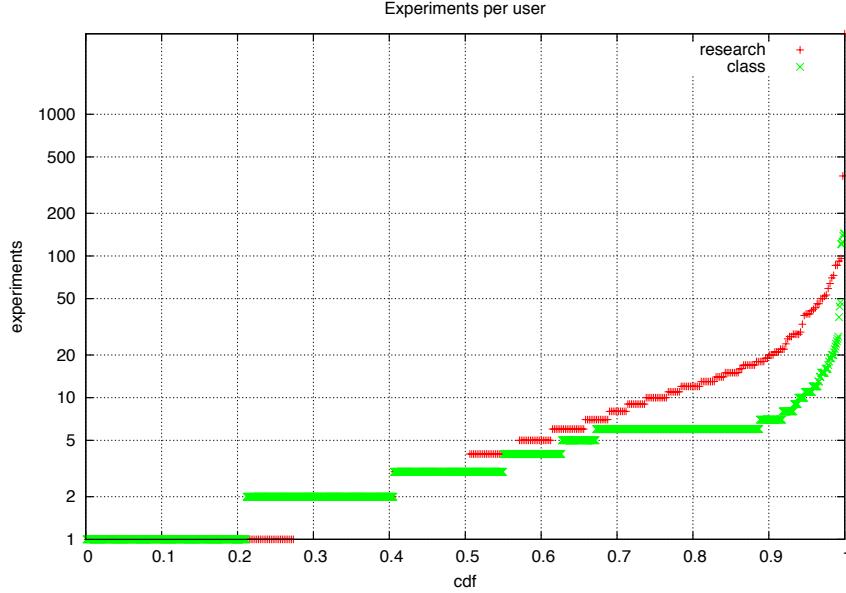


Figure 10: Experiments per user

Figure 11 shows the number of projects that a user belongs to. Around 90% of research users and 97% of class users belong to only one project. **Check if multiple-project users are in fact PIs**

Figure 12 shows the distribution of a user’s manipulation lifetime, i.e. it shows the elapsed time between the first and the last experiment manipulation by the user. Around 22% of class users and 35% of research users never manipulate an experiment, which is indicated by a lifetime on our x-axis. Half of research users and 35% of class users are active for a week or less. Around 82% of research users and almost all class users are active for a year or less. **It’s surprising that class users are longer active than research users – investigate this further**

We further define “active lifetime” of a user as the sum of all time when an experiment swapped in by this user was holding testbed resources. Figure ?? shows the distribution of this time for research and class experiments. 60% of class and research users are active for under one day. 90% of class users are active for under two weeks, and 90% of research users are active for under two months.

### 5.3 Idleness

**Explain that we want to investigate how well are resources used and if we could have a better idle-swap policy to serve more users. Explain how we calculate idle time and how we gather statistics.**

These are statistics from September 13, 2010 to April 1, 2011. 54% of timeslots had a free node that was not allocated to an experiment. 26% of timeslots had an idle node. 12% had a node with only a network activity, 3% had a node with only a CPU activity and 1% had a node with network and CPU activity.

We next seek to understand the cause for idleness and look for two different patterns: (1) instances where some nodes in an experiment were never used and (2) instances where all nodes in an experiment were idle for some time. We now treat class and research projects together.

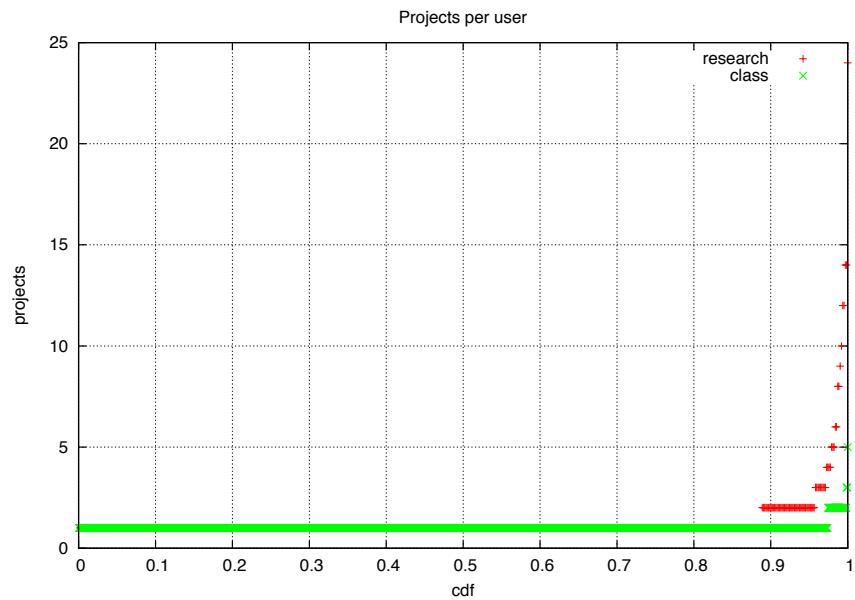


Figure 11: Projects per user

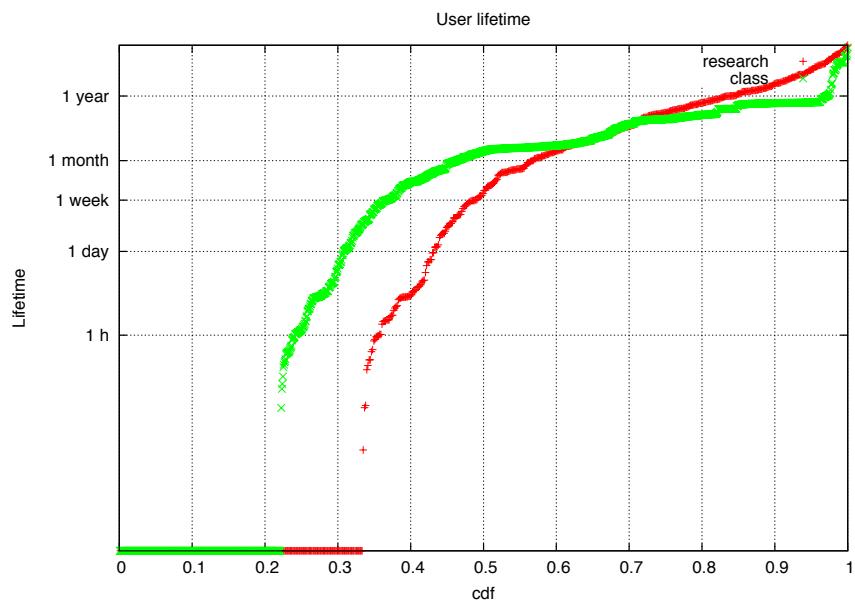


Figure 12: User manipulation lifetime

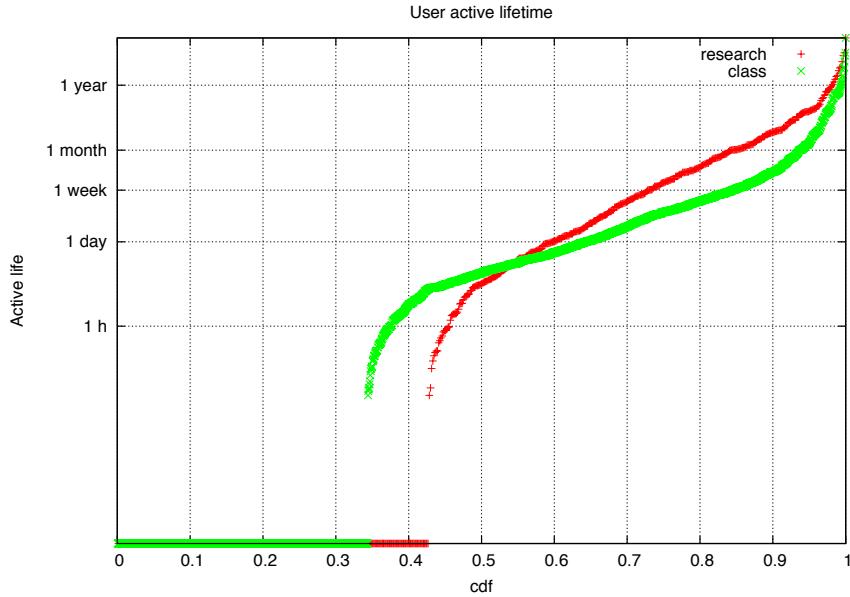


Figure 13: User active lifetime

Only 0.5% of experiment instances had one or more idle nodes. This means that users can usually estimate well what size of network they need for experimentation.

Figure 13 shows the distribution of the percentage of idle time in an experiment. A little more than 40% of experiments are never idle. The idleness distribution then changes almost linearly until it reaches 100%.

We further investigate the duration of idle gaps seeking to understand if testbeds could be better utilized if idle experiments were auto-swapped out after only a brief idle time and swapped back in later. Figure 14 shows the longest idle interval for each experiment, where the number of idle timeslots was positive. Around 65% of experiments are idle less than 1h, and 96% of them are idle less than 4h. This means that the idle-swap time set at 4h is a good estimate.

Figure 15 shows the duration of the idle interval before the experiment is swapped out. About a quarter of experiments are idle less than 1h before being swapped out. There is an almost flat line near the 1h mark going from 25% to near 50% of experiments. These are mostly class experiments for which the idle-swap time is automatically set to 1h. A similar flat line is visible around 4h from 62% to 96%. These are research experiments for which the idle-swap time is automatically set to 4h. Again these figures mean that our settings are correct.

Failed swapin data (TODO)

Planetlab (TODO)

Starbed (TODO)

## 6 Experiment Characteristics

Topology (TODO)

OS (TODO)

Link characteristics (TODO)

SEER (TODO)

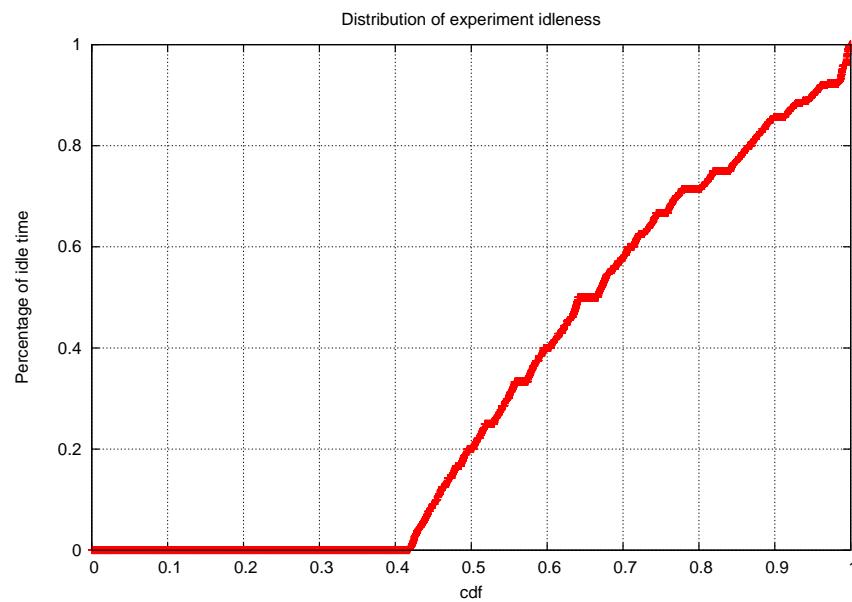


Figure 14: Distribution of experiment idleness

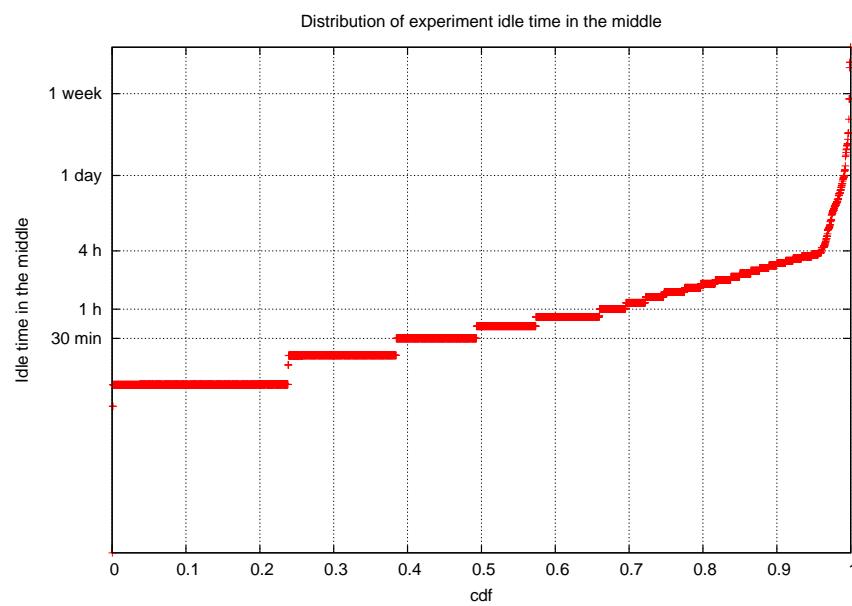


Figure 15: Distribution of experiment idle time in the middle

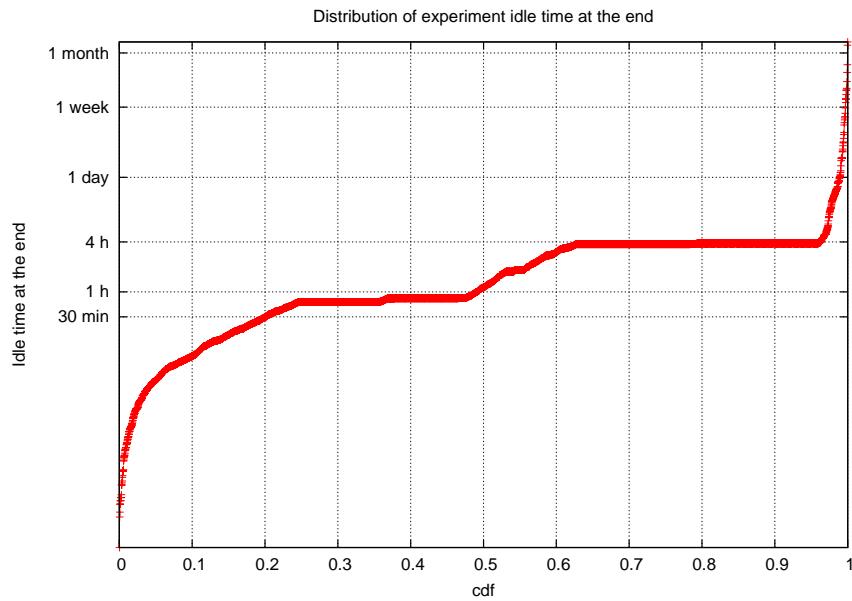


Figure 16: Distribution of experiment idle time at the end

## 7 What hinders testbed use

Survey results (TODO)

## 8 Implications for the Future

## 9 Conclusions