

Neuroscience runs on GNU/Linux

Michael Hanke^{*1,2,3} and Yaroslav O. Halchenko^{*1,2}

¹*Center for Cognitive Neuroscience, Dartmouth College, Hanover, New Hampshire, USA*

²*Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire, USA*

³*Department of Experimental Psychology, University of Magdeburg, Magdeburg, Germany*

June 7, 2011

Running title: GNU/Linux prevalence in neuroscience computing

Correspondence

Dr. Michael Hanke & Dr. Yaroslav O. Halchenko
Dartmouth College
Department of Psychological and Brain Sciences
419 Moore Hall
Hanover, NH, 03755, USA
michael.hanke@gmail.com
yaroslav.o.halchenko@onerussian.com

Acknowledgements

We are grateful to Andrew Connolly, Swaroop Guntupalli, Rajeev Raizada, and Jim Haxby for their support and feedback on the survey and this manuscript.

*Both authors contributed equally to this article.

In response to a recent grant application for a software development project, we received some reviewer comments that questioned the prevalence of GNU/Linux systems as a computing platform in neuroscience. Moreover, a concern was raised that virtualization¹ is not a feasible solution to overcome limitations of any particular platform or to provide a convenient multi-platform working environment. We were surprised by these comments, because they are in contrast to what we experience daily while working with software developers worldwide to integrate neuroscience software into the NeuroDebian² project.

In an attempt to replace subjective experience with facts, in May 2011, we conducted an online survey³ in which we asked neuroscientists to share some details about their computing environments. We tried to avoid a selection bias among participants by using an uninformative short-URL in the request for participation, and posted it to numerous neuroscience-related mailing lists, including thematic lists (*e.g.*, **comp-neuro**⁴, **connectionist**⁵), as well as project-specific mailing lists for popular, cross-platform tools in various subfields of neuroscience. Within 12 days, a total of 583 participants from 44 countries responded to the survey (three empty submissions were removed and 14 additional submissions were excluded after being identified as identical duplicates, sent from the same machine in short succession).

In the survey we asked participants to describe three computing environments they might be using: *personal* – a system with an operating system of their own choice, where they have permission to install arbitrary research software; *managed* – an environment that is provided and maintained by someone else (*e.g.*, dedicated IT staff), without general permission to install arbitrary software; *virtual* – an environment that runs in a virtual machine, possibly with multiple instances of operating systems running simultaneously on the same hardware.

The most striking result was that GNU/Linux-based operating systems are the most commonly reported computing platform in our sample of neuroscience researchers. A total of 68% (95%-*CI* [64, 72]) of all participants reported to be using such an OS in at least one of the described computing environments. For comparison, this statistic yields 52% for Windows (*CI* [48, 56]) and 26% for Mac OS X (*CI* [23, 30]). This figure was even higher for participants that described themselves as “*developing software to be used by other researchers*” (n=237; 75% Linux users, *CI* [69, 80]), than other non-developer participants (n=346; 64% Linux users, *CI* [58, 68]). Moreover, the prevalence of GNU/Linux was evident across researchers working with any of the 9 different data modalities that were assessed by the survey (magnetic resonance imaging, *CI* [70, 79]; magneto/electro-encephalography, *CI* [61, 73]; electrophysiology, *CI* [55, 71]; behavioral data, *CI* [59, 70], simulations, *CI* [65, 78]; remaining modalities: *CI* [68, 84]).

Taking a closer look at the three types of computing environments, the survey revealed that in both, *personal* (n=566) and *managed* (n=371) computing environments

¹http://en.wikipedia.org/wiki/Hardware_virtualization

²<http://neuro.debian.net>

³<http://neuro.debian.net/survey/2011>

⁴<http://www.neuroinf.org/mailman/listinfo/comp-neuro>

⁵<https://mailman.srv.cs.cmu.edu/mailman/listinfo/connectionists>

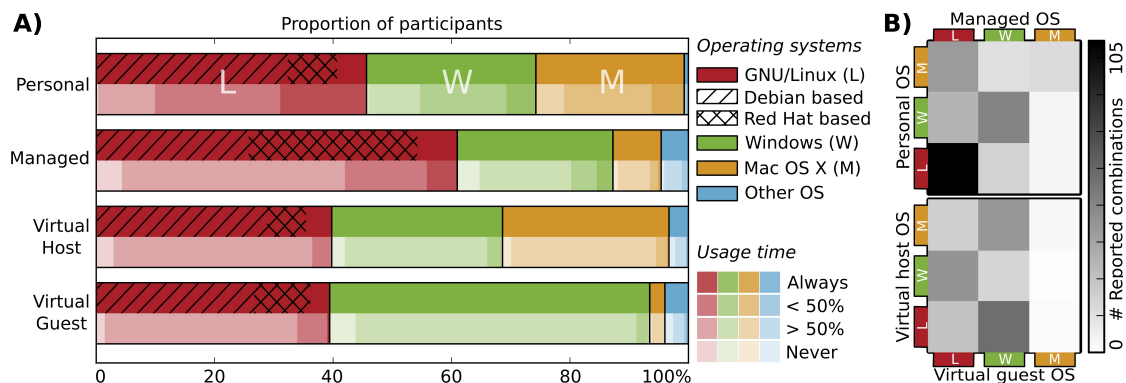


Figure 1: A) Distribution of operating systems (OS) by computing environment type. Each horizontal bar indicates the proportion of survey participants that reported to be using the respective operating system. The lower half of each bar displays the distribution of participants’ responses to the question “*What fraction of your research activity time do you spend in this software environment as opposed to any other environment that you might have access to?*” (four-level answer). B) Total number of reported combinations of operating systems in *personal* and *managed* environments (top), as well as host and guest operating systems of a *virtual* environment (bottom). Only those OS combinations were considered in this figure that were reported to be actually used, and not those that were merely indicated to be available to a particular researchers.

the largest fraction of participants use a GNU/Linux OS. In the *managed* environment it is a majority of 61% (*CI* [56, 65]; Figure 1A). The data showed an expected difference between *personal* and *managed* computing environments: while 92% of all participants reported to use laptops, commodity desktops, or workstations for their *personal* environment, 68% indicated to be using high-end workstations, compute clusters, or grid/cloud computing infrastructure in *managed* environments. Of all participants that reported to be using both *personal* and *managed* computing environments, the majority is using a GNU/Linux OS in both environments, followed by researchers that exclusively use Windows (Fig. 1B top panel).

We also asked participants to rate their individual computing environments regarding various aspects by indicating how much they agree to a particular statement (4-level answer: definitely agree/mostly agree/disagree, encoded as evenly-spaced numerical values within $[-1, 1]$; disagreement being negative). We analyzed the data with respect to differences between users of individual operating systems – grouped into the major families: GNU/Linux, Windows⁶ and Mac OS X⁷. All statistical analyses were implemented as ANOVA contrasts, and every test report includes the 95% confidence intervals of parameter estimates for each OS group. First and foremost, we found that preference of GNU/Linux, Mac OS X, or Windows was *not* differentially motivated by adequacy of

⁶<http://www.windows.com>

⁷<http://www.apple.com/macosex>

hardware support, and availability of free support via web-forums and similar channels. Interestingly, we observe that, in comparison to Mac OS X and Windows, GNU/Linux users are more likely to prefer this OS, because they see themselves as *“having the necessary technical skills to maintain this environment themselves”* ($t=2.796$, $df=553$, $p<0.01$, CI_L [0.41, 0.54], CI_M [0.24, 0.42], CI_W [0.27, 0.44]). Moreover, we found that they also, more than Mac- or Windows-users, see the *“variety of available research software”* as a reason for their platform preference ($t=4.456$, $df=552$, $p<0.001$, CI_L [0.44, 0.57], CI_M [0.24, 0.41], CI_W [0.20, 0.36]). There is evidence that Windows-users tend to be more exposed to *vendor lock-in* situations. They, more than users of any other major operating system, indicate that they *“rely on a particular application that runs in this environment only”* ($t=3.245$, $df=553$, $p<0.001$, CI_L [-0.18, -0.03], CI_M [-0.17, 0.03], CI_W [0.01, 0.20]). At the same time, researchers using Windows in *managed* environments are less likely to agree that it *“provides them with the best available tools for their research”*, in comparison to GNU/Linux and Mac OS X ($t=3.802$, $df=361$, $p<0.001$, CI_L [0.22, 0.36], CI_M [0.12, 0.50], CI_W [-0.10, 0.12]). Moreover, these researchers are also less likely to agree that *“the support staff solves all their technical problems and addresses their demands in a timely fashion”* ($t=2.248$, $df=360$, $p<0.05$, CI_L [0.09, 0.24], CI_M [-0.01, 0.41], CI_W [-0.12, 0.12]) and that *“there are always enough licenses for essential commercial software tools”* ($t=1.700$, $df=354$, $p<0.05$, CI_L [0.03, 0.17], CI_M [-0.09, 0.29], CI_W [-0.14, 0.08]).

Even though we expected hardware virtualization to be a common tool on computing platforms in neuroscience, we were still surprised by the survey results. A total of 44% (CI [40, 48]) of all participants reported to be using virtual machines (VM) as part of their research activities. Among these participants, we could identify two distinct usage patterns. The majority (79%) uses different OS types for host and guest OS (*i.e.*, Windows on a Mac). However, about 21% of all participants report to be using the same OS family inside and outside a VM (Fig. 1B bottom panel). The latter group of researchers is more likely to state that a VM provides them with the ability to *“to easily create a snapshot of a whole analysis environment”* ($t=3.678$, $df=259$, $p<0.001$, CI_{same} [-0.05, 0.27], CI_{diff} [-0.30, -0.14]) and that they *“can take a complete analysis environment with them and run it on different machines”* ($t=3.073$, $df=261$, $p<0.01$, CI_{same} [0.08, 0.43], CI_{diff} [-0.14, 0.04]). On the other hand, researchers using different OS inside and outside the VM are more likely to state that a VM allows them to *“run software that is otherwise incompatible with their system”* ($t=5.047$, $df=262$, $p<0.001$, CI_{same} [0.03, 0.34], CI_{diff} [0.56, 0.72]). Windows is the most frequently used operating system inside a VM. Windows-users themselves most often run a GNU/Linux OS inside a VM (Fig. 1A).

We believe that these results provide ample support for considering GNU/Linux as the current standard computing environment in neuroscience research. Apparently, this platform has come a long way from being a playground for technically skilled “geeks”, to a robust and reliable environment for day to day research activities. While it used to be that using GNU/Linux was only feasible with a good local system administrator, today the situation has changed and Windows or Mac OS X users are more likely to report *“many of my colleagues use something similar”* as a motivation for their platform

choice than GNU/Linux users. ($t=3.779$, $df=552$, $p<0.001$, CI_L $[-0.05, 0.08]$, CI_M $[0.14, 0.30]$, CI_W $[0.06, 0.22]$). Moreover, excluding all reports from dedicated IT-staff, GNU/Linux users report the lowest average time they need to invest in maintenance of their personal computing environment (5.77 h/month). This is not significantly different from the average investment of a user of Apple’s Mac OS X (6.44 h/month) – a system that is widely known for its ease of maintenance. In contrast, Windows-users spend on average 13.97 h/month on system maintenance, significantly more than GNU/Linux or Mac OS X users ($t=3.356$, $df=497$, $p<0.001$, CI_L $[1.92, 8.02]$, CI_M $[1.70, 9.54]$, CI_W $[9.49, 16.97]$; again excluding all system administrators).

It is our impression that despite a clear user preference, commercial software and hardware vendors often do not provide adequate support for the GNU/Linux platform. GNU/Linux is often perceived as a huge heterogeneous family of distributions that is impossible to support as a whole. However, our data show that the vast majority of all GNU/Linux-based neuroscientists use only two flavors of this platform: Red Hat⁸-based, and Debian⁹-based GNU/Linux distributions, with a preference for Debian-based systems in the *personal* environment (Fig. 1A). Both flavors are known to offer a stable platform with predictable release cycles. Moreover, they follow common standards (*e.g.*, Linux Standard Base¹⁰, Freedesktop¹¹) that make them very similar to each other (despite a different native package format). We think that it is both feasible and in the interest of vendors to make themselves familiar with the GNU/Linux platform and support it for their products.

All survey data, as well as results of supplementary analyses are publicly available on the survey website¹². We plan to run a future version of this survey in spring of 2012 to track changes in this field and further investigate the details of the widespread use of virtualization in neuroscience research.

⁸<http://www.redhat.com>

⁹<http://www.debian.org>

¹⁰<http://www.linuxfoundation.org/collaborate/workgroups/lsb>

¹¹<http://www.freedesktop.org>

¹²<http://neuro.debian.net/survey/2011>