A Comparative Study of Network Testbed Usage Characteristics

Jelena Mirkovic, Hao Shi, Alefiya Hussain
USC/Information Sciences Institute
mirkovic,haoshi,hussain@isi.edu

ABSTRACT

1. INTRODUCTION

2. RELATED WORK

3. BACKGROUND

We define a *network testbed* as any collection of physical or virtual machines that is shared by multiple users with the goal to support research or education activities. This is a broad definition that encompasses Emulab-like testbeds such as Emulab [**?**] or DETER [**?**] where users gain exclusive, superuser access to physical nodes, VM-based testbeds such as Planetlab [**?**] where multiple users share access to a physical node and private testbeds in research labs and companies. The goal of this paper is to understand how network testbeds are used and to identify places where improvements are needed.

Discuss how testbeds differ and their modes of use. Discuss project, slice, experiment, etc. Say experiments and projects have descriptions. We will show mapping as we go and in a table.

Say how we lack info about what users do with machines.

3.1 Testbed Uses

Two main testbed use modes we investigate in this paper are for research and for classes. In research use model a group of users investigates a common research problem and uses testbed to evaluate the problem or their solutions. The users could be from academia (faculty and students), a government lab (researchers, staff) or industry (company employees). Their final goal may be to produce research papers, white papers, to change government policies or to test a product. This variety of research users makes it difficult for testbed owners to evaluate the productivity of their users, ultimately quantifying how useful testbeds are for research, because some research outcomes may not be public. In classes teachers use testbeds to either illustrate a concept taught in class or to assign a practical project to students. This project may be of research nature, especially for upper division graduate classes, which blurs the line between class and research.

3.2 Terminology

An **experiment definition** is the input submitted to the testbed (one or more times) describing resource needs and setup operations. The experiment definition is closely tied to *one particular purpose*, e.g. a research question or a class project. It is uniquely identified by its name and can be modified while still keeping its identity.

An **experiment instance** denotes the combination of experiment definition and testbed resources allocated to it. Return of the resources to the testbed denotes the end of this instance. On the other hand, a modification of the experiment definition which requires release and reallocation of resources (e.g. to change experiment topology or add more nodes) still belongs to the same instance. Thus an experiment instance may be associated with more than one resource set.

A **project** is a collection of experiment definitions and people, under a single Principal Investigator (PI) aiming to investigate one specific research direction or taking one specific class.

In our investigation of testbed use patterns we have discovered a broad range of activity levels in every dimension. Some projects and some users were very active, while others generated none or a few experiment instances. Some experiment instances were long-lived (months) while others were extremely short (under 10 minutes). From pure use pattern data it was impossible to understand the causes of such broad range of uses. To disambiguate and possibly explain this data we have attempted to somehow quantify from project descriptions, experiment definition descriptions, personal conversations with the PIs and publications co-authored by project members how useful the testbed was for each given project. We could do so only for DETER data as the data we have from other testbeds is anonymized due to privacy concerns.

The next several definitions apply to any testbed.

An **experiment manipulation** is any interaction between the user and the testbed's control server. It usually results in allocation or deallocation of testbed resources, either physical resources or database entries.

An **active project** is a project associated with at least one experiment manipulation. Similarly an **active user** is a user that has manipulated at least one experiment. The rest of the projects and users are labeled **inactive**. Further, users that do not belong to any project are labeled **orphan users**. While it is possible for an inactive but non-orphan user to still do some useful work with the testbed, e.g., by logging into an experiment instance allocated by another member of its project, orphan users cannot do any useful work on the testbed.

We now introduce a few more definitions so we could attempt to separate those projects and users where we see little associated activity due to their short presence on the testbed from those where low activity is related to poor match of the testbed with their research needs

A **warm-up time** is the time elapsed between a project's creation and the first experiment manipulation.

testbed for less than a year. Similarly an **early user** is a user that has been present on the testbed (has an account) for less than a year. We have selected the threshold value of a year by observing the distribution of warm up time for *those research projects that resulted in measurable outcome acknowledging DETER use (aka "outcome projects", see below)*. This distribution is shown in Figure **??**. The highest warmup time we see is a little below 8 months. To be conservative we select 1 year as the threshold for the project to become active.

active. Similarly a **stale user** is a user that is neither early nor active.

We further divide active projects on DETER into several categories. As mentioned above, due to privacy concerns, we cannot perform similar classification for projects on other testbeds.

1. **Internal projects** are those projects created for monitoring and development of DETER testbed. **Internal users** are those that are members of any internal project. While many of those users also lead research projects on DETER we have found through manual investigation that PIs have hard time separating their activities into multiple projects, i.e. some internal projects may be used to do research about topics unrelated to the DETER testbed and vice versa. To avoid bias in our data, since internal users have vested interest in the testbed and are likely to be very active, we exclude both internal projects and internal users from all DETER statistics. Emulab as well. We do not exclude them for project size/activity calculation

2. **Outcome projects** are those projects where we can clearly attribute some measurable outcome of the project to its use of DETER. We classify research projects as outcome projects if they have produced at least one peer-reviewed publication (this includes MS and PhD thesis), which acknowledges use of the DETER testbed. We classify class projects as outcome projects if they have more than three members (one PI and two TAs), which indicates that students taking the class have used DETER for their class work. There is one project that has 8 users and we classified it as no-outcome because we know these are PIs exchanging materials. We selected the threshold value of three members empirically. The rest of the projects are labeled as **no-outcome projects**.

3. **Try-and-leave projects** are those no-outcome projects that have created a small number of experiment instances, followed by at least a year of inactivity and where descriptions of experiment instances suggest attempts to learn about the testbed (e.g. "learning how to use DETER","trying out SEER"). We believe that this use pattern indicates that the testbed did not match the users' needs.

4. **Hard-to-tell projects** are those that do not fall into either of the above categories. That is, they exhibit usage patterns that are suggestive of performing research of developing class materials with DETER but they have not yet generated a measurable outcome. In case of research projects there are many reasons for this effect. First, some research may take years to mature to publication, i.e. it may just be too early to tell. Second, the project may be industrial or government project and thus we cannot expect it to produce a peer-reviewed publication output. Since we know project and user affiliations on DETER we can identify such projects (and we do later in text). Third, some use of the testbed may generate negative results, e.g., a user believed he could test some hypothesis on the testbed but concluded otherwise.

4. TESTBED DATA

Authors are closely involved with the DETER testbed [**?**] and a lot of data presented in this paper is derived from monitoring DETER usage. As such some of our conclusions may apply just to the DETER testbed or to Emulab-like testbeds. Where possible we supplement our study with data from other testbeds, specifically Emulab, Schooner, Planetlab and Starbed. This data is not entirely compatible with data we have from DETER for three reasons:

1. Due to privacy reasons other testbeds could not share with us data identifying their users and projects.

2. Use model of Planetlab and Starbed is much different than use model of Emulab-like testbeds so it is difficult to establish a mapping between units of work between different testbeds.

In the text we clearly identify which data we used for our investigation and where and how our conclusions apply.

Here is all the data we have:

1. DETER: Data about each user, their affiliation and experiment manipulation activity. Data about project topics, project membership and experiment manipulation activity. Experiment durations. We have the preceding data from 2004 till today. Experiment topologies - only current snapshot for experiments that are not terminated. Machine allocations to experiments and activity (coarse-grained) for the past 6 months only. Data about publications that we can link back to projects.

2. Emulab: Publicly available data about project topics and project activity (coarse grained: some vs none). Data about user and project activity (anonymized) and experiment sizes and durations. From 2002 till today. affiliation and experiment manipulation activity (coarse grained, no data about experiments). Data about project topics, project membership and experiment manipulation activity (coarse grained, no data about experiments).

3. Planetlab: Data about allocations and resource usage for each project.

4. Starbed: User resource reservations.

4.1 Privacy and Anonymization

Emulab data is anonymized, we know class and research classification

PlanetLab data is anonymized, we know only sliceid

4.2 Cleaning the data

Table 4.2 shows the breakdown of projects and users per categories introduced in section 3.2 for DETER and Emulab data.

Look into inactive projects

Look into orphan users

Look into inactive users

Explain why inactivity happens. Check how inactive projects distribute over categories.

Explain how we account for class users vs Emulab (we recycle uids)

Now explain why we wanted but couldn't classify inactive projects into early and stale and show the graphs below.

5. USAGE CHARACTERISTICS

5.1 Experiment Distributions

If we use the maximum value of the warmup time for the threshold then the starred rows in the table 4.2 apply.

We now look at number of projects per research category. This is shown in table 2. We took DDoS, worm and botnet out of attack category because they are popular topics in recent years.

Say exps from outcome projects follow same trends as all so we don't show them here.

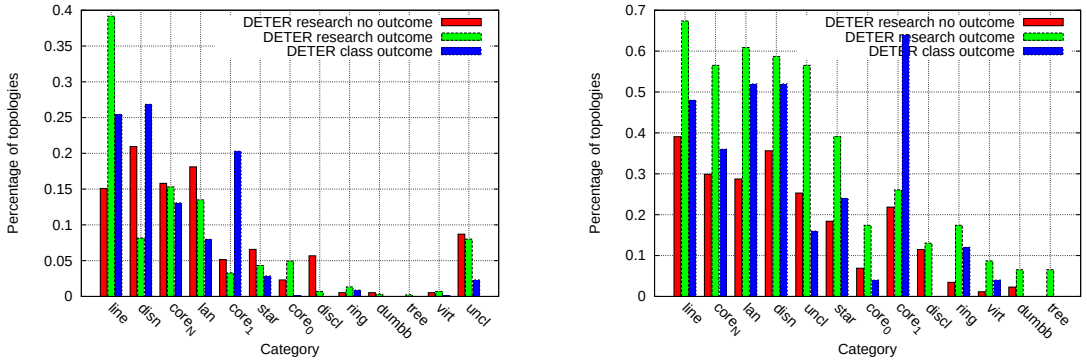

Figure 13: Experiment topologies in DETER: experiment vs project distribution

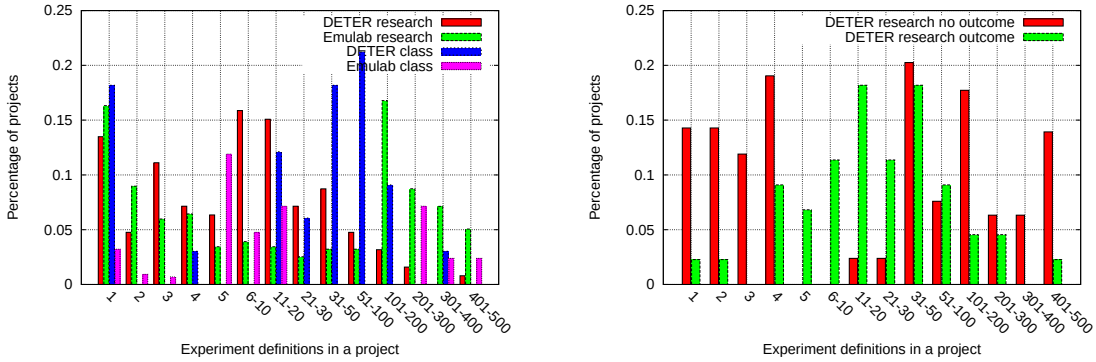

Figure 14: Experiments per project. Left: DETER vs Emulab, Right: All vs outcome

Projects	DETER	Emulab
Total	234	736
Active	179 (76% of T)	534 (76% of T)
Active no alloc	2	15
Inactive	55	202
Early∗	30 (55% of I)	150 (74% of I)
Stale∗	26	72
Internal and active	12 (7% of A)	25 (5% of A)
Working set	167 (93% of A)	509 (95% of A)
Class	34 (20% of WS)	42 (8% of WS)
Outcome (class)	25 (74% of C)	35 (83% of C)
Research	133 (80% of WS)	467 (92% of WS)
Outcome (research)	46 (35% of R)	
Users	DETER	Emulab
Total	2,345	3,607
Orphan	245 (10% of T)	210 (6% of T)
Non-orphan	2,100	3,397
Active	1,579 (75% of NO)	1,966 (58% of NO)
Inactive	521	1,432
Early∗	463 (89% of I)	1,292 (90% of I)
Stale∗	98	674
Internal	73 (5% of A)	194 (10% of A)
Working set	1,506 (95% of A)	1,772 (90% of A)
Class	1,132 (75% of WS)	443 (25% of WS)
Research	356 (24% of WS)	1,267 (72% of WS)
Mixed	18 (1% of WS)	62 (3% of WS)

Table 1: Breakdown of project and user data per category. Starred rows are generated by taking the maximum warmup time for working-set projects/users as a threshold for declaring a project/user as early

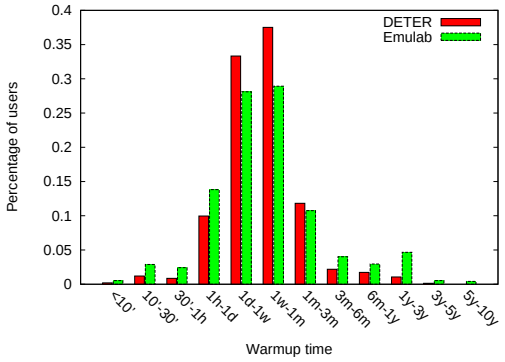

Figure 1: User warmup time in DETER and Emulab

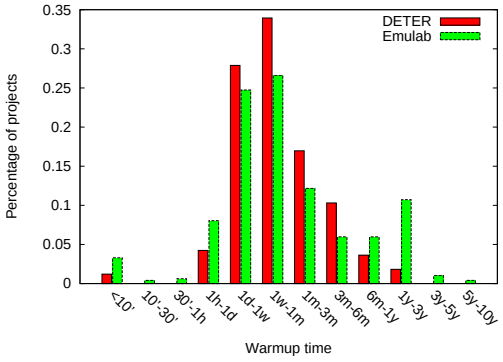

Figure 2: Project warmup time in DETER and Emulab

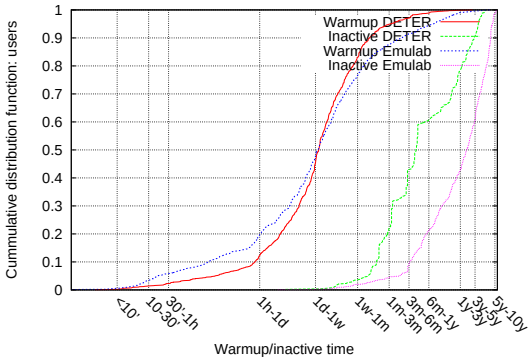

Figure 3: User warmup and inactive time

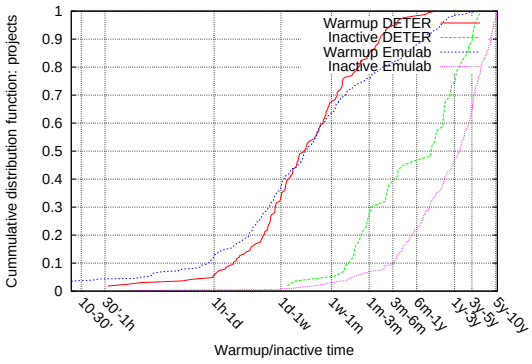

Figure 4: Project warmup and inactive time

Research category	Projects
Attacks	48
DDoS	18
Architecture	14
Infrastructure	12
Testbeds	12
Worms	11
Evaluation	8
Privacy	6
Botnets	4

Table 2: Projects per research category

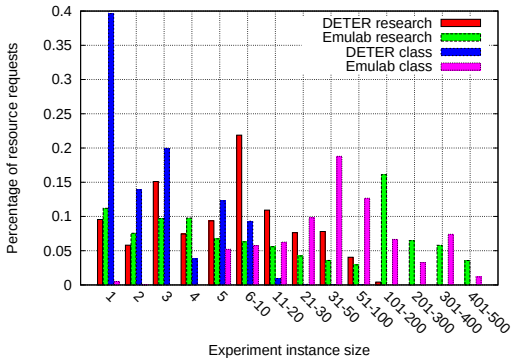

Figure 5: Experiment instance size

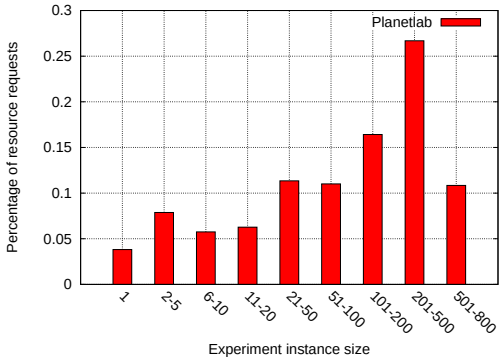

Figure 6: Experiment instance size in Planetlab

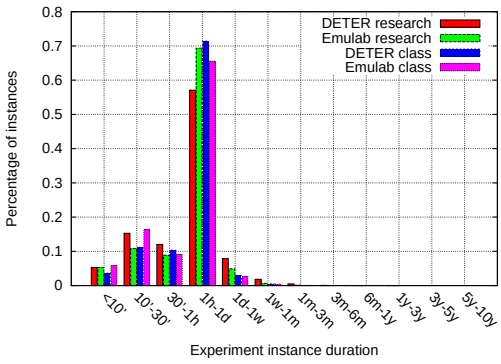

Figure 7: Experiment instance duration

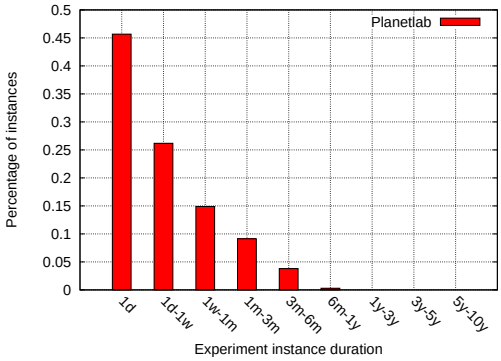

Figure 8: Experiment instance duration in Planetlab

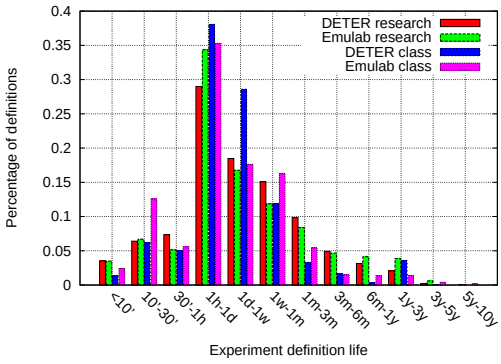

Figure 9: Experiment definition life

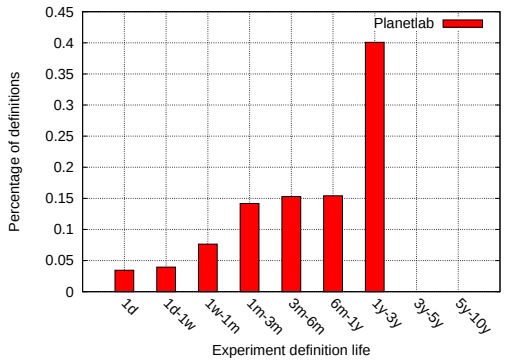

Figure 10: Experiment definition life in Planetlab

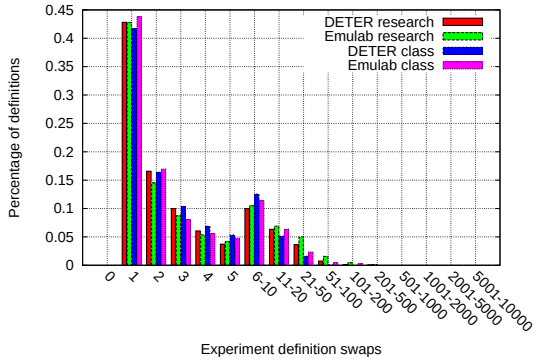

Figure 11: Experiment definition swaps

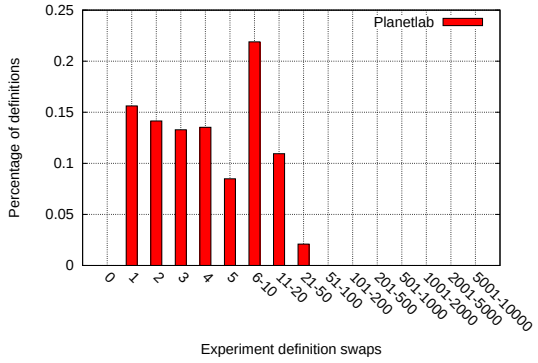

Figure 12: Experiment definition swaps in Planetlab

5.2 Project Distributions

In this case trends differ between all and outcome projects. Those with outcome are definitely more active.

Are outcome projects more active because they have been around longer or because they work harder? Both. Figure 18 shows that. Figure 19 shows that in large research projects with outcome almost all users are active, while in smaller projects only a few may be active. What's surprising is that for classes with outcome sometimes only half of the users are active. This may be due to them having a choice of using DETER vs not or maybe because we create more accounts than needed for the class.

Explain project evolution from inception to publication. Quantify how many projects do not result in productive use. Some don't manipulate experiments at all. Some are idle. Quantify this for users as well. Quantify this for experiments and for idleness as well. Not sure if idleness goes here.

5.3 User Distributions

Members can be deleted and we don't account for that

5.4 Geolocation Distribution

Table 3: Node activity

Activity	Percentage
Idle	57%
Network	30%
CPU	10
Interactive	3

57% of timeslots reported by allocated node were idle but per experiment all nodes are used together. There are really no huge inactive periods so inactivity is spread through the experiment. 26% of timeslots report a network activity. 6.8% report a CPU activity and 2% report a CPU and network activity.

cummulative per project idle nodes correlate with experiment duration? correlate with experiment history?

6. FUNCTIONAL PATTERNS

Per category, what are the functional uses

7. CONCLUSION

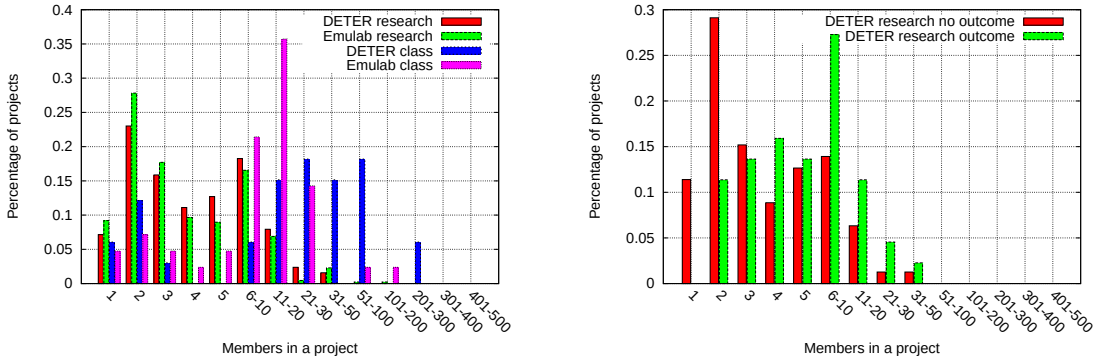

Figure 15: Members per project. Left: DETER vs Emulab, Right: All vs outcome

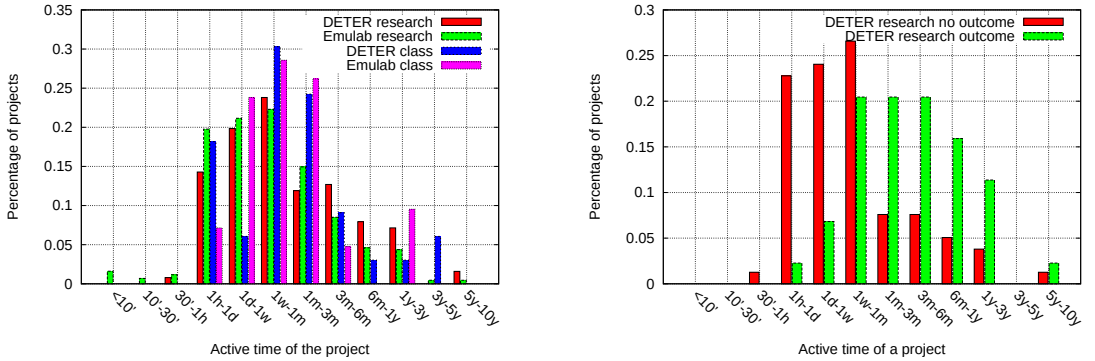

Figure 16: Active time of a project. Left: DETER vs Emulab, Right: All vs outcome

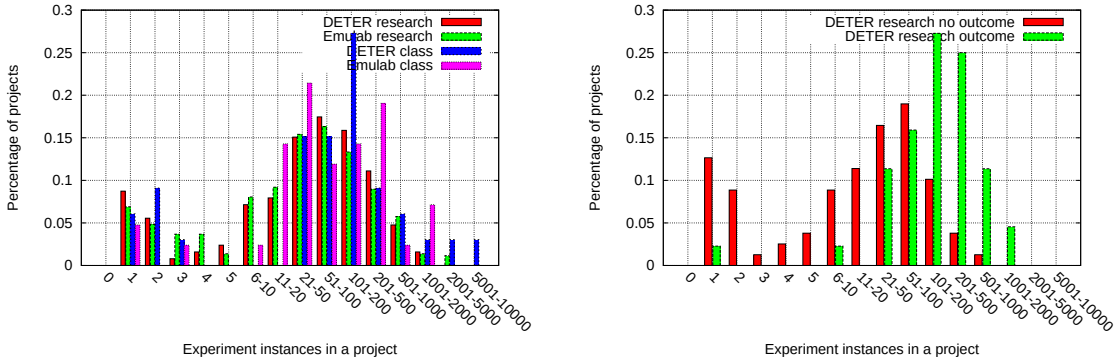

Figure 17: Experiment instances in a project. Left: DETER vs Emulab, Right: All vs outcome

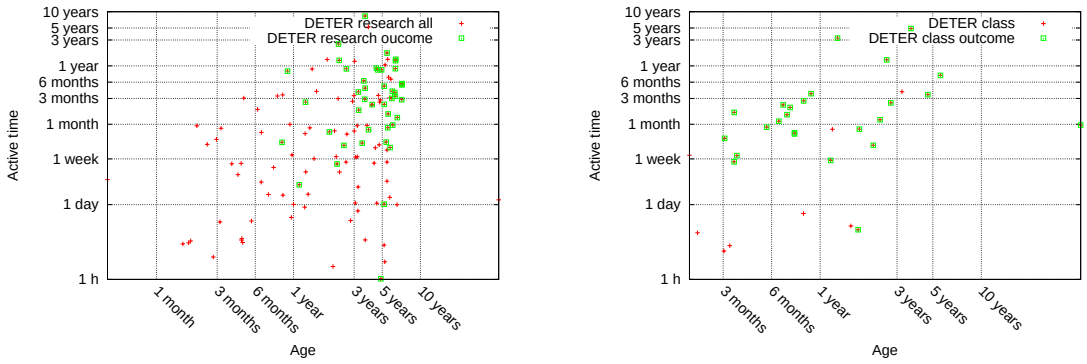

Figure 18: Active time vs age of a project. Left: DETER research, right: DETER class

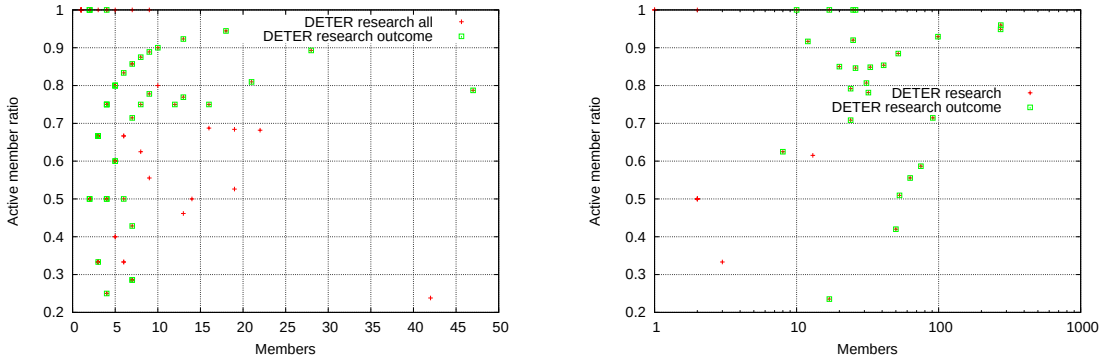

Figure 19: Active member ratio vs all members of a project. Left: DETER research, right: DETER class

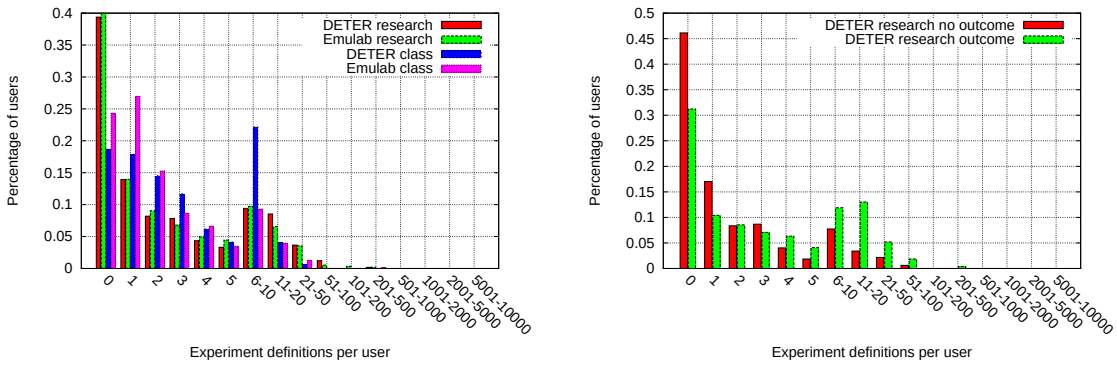

Figure 20: Experiment definitions per user. Left: DETER vs Emulab, Right: All vs outcome

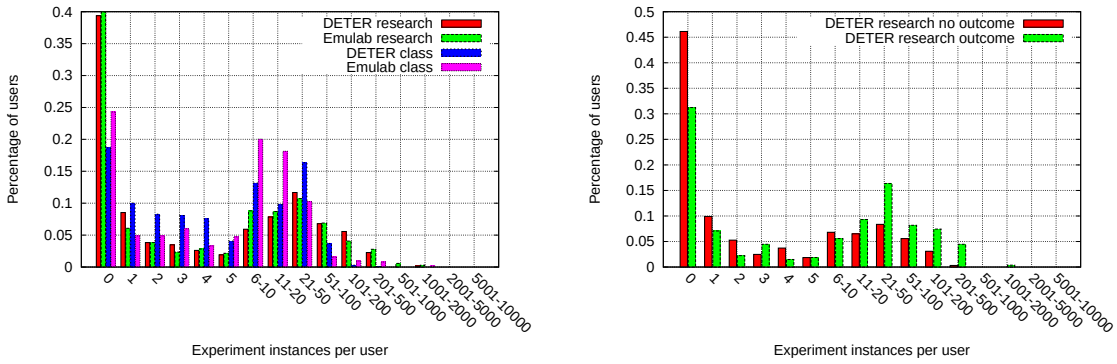

Figure 21: Experiment instances per user. Left: DETER vs Emulab, Right: All vs outcome

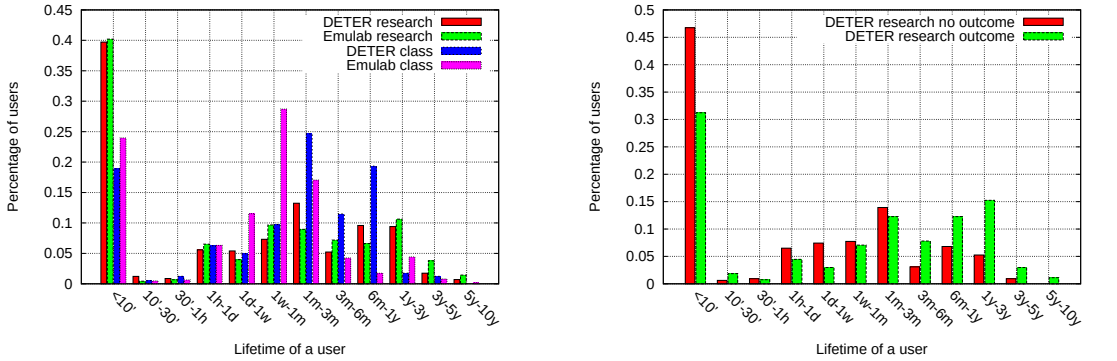

Figure 22: User lifetime. Left: DETER vs Emulab, Right: All vs outcome

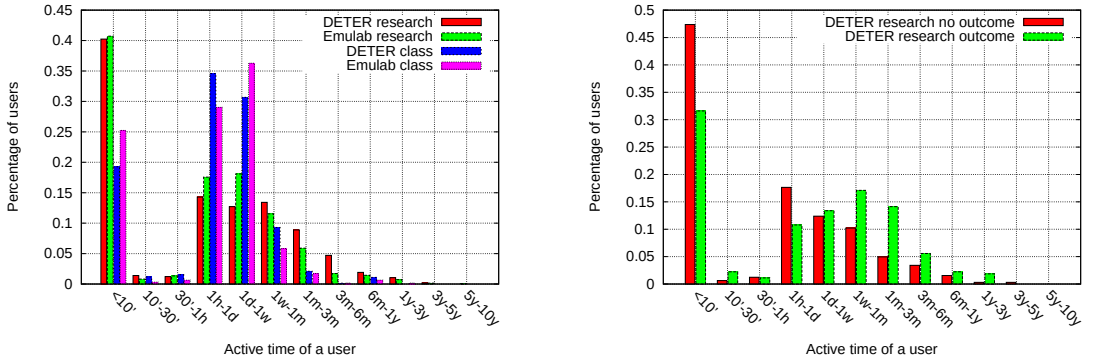

Figure 23: User active time. Left: DETER vs Emulab, Right: All vs outcome

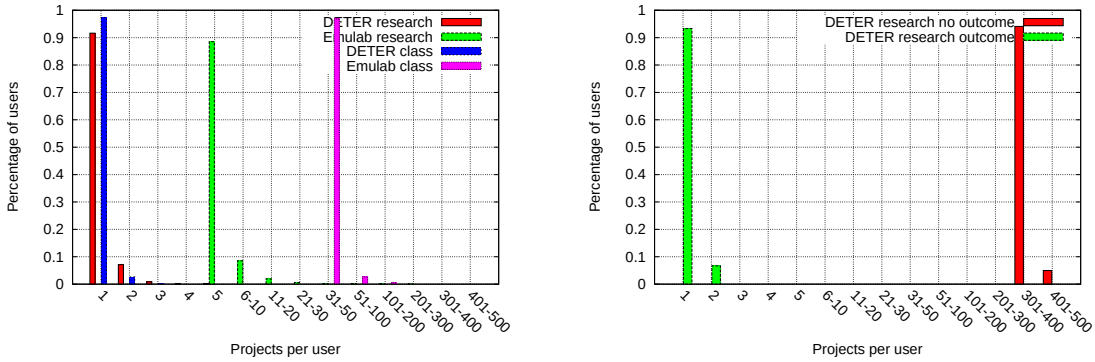

Figure 24: Projects per user. Left: DETER vs Emulab, Right: All vs outcome

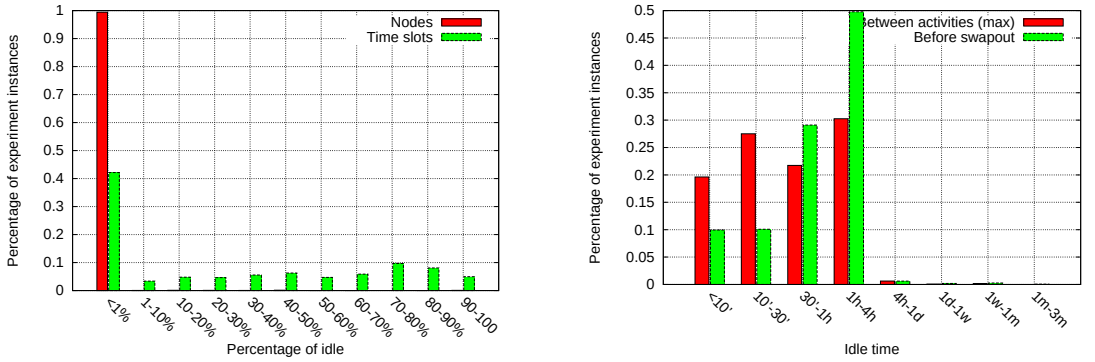

Figure 25: Idleness per experiment

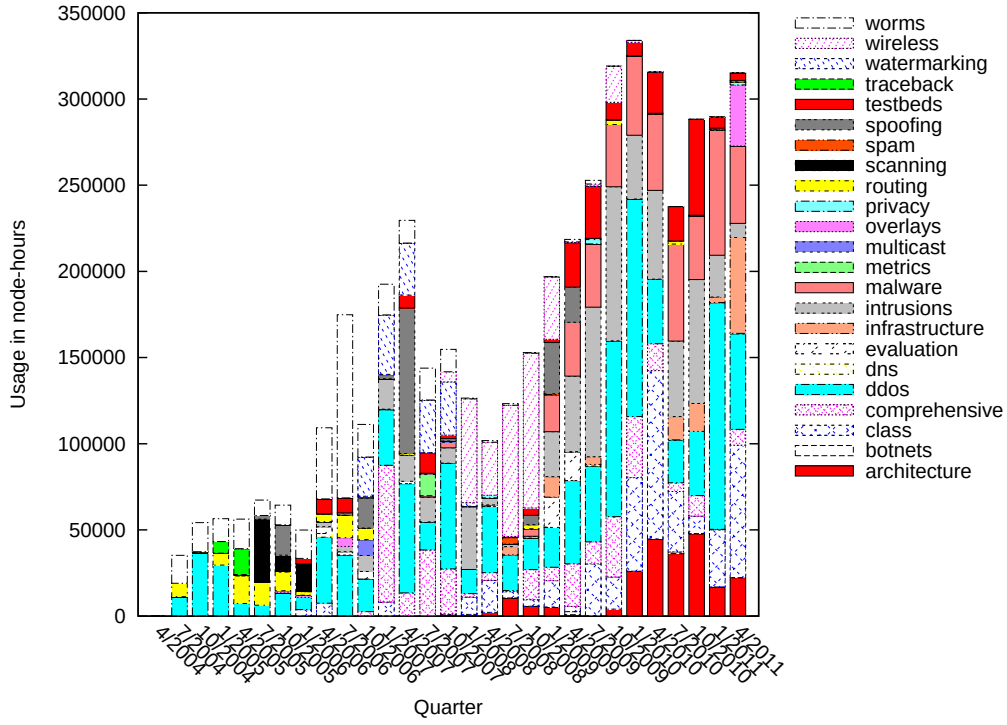

Figure 26: Usage per project category in node-hours in DETER

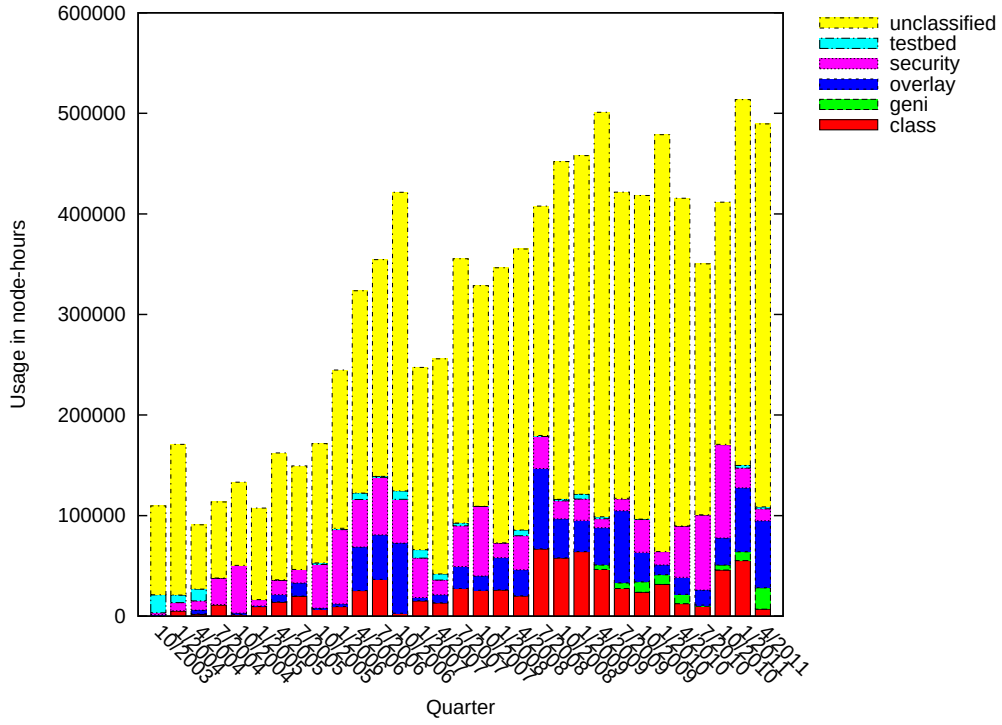

Figure 27: Usage per project category in node-hours in Emulab