Understanding and Resolving Conflicts on PlanetLab

Larry Peterson
Princeton University

November 13, 2008

Abstract

This paper describes our experiences operating the PlanetLab network testbed over the past six years, and in particular, how we resolved disputes between the network research community that uses PlanetLab, universities and companies that host PlanetLab nodes, and unaffiliated Internet users impacted by PlanetLab experiments. It also reports the principles that have guided the resolution of these disputes, and in doing so, attempts to inform the community's on-going dialogue about policies governing the uses of experimental network infrastructures.[1]

1 Introduction

PlanetLab is a global testbed for network and distributed systems research. Created in 2002, PlanetLab has grown to nearly 1000 machines spanning over 470 sites and 40 countries. Over its six-year lifetime, PlanetLab has been used by over 4700 researchers as an observatory from which today's Internet can be studied, and as a platform for evaluating the network services and protocols of tomorrow's Internet [2].

PlanetLab is an overlay on today's Internet, which means the experiments it hosts have the opportunity to run "in the wild," that is, send packets to and receive packets from regular Internet hosts. This results in rich interactions between experiments and various network stakeholders, and as a consequence, PlanetLab provides visibility into the tensions and disputes that arise among these stakeholders.

This paper describes the disputes that have arisen over the last six years, analyzes the social and legal tensions at their root, and reports the principles that have guided the resolution of these disputes in the context of PlanetLab. In doing so, this paper informs the community's on-going dialogue about policies governing the uses of experimental network infrastructures, such as GENI [1].

2 Background

PlanetLab is a community-based effort. Research institutions contribute machines and connect them to their campus networks, and in return, researchers at those institutions are granted access to the global pool of machines.

Every PlanetLab node runs a Linux-based operating system that has been modified to implement a *slice* abstraction—a network-wide resource container that hosts experiments. A control center—known as PlanetLab Central (PLC)—provides researchers with an interface to create and control their own slices, and operators with an interface to remotely manage the nodes (e.g., install and update software, monitor node

[1] This is a working document. Please send comments and suggestions to the author at *llp@cs.princeton.edu.*

health, allocate resources to slices, and audit network traffic). This section outlines the aspects of PlanetLab relevant to our discussion; additional detail can be found elsewhere [14].

2.1 Actors

PlanetLab includes four principle actors:

- **Hosting Sites**: Organizations that own and host nodes. Hosting sites share responsibility for their nodes with PlanetLab Central.
- **PlanetLab Central** (also known as PLC): An operations center that provides on-going support for PlanetLab nodes. PLC acts as a trusted intermediary between hosting sites and researchers.
- **Researchers**: Individuals and project teams that are granted a slice of PlanetLab's resources in which they run measurement experiments and experimental network services.
- **Third Parties** (also known as users): Individuals and organizations not affiliated with PlanetLab. These include users that take advantage of services deployed on PlanetLab, content providers that deliver content via PlanetLab, and ISPs that carry data between PlanetLab and the rest of the Internet.

2.2 Hosting Nodes

Each hosting site connects a set of PlanetLab nodes (usually two or three) to their campus network, giving them connectivity to local users and machines (via the campus net), as well as to arbitrary Internet hosts (via the site's commodity and research service providers). This includes assigning a static IP address to each node, and resolving DNS names (both forward and reverse lookups) for each node.

Hosting sites are expected to locate their PlanetLab machines outside the campus firewall, but behind the campus border router—on a so called *DMZ network*. Keeping the machines outside the firewall offers standard protection to internal network assets, while giving the experiments running on the nodes unfettered access to the rest of the Internet. The expectation is that hosting sites will not filter ports or ping packets for the nodes they host.

2.3 Trust Assumptions

PlanetLab's software architecture is designed around the following four trust assumptions—the assumptions that, if they hold, result in the system working as expected. Our experience to date is that all the parties are living up to their obligations.

- PLC expresses trust in a researcher by issuing it credentials that let it access slices. This means that the researcher must adequately convince PLC of its identity (e.g., affiliation) and intent to adhere to best practices.
- A researcher trusts PLC to create slices on its behalf and check credentials so that only that researcher can install and modify the software running in its slice.
- A hosting site trusts PLC to manage the nodes, for example by installing software on its nodes that isolate resource usage of slices and bound/limit slice behavior.

- PLC trusts hosting sites to keep their nodes physically secure. It is in the best interest of hosting sites to not circumvent PLC (upon which it depends for accurate management of its nodes). PLC must also verify that every node it manages actually belongs to a member hosting site.

Note that those trust assumptions that fall on PLC (the second and third) effectively establish the requirements placed on the PlanetLab code base.

2.4 Membership Agreement

Many of the relationships and assumptions described in this section are spelled out in a membership agreement signed by research organizations that join the PlanetLab Consortium [16]. An important aspect of that agreement is that the behavior of a slice is the responsibility of the researcher that creates the experiment running in that slice, along with that researcher's home institution. Hosting sites host nodes, but they are not responsible for the slices running on those nodes. PLC manages nodes, but it is not responsible for slices running on those nodes.

3 Resolving Conflicts

PLC fields and responds to complaints from various sources. This section discusses the sources, and outlines the actions taken in response to each. One of the goals of this section is to identify the mechanisms that have been designed to provide a technical solution to these complaints.

3.1 Third Parties

When a third-party receives an unwanted packet from a PlanetLab node, it either (a) contacts the hosting site, which forwards the complaint to PLC; or (b) contacts the web server running on the PlanetLab node, which helps them formulate a complaint that is sent directly to PLC. In both cases, PLC uses an auditing tool to determine the responsible slice, and asks the corresponding researchers for an explanation. The following outlines typical actions:

- Hosting site takes the nodes off-line or blocks traffic without waiting for PLC to respond.
- PLC suspends the offending slice until corrective action is taken.
- Researchers modify their experiment to eliminate questionable behavior.
- PlanetLab blacklists the third-party site, so they receive no more packets.
- Researchers consult their slice-specific logs to learn about any user was using/abusing their service, and when appropriate, share this information with complainant.
- In the rare case that PLC suspects the node itself has been compromised, PLC brings the node into *safe mode*, which allows the support staff to safely access and inspect the node, but disables all slice-related activity.

Three key technical mechanisms allow PLC to do its job. One is isolation machinery (virtual machine monitor) that bounds the resources any give slice is able to consume [18]. The second is an audit mechanism that allows PLC to trace network activity back to the responsible researchers [11]. This is important because it allows the response to affect only the responsible slice, rather than all experiments running on the offending node, which is the undesired consequence of bring the entire node down. The third is a secure remote boot mechanism that allows PLC to safely inspect the node, even if the kernel has been compromised.

3.2 Hosting Sites

In addition to satisfying the original complainant, PLC must also convince the hosting site that they should keep their nodes on-line, or bring them back on-line if they initially took them down in response to the complaint. This recovery process is made all the more urgent by the fact a single incident often causes the third-party to receive unwanted packets from multiple PlanetLab nodes, which results in multiple hosting sites taking their nodes off-line.

Quickly resolving third-party complaints is usually sufficient, but PLC also plays the role of a research advocate—helping researchers at the hosting site, who have a vested interest in their site remaining connected, convince the site's security response team that the benefits of having access to PlanetLab outweigh the risks of hosting nodes. (PLC also plays this role when helping a researcher make a case for joining PlanetLab in the first place.) Hosting sites are often worried about avoiding the next complaint, once the current complaint has been resolved. This problem is amplified by third parties that claim much more than the law allows, which has a "chilling effect" on hosting sites.

In some cases, hosting sites will object to the behavior of experiments without first being contacted by a third party. There are two general situations in which this happens:

- The PlanetLab nodes they host are consuming an unusual amount of site resources (e.g., bandwidth). PLC supports a fourth mechanism—bandwidth throttles—that limit such consumption. In some cases, PLC pushes back by explaining that contributing these resources is the cost of being able to use PlanetLab. Some sites have to deal with prohibitively expensive bandwidth charging models, which complicates the resolution of this issue.

- The site's Intrusion Detection System (IDS) flags a PlanetLab node as having been compromised. These are almost always false alarms since IDS usually look for Windows vulnerabilities, but PlanetLab runs Linux. The exceptions are when the hosting site objects to an experimental service running on their nodes (e.g., an open proxy on port 3128). Assurances from the researchers sometimes work, but in a few cases, the researcher agrees to not run their service on that site's nodes. PlanetLab includes a fifth mechanism—per-slice resource limits—to implement such bans, but PLC prefers researchers voluntarily avoiding certain sites for the sake of not establishing the precedence of sites being able to pick-and-chose which experiments they want to allow.

To complete the story on hosting sites, it is important to understand that many are connected to national research and educational networks, that in turn impose their own terms and conditions on network traffic. For example, many US universities are connected to Internet2, and universities in the UK are connected to JANET. When these networks object to traffic originating on PlanetLab nodes being hosted at member sites, PLC must work with the national network to establish a common understanding of acceptable usage.

One example involves Internet2's original policy that its backbone not be used to carry traffic to/from the commodity Internet. Many experimental network services, however, benefit from exactly that capability—being able to study and measure user traffic is essential to this research. To better support network research (as opposed to carrying only large data sets on behalf of other science disciplines), Internet2 relaxed this restriction.

As a second example, flows generated by some PlanetLab experiments triggered alarms on flow-monitoring systems used by JANET(UK), the organization that runs the UK national research and education network, to detect the similar flow patterns generated in the early stages of worm propagation and spam distribution. Following discussions between PLC and JANET(UK) it was agreed to whitelist PlanetLab nodes for which the detection algorithms could be modified to reduce their sensitivity. To ensure that this did not increase the risk either to JANET customers or other Internet users, it was agreed that any reports of incidents involving these nodes would be forwarded directly to PLC, as well as to the hosting sites. Although the host site remains responsible for ensuring the security of all its networked computers, including their PlanetLab nodes, it was hoped that speeding up the investigation in this way would allow the host site to reach a quick and accurate assessment of the nature of any unexpected traffic.

However, this discussion also identified a potential policy problem since the JANET Acceptable Use Policy prohibits “deliberate unauthorised access to networked facilities or services.” It was therefore necessary to clarify to what extent PlanetLab probe packets were, or were not, authorized. The relevant definition of “authorization” is that used by the UK’s Computer Misuse Act [12], and a recent UK court decision (DPP v Lennon) on denial of service attacks suggests that connecting a machine to a public Internet address does not give carte blanche authorization for any and all traffic to be sent to the host. However, advertising an IP address as providing a service does appear to authorize reasonable attempts to send packets that would normally be part of the use of that service. PLC was therefore requested, so far as possible, to ensure that experiments run on JANET-hosted nodes do not probe hosts or ports picked at random, but only those advertised as providing the relevant service.

3.3 Researchers

Researchers are interested in the operational stability of the system, and how to effectively use it for their research. The former concerns are outside the scope of this paper. With respect to the latter, most can be traced to limitations placed on PlanetLab nodes by hosting sites. For example, a site might take its nodes off-line due to an earlier complaint, or it may place various restrictions (e.g., firewall blocking certain TCP ports) on the kinds of experiments that can be performed. PLC plays an important role in balancing these two perspectives: pushing researcher demands too hard risks losing hosting sites (thereby diminishing the research value of the platform), while pushing site demands too hard risks rendering the platform less valuable to researchers.

An illustrative example of this balancing act concerns a measurement experiment that participated in BitTorrent swarms, up to the point of downloading the first few kilobytes of files for the sake of measuring throughput rates to various Internet sites. Only the first few kilobytes of these files were downloaded, and so the contents were not usable, but in certain cases these partial downloads triggered copyright infringement complaints from the MPAA. These complaints were directed at the hosting sites’ DMCA agent, which immediately instructed that the site’s PlanetLab nodes be taken off-line. The legal counsel at the researcher’s home institution was willing to defend the research if it ran locally, but because other hosting sites demonstrated an unwillingness to accept the same level of risk, the price to PlanetLab as a whole—in terms of sites taking their nodes off-line—was deemed too high to allow the experiment to run across the whole platform. Subsequently, the company monitoring the Internet for copyright infringement on behalf of the MPAA agreed to white-list PlanetLab nodes, but it is clear that the threat of DMCA-related action has a chilling effect on many Universities.

4 Usage Scenarios

This section catalogs the disputes that PLC has encountered, organized according to the kind of research being done; i.e., according to the scientific and technical context in which they emerge. It also gives a preliminary analysis of the social and legal concepts that seem most relevant, often borrowing from the language used by those registering the complaint. For each type of complaint, we briefly comment on its frequency.

4.1 Measurement Studies

Researchers conduct measurement studies that probe the Internet (send packets into the Internet) in an effort to learn about its structure, performance characteristics, loss properties, and so on. Third parties—either the subject of a probe or an ISP speaking on its behalf—sometimes object to these probes. Probes can be classified as being of three general types.

- **Network-Level Probes:** An experiment running on a PlanetLab node sends a probe to some set of unsuspecting Internet hosts. Third parties object to the rate of the probes (suspecting a DDoS attack) or to the specific port being probed (suspecting an attempt to break into the machine). The network research community has established best practices that avoid most of these complaints [19], but some sites complain about even these "acceptable" probes, and at the same time, some researchers fail to adhere to the best practices. Such complaints occur on a regular basis, with one site threatening legal action after receiving three probe messages.

- **Application-Level Probes:** An experiment running on a PlanetLab node participates in—up to a point—an Internet application. Such experiments interact with a third-party by imitating a "real" client or server, but with the objective of measuring some network property rather than engaging the service. Three examples illustrate:

  - **HTTP:** The researcher installs a "web bug" in a legitimate web page so a client unknowingly contacts a PlanetLab node. Such experiments require the cooperation of the content provider for the legitimate page. There have been no known complaints about web bugs.
  - **DNS:** An experiment queries a DNS server to resolve a domain name, and then asks the same server to forward a query to a third DNS server in an effort to measure the latency between the server node and the end target node [10]. Occasional complaints of this sort have occurred, with one very upset tier-5 ISP threatening legal action and initiating a DDoS attack on PLC.[2]
  - **P2P:** An experiment participates in BitTorrent or Gnutella, up to some point, but without downloading a (possibly illegal) file. There has been occasional threat of legal action from agents acting on behalf of copyright holder, plus occasional objection from hosting sites that prohibit P2P activity of any type.

- **Edge Probes:** An experiment running on a PlanetLab node installs "active content" in a client's web browser. This code sends probes from the clients machine into the Internet, and then eventually packages the results and sends them back to the researcher. This technique is useful for measuring the

[2] This episode was rare in the level to which the third-party escalated the complaint, and while he was never satisfied with PLC's response (we black-listed the site and had the researchers modify their experiment to be less aggressive), there was insufficient harm to the ISP (it received roughly two DNS requests per second for a period of a few days) to pursue it any further.

usage of middleboxes (e.g., NATs), which can only be detected from the probes that originate from the edge of the net. This active content is deployed using a method similar to the web bug described above, except a CDN (next section) is able to do so without the cooperation of the content provider by rewriting pages it caches. This is not a common practice, and there have been no known complaints received to date.

With respect to edge probes, PlanetLab is pursuing an opportunity to work with companies that market user devices and applications to instrument those edge points-of-presence with probes back to servers running in PlanetLab. This will give researchers an unprecedented view of the Internet, but has the potential to dramatically increase the volume of third-party complaints.

4.2 Deployment Studies

Researchers also deploy continuously running network services that attract a user community. This is generally a symbiotic relationship: users gain some benefit by using the service (e.g., faster downloads) and researchers learn how the service behaves under real workloads. To give a sense of the magnitude of such services, PlanetLab typically carries 6TB of Internet-bound traffic on behalf of 1 million users each day.

Content Distribution Networks (CDNs) are the most successful example of continuously running services on PlanetLab. These CDNs effectively interpose themselves between clients that access content and servers that provide content. Interpositioning is a straightforward thing to do, either by setting a web browser parameter (on the client side) or rewriting a URL (on the server side). For other experimental services—e.g., alternative implementations of BitTorrent [15]—researchers might provide an "agent" that users willingly download and run in order to take advantage of the service.

In deploying and operating such experimental services, several issues arise:

- **Opt-in Strategy:** How the service attracts real users and workload. Three common interposition techniques are applicable to CDNs; services that require a user to download and run an agent are essentially equivalent to client opt-in. Each generates its own set of complaints:
  - **Client Opt-in:** Clients point their web browser at a PlanetLab node, treating it as a proxy. In this case, content providers sometimes object that they have no visibility into who is really accessing their content. Complaints of this form are received on a regular basis.
  - **Server Opt-in:** Content providers make their content available through a PlanetLab service, usually by rewriting the URLs for the pages they want delivered via the experimental service. In this case, clients complain that they don't understand why they are pulling content from PlanetLab nodes rather than the origin server. Complaints of this form are received on a regular basis, often from users that suspect spyware has been installed on their machines. [3]
  - **Transparent Opt-in:** Third parties (neither client nor server) force traffic through a PlanetLab service without either the consumer or provider being aware. Complaints of this form are received on a regular basis.
- **Unwanted Probes:** In an effort to make their services more effective, experimental services sometime probe the client that originally requested the service, for example, to select the most appropriate ingress node for that client. Such probes result in the same sort of complaints as those generated from

[3] Such complaints are not unique to experimental services, but are also reported by commercial services like Akamai.

measurement studies, except in this case, the response “you contacted us first” generally resolves the issue. Ironically, one can point to inadequacies in today’s network architecture as resulting in the need for such probes. Such complaints occur on a fairly regular basis.

- **Illegal Content:** Third parties complain about illegal content being carried on PlanetLab. When someone complains about illegal content (sometimes a hosting site, sometimes the copyright owner, sometimes a law enforcement official), PlanetLab and the service in question turn over log files. All three situations happen occasionally.

- **Privacy:** Experimental services typically maintain log files. These logs provide a wealth of research data, but are also used to respond to subpoenas and complaints. These log files are not made public, and would certainly need to be anonymized before doing so (if at all). Whenever a client explicitly opts-in, he or she sees a page telling them that their use of the service should not be assumed private, but users are typically unaware that their requests are being logged when they access content made available as a result of server (content provider) opt-in. Both PLC and experimental service providers have handed logs over to the authorities under subpoena or other compulsory process, and have voluntarily cooperated with third parties that were judged to have a legitimate complaint about some user that is abusing a service (see next point).

- **Abusing Services:** Users will find ways to abuse network services, for example, to access illegal content, post anonymous email, auto-click ads, access proprietary material, and obfuscate phishing sites. Researchers have established best practices to thwart such abuses [20], but this is an arms race [13]. Interestingly, PLC also receives complaints from users about these abuse-prevention mechanisms keeping them from taking advantage of the services in questionable ways. Complaints about not being able to use PlanetLab services for purposes outside what the service provider deems acceptable are surprisingly common.

- **Bandwidth Shifting:** It is possible for content providers to ship their content via PlanetLab, which takes advantage of network bandwidth made available for research purposes. Server-side opt-in is most problematic in this regard. In some cases, technical mechanisms limit usage to PlanetLab sites, non-profits, and so on. It is difficult to know how much bandwidth shifting happens in practice, and while there has been discussion in the community about this possibility, no serious objections have been raised to date.

- **Financial Gain:** There is always the possibility that researchers will want to transition their work into a commercial setting, where conflict-of-interest is a well-understood issue. A somewhat newer question, however, is whether researchers should be able to benefit financially from the log files and other measurement data they’ve collected on a research facility like PlanetLab. The possibility of inadequately-anonymized logs being distributed only compounds this scenario. Some initial discussion on this point is starting to happen in the community, but there as been no “test case” to date.

5 Policy Positions

Settling disputes between researchers, third-parties and hosting sites—and more generally, setting policy for how PlanetLab can and cannot be used—is an exercise in balancing competing rights. Where the law is clearly applicable, we follow it. More often than not, however, the issues of the law itself are not black-and-white (see [17, 8] for an excellent summary of several legal issues related to network research), in which case there are often ethical issues to consider—balancing the rights of privacy, autonomy, and property of sites and third-parties against the public good of the research being conducted.

Over time, we have evolved a set of policies/positions that guide our decision-making process. This section summarizes these positions, organized around the three primary stakeholders with which PLC interacts. The following section fleshes out some of the philosophical arguments that underlie these positions.

5.1 Third Parties

Four policies guide PLC's interactions with third-parties that believe they have been affected by PlanetLab activity.

- **Do not police content.** PLC does not police or censor content. Policing content is problematic for two reasons. First, it raises the expectation that the policing will be perfect, so that objectionable or illegal content will not get through. We do not trust the available technology to achieve this standard. Second, policing content is vulnerable to complaints of censorship, and puts PlanetLab in the untenable role of Internet censor.

- **Prefer opt-out to opt-in.** PLC does not require researchers to receive explicit permission to include an Internet host in a probing experiment (i.e., explicit opt-in is not required), but it must avoid contacting sites that ask to not be contacted (i.e., opt-out must be honored). To this end, PlanetLab maintains a blacklist of Internet addresses. The rationale is that a site that connects a machine to the public Internet implicitly accepts that others will attempt to connect to it. The site has the right to object to (and take steps to thwart) unwanted traffic, such as a denial of service attack, but it cannot reasonably expect the rest of the Internet to ask permission to contact its machine before doing so the first time.

- **Users need not explicitly opt-in.** Many experimental services running on PlanetLab carry traffic generated by users. PLC does not require that researchers give these users an opportunity to explicitly opt-into their services. PLC does not view these users subject to Institutional Research Board (IRB) oversight [5], although any data that might indirectly reveal private information about these users must be protected from public disclosure.

- **Solicit third-party exemptions.** Third parties that recognize the value of research being done on PlanetLab, or are at least accept that the experiments are doing no harm, are sometimes willing to explicitly exempt PlanetLab from their monitoring tools, thereby avoiding the false positive that result from PlanetLab-generated traffic. PLC takes every opportunity to convince ISPs to not trigger complaints about probe packets from PlanetLab nodes, and the MPAA to whitelist PlanetLab nodes on their DMCA detection systems. There are successful examples on both counts.

Implicit in all third-party interactions is the underlying principle of being responsive. PLC does not attempt to prevent all complaints, but is responsive when complaints do occur.

5.2 Hosting Sites

Interacting with hosting sites is similar to interacting with third parties, since the former are often acting in response to (or in anticipation of) complaints from the latter. Here, we identify three policy positions:

- **Do not police content.** This is a restatement of the same principle in the previous section. Hosting sites may be uncomfortable with certain content, but in most cases have come to understand that proactive filtering is not a viable solution. Moreover, research usually benefits from being able to carry real content from real users rather than run only synthetic workloads.

- **Support research.** This is a hosting site variant of the opt-out principle. Hosting sites may refuse to host experiments that they object to (opt-out), but PLC does not provide a mechanism for them to explicitly select the experiments they do wish to host (opt-in). In practice, PLC goes a step further by articulating the case for hosting sites to accept all experiments that adhere to best practices. The key point of leverage is that the site's researchers are allowed to use all of PlanetLab's nodes as long as it allows other sites' well-behaved researchers to use its nodes.

- **Provide PR cover.** For some hosting sites, bad publicity is also a serious risk. PLC attempts to shield hosting sites from bad publicity as much as possible, for example, by allowing companies to place the nodes they host in the planet-lab.org domain. On the other hand, potential embarrassment is not viewed as a valid reason for shutting down an experiment as long as the research is defensible.

This discussion focuses on what hosting sites can and cannot expect from PLC, but it does not directly address what PLC should be able to expect from hosting sites. Although hosting sites officially accept certain terms when they connect to PlanetLab (e.g., that they will not filter traffic), virtually no hosting site understands the implications of hosting PlanetLab nodes until the complaints start to arrive and the IDSes start to fire. Ideally, hosting sites would acknowledge the risks and hassles outlined in this report, and agree to support the research community (i.e., keep their nodes on-line and their filters turned off) as long as PLC consistently implements the policies outlined in this section.

5.3 Researchers

The primary interaction in PlanetLab is between PLC and the research community. We identify six guiding positions on this relationship.

- **Researchers are responsible for their experiments.** PLC cannot respond to third-party complaints without involvement of the researchers whose experiments trigger the complaints. PLC identifies the responsible researchers as quickly as possible. The auditing mechanism makes this possible. Also, researchers that commonly trigger complaints are required to subscribe to the support mailing list so they can respond to incidents immediately. Once the responsible researchers are identified, PLC's main role is to make sure that they respond to the complaint in a satisfactory manner, and when appropriate, defend their experiments to the hosting site(s) cited in the complaint. Researchers, however, are ultimately responsible for the behavior of their experiments.

- **Expect best-practices from researchers.** Because poorly designed experiments hurt everyone—they put PlanetLab at risk of having nodes taken off-line—both PLC and the broader research community have little tolerance for naive researchers. Slices running poorly designed experiments are suspended until the responsible researchers can satisfy PLC that their experiments are consistent with best practices. Note that best practices vary on an experiment-by-experiment basis [19, 20], but one practice is common across all of them: researchers should slowly ramp up their experiments so that the impact of a mistake is minimal. This includes demonstrating that the experiment runs without complaints at the researcher's home site before porting it to remote sites.

- **Expect new activity will cause complaints.** The nature of research conducted on PlanetLab is that it will trigger complaints, if for no other reason than that the generated traffic is not typical. PLC's threshold for allowing an experiment is not that it triggers no complaints, but rather, that the experiment is defensible on its technical merits, and that the risks to PlanetLab are tolerable (i.e., will not cause a significant number of sites to take their nodes off-line). In a sense, the hosting sites play the role canaries in the mine-shaft.

- **Preserve privacy of log files.** Many experimental services collect log files. They are needed to provide operational support for the service, and represent an important research asset (i.e., they are used to improve the service). To date, however, anonymized versions of these logs have not been published for use by the broader community, even though there is general agreement that they contain valuable research data. One key obstacle is a fear that anonymization techniques are not foolproof. Because of the importance of protecting the privacy of users that might be indirectly identified in these logs, PLC views controlled access to this data as a best practice that must be followed.

- **Prevent service abuse.** Another best practice PLC enforces on researchers is that they not allow their experimental services to be abused, that is, used to impact users in ways that those users object to. Researchers must also allow users to opt-out of their service should they be inconvenienced in any way.

- **Leverage PlanetLab's unique properties.** An experiment that would have been just as effective (had just as much fidelity) if run from a single site rather than the several hundred points-of-presence offered by PlanetLab is not an appropriate use of PlanetLab, and so any risk (in terms of complaints raised about the experiment) is not worth taking. Simply reducing the running time of an experiment—because it can be run in parallel at 1000 machines—is not an adequate justification for an experiment that triggers substantive complaints.

6 Discussion

The policies outlined in the previous section are based on an implicit foundation of legal and philosophical assumptions. This section expands on some of these assumptions, not only to show in more detail how they lead to specific policies we have adopted in PlanetLab, but more importantly, to reveal how to adapt these principles to future situations that will undoubtedly occur on PlanetLab and similar network experimentation platforms.

6.1 What's Distinctive about Research?

It is useful to distinguish between issues that turn on the distinctive character of the Internet, and those that turn on the distinctive requirements of conducting research on top of (embedded in) the Internet.

In the former, tensions arise over behaviors the network's design enables, which certain actors find objectionable. These tensions may take various forms and, in our view, deserve to be studied. At one extreme, actors may object to their machines being accessed in ways that are not only enabled, but in fact required for the proper functioning of the network. At the other extreme, actors may defend objectionable behaviors on grounds that they have merely found clever ways to exploit present-day design characteristics of the network to full advantage, and have not "broken" or "misused" the technology. Between these two extremes is a range of cases in which disputing actors believe that others are behaving inappropriately, harming them, threatening them, or violating their rights, while those accused cite requirements and technical properties of the network, their own competing claims, or the reasonableness of their behaviors in light of expectations embodied in the network architecture.

With respect to the distinctive requirements of conducting network research, tensions arise when researchers take advantage of network capabilities. For example, measurement experiments that send probe packets to Internet hosts are required to honor opt-out requests, but need not secure explicit permission before including a host in an experiment. The rationale is that a site that connects a machine to the public Internet implicitly accepts that others will attempt to connect to it. Today's accepted practice grants the site the right to object to (and take steps to thwart) unwanted traffic, but it does not reasonably expect the rest of the Internet to ask permission to contact its machine before doing so the first time. This position is consistent with the UK legal opinion outlined at the end of Section 3.2, which allows for hosts that advertise a service being the target of Internet packets, but randomly selected hosts are not. This is also consistent with how we enforce PlanetLab's current AUP.

The reason it is useful to distinguish the two sets of issues is that some of the push back PlanetLab researchers have experienced is due to disagreements of the first kind—namely, network actors believing that researchers are overstepping the lines of proper behavior—whereas others call into question the privileges and obligations of researchers specifically. It is hard to say why network researchers not be given at least as much leeway as the average network citizen, although what additional leeway and under what conditions is important to investigate. For example, exactly how much additional traffic a site should expect (tolerate) in the name of research is open to debate, where the costs borne by a site are weighed against the value of the research for promoting the public good. It is our hope that the answer to these questions, and the establishment of best practices, can be left to the network experts.

6.2 Public Good

The cornerstone of any argument in favor of accommodating research is that it has the potential to benefit the general public—that it be of some *public good* or serve the public interest. In relation to network research, we characterize the public good as increased transparency of the network (i.e., improves society's understanding of the inner-workings of the Internet) and helping engineers build a better network, where "better" includes faster, more robust, safer, more secure, more open, more accessible, and so on. It is worth noting that we are not presuming to define "better," but rather are accepting as a starting point the generally held view of network engineers, who are also responding to a sense of the broader good for the societies in which they live. For engineers in the United States, for example, we imagine they are responsive to the values commitments of liberal democracy.

Although transparency and improved performance, security, or robustness are part of what constitutes public good for network users, a further requirement is that the benefits, in principle, be available to all and not only a privileged few. Making special accommodation for network research is warranted only when its benefits flow equitably to network users, in general, and not only to those with special, or vested interests. It is important to note that this distinction is not equivalent to a distinction between non-profit and for-profit. That someone is able to translate transparency available to all into commercial advantage, or commercialize a better-engineered network service does not negate the public good of the relevant research, as long as the better-engineered network is available to anyone wanting to take advantage of it. For this reason, we have no qualms about commercial involvement in PlanetLab, or for-profit spinoffs from PlanetLab, as long as this involvement does not prevent the unrestricted flow of research findings.

6.3 Human Subjects and Consent

The US federal government defines regulations and guidelines governing the involvement of human subjects in biomedical and behavior research [5]. This includes oversight by an Institutional Review Board (IRB) at the researcher's home institution. Unlike health-related research [9], however, there are no guidelines or exceptions spelled out for network research.

As a general matter, we do not believe PlanetLab experiments are subject to these regulations. IRB oversight is intended to protect humans that are the subject of biomedical or behavioral studies, but the experiments discussed in this report are fundamentally different—they involve the study of network systems, not the study of humans. People inject work (packets) into the system, but this is done for the purpose of identifying performance bottlenecks, security vulnerabilities, functional deficiencies, and so on. These studies are not for the purpose of understanding humans, and as such, we believe they do not fall under the purview of IRBs.

For a given researcher who believes his or her particular experiment might be subject to IRB review, the following factors are relevant to the IRB's consideration. First, approval is typically granted as long as the experiment has some research value and the impact on users is negligible; a well-designed experiment should not disrupt users in any measurable way. Second, the arguments against the need to obtain informed consent are compelling: (1) in many cases it is not practical to obtain such consent without jeopardizing the validity of the results; (2) by policy, any incidental user-oriented data that might be collected must be carefully protected; and (3) by policy, any user inconvenienced by the experiment has recourse—they can elect to opt-out of the experimental service.

Regarding point (1), the impractical nature of consent is implicit in the Internet's design—user traffic is potentially carried by multiple Internet Service Providers (ISP), only one of which was explicitly selected by the user. Other ISPs provide transit service without consent. An experimental service is no different than a commercial service in this respect—it provides a (logical) transit service that may be unaffiliated with either the source and the destination of the traffic. Of course, there are also cases where user consent is either explicitly given (e.g., the user elects to use an experimental service) or implied (e.g., the user interacts with an "agent" of a content provider whenever a content provider elects to publish its content through some experimental service).

Regarding point (2), a commercial transit service may collect data about the traffic it carries for the purpose of improving the offered service; an experimental service may collect and analyze traffic for the same purpose (see next subsection). The only difference between a research-provided service and a commercially provided service is that the former is more likely than the latter to publish a paper describing the service. Both, however, are obligated to maintain the confidentiality of any incidental data collected about the users that generated that traffic. Again, PlanetLab's policy is that any such data must be carefully protected.

6.4 Privacy

PlanetLab's policy is that any incidental information collected about individuals in the course of running an experiment must be protected from public disclosure. Analysis can be done on this data to improve a service—as with a commercial service—and the results of this analysis can be published, but any data that directly or indirectly reveals personal information must not be made publicly available. This position is a straightforward application of the privacy expectation placed on all network carriers, as defined by the Electronic Communications Protection Act (ECPA) and the Stored Communications Act (SCA) [3, 7]. It is not specific to network research.

Beyond this position, there are three other points to make. First, various approaches to analyzing data require more thought. One is to post only anonymized data, but whether or not anonymization software adequately protects this data is still open to debate. A second is to establish a community gatekeeper that permits access to raw data to only those researchers that meet certain criteria, but the CS community has yet to define any such protocols. A third is for the researchers that collect the data to do analysis on behalf of the community, but this puts a burden on those researchers and they currrently no incentive in providing such a service.

Second, network data, such as topology maps, can be published. We view this information as already public. Said another way, we view the machines connected to the Internet as "publicly visible," and so by analogy, observing properties of their connectivity is no different than observing traffic patterns on a public street—there is no need for consent to probe these machines and there is no private information to protect.

Third, while our primary focus is on experiments that run on top of PlanetLab, this experimental traffic also "flows through" PlanetLab, the platform. PLC records only per-flow byte and packet counts for auditing purposes (i.e., to determine what slice is responsible for a given packet). This data includes only summary information for outbound flows and PLC allows only limited queries on this data. Whether there is risk of revealing information about individuals requires further analysis.

6.5 Other Legal Issues

Three other legal issues impact PlanetLab. First, one of the rationales behind ECPA and SCA is the desire to limit government access to data collected on telecommunications networks—data that is made public can also be read by government officials. On the other hand, there are certain times when making such information available to law enforcement officials is in the public interest, and to this end, the Communication Assistance for Law Enforcement Act (CALEA) [4] dictates that telecommunication providers must be able to comply with wiretap orders. Similar provisions are made for National Security Letters authorized by the Patriot Act. The question, then, is whether it is necessary for researchers to build their experimental services so they too can respond to such orders. If researchers have recorded useful information in the logs they keep, they should certainly make this information available, assuming they are so advised by their institution's counsel. On the other hand, engineering an experimental service to be CALEA-compliant with the same efficacy as a commercial telecommunications provider places an undue burden on researchers. It is not reasonable to expect such compliance, and to the extent a given experiment is pressed to implement a wiretap, the option of terminating the experiment at that time is always available.

Second, there is a concern that an experiment might inadvertently damage, or cause loss to, Internet users, in violation of the Computer Fraud and Abuse Act (CFAA) [6]. For example, an experiment evaluating defenses against malware might accidentally release the malware on the Internet. As another example, a well-intentioned experimental service might be abused for a malicious purpose, and so indirectly enable a cyber crime. With respect to the first example, it would be a violation of PlanetLab policy to knowingly risk launching a virus or DoS attack. Regarding the second example, we cannot expect researchers to anticipate all possible abuses of their experimental services, but they should take reasonable steps to prohibit known abuses (e.g., learn from the best practices of similar systems [20]), and they should respond quickly to complaints that bring new abuses to light.

Third, issues of the law are complicated by the fact PlanetLab is global, meaning it spans multiple jurisdictions. This report focuses primarily on US law, which represents the vast majority of our experiences to date, the primary exception being the interpretation of UK law described in Section 3.2.

7 Concluding Remarks

One major conclusion we can draw from this experience is that it is necessary to proactively manage relationships with all the stakeholders—individual researchers, research advocacy groups, University CIOs and their security response teams, commercial ISPs, national education and research ISPs, content providers, and national computer security response teams (CERT/CSIRT). This report is intended to contribute to that dialogue.

Acknowledgements

This paper has benefited greatly from conversations with Helen Nissenbaum, Ed Felten, Michael Wonham, Joseph Hall, Bruce Maggs, and Clayton Marsh.

References

[1] GENI: Global Environment for Network Innovations. `http://www.geni.net`.

[2] PlanetLab Bibliography. `http://www.planet-lab.org/biblio`.

[3] Electronic Communications Privacy Act. 18 U.S.C., Section 2510-2522, 1986.

[4] CALEA: First Report and Order. FCC 05-153, September 2005.

[5] TITLE 45 PUBLIC WELFARE, DEPARTMENT OF HEALTH AND HUMAN SERVICES, PART 46, PROTECTION OF HUMAN SUBJECTS. Code of Federal Regulations, June 2005.

[6] Computer Fraud and Abuse Act. 18 U.S.C., Section 1030, 2007.

[7] Stored Communications Act. 18 U.S.C., Section 2701-2711, 2007.

[8] Burstein, A. Conducting cybersecurity research legally and ethically. *USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET '08)* (Apr. 2008).

[9] Department of Health and Human Services. Health Insurance Portability and Accountability Act. `http://www.hhs.gov/ocr/hipaa/`.

[10] Gummadi, K., Saroiu, S., and Gribble, S. King: estimating latency between arbitrary internet end hosts. In *Proceedings of the 2nd ACM Workshop on Internet Measurement* (Marseille, France, 2002).

[11] Huang, M., Bavier, A., and Peterson, L. PlanetFlow: Maintaining Accountability for Network Services. *ACM SIGOPS Operating Systems Review 40*, 1 (Jan 2006).

[12] Information Commissioner's Office. Data Protection Good Practice Note. Code of Federal Regulations, June 2005.

[13] Park, K., Pai, V. S., Lee, K.-W., and Calo, S. Securing web service by automatic robot detection. In *Proceedings of the 2006 Usenix Annual Technical Conference* (Boston, MA, June 2006).

[14] Peterson, L., Bavier, A., Fiuczynski, M., and Muir, S. Experiences Building PlanetLab. In *Proc. 7th OSDI* (Seattle, WA, Nov 2006).

[15] Piatek, M., Isdal, T., Anderson, T., Krishnamurthy, A., and Venkataramani, A. Do Incentives Build Robustness in BitTorrent? In *NSDI* (2007).

[16] PLANETLAB CONSORTIUM. PlanetLab Membership Agreement. http://www.planet-lab.org/consortium/.

[17] SICKER, D., OHM, P., AND GRUNWALD, D. Legal issues surrounding monitoring during network research. In *In Proceedings of Internet Measurement Conference (IMC)* (October 2007).

[18] SOLTESZ, S., POTZL, H., FIUCZYNSKI, M., BAVIER, A., AND PETERSON, L. Container-based Operating System Virtualization: A Scalable, High-Performance Alternative to Hypervisors. In *Proc. EuroSys 2007* (Lisbon, Portugal, Mar 2007).

[19] SPRING, N., BAVIER, A., PETERSON, L., AND PAI, V. S. Using PlanetLab for Network Research: Myths, Realities, and Best Practices. In *Proc. 2nd WORLDS* (San Francisco, CA, Dec 2005).

[20] WANG, L., PARK, K., PANG, R., PAI, V. S., AND PETERSON, L. Reliability and Security in the CoDeeN Content Distribution Network. In *Proc. USENIX '04* (Boston, MA, Jun 2004).