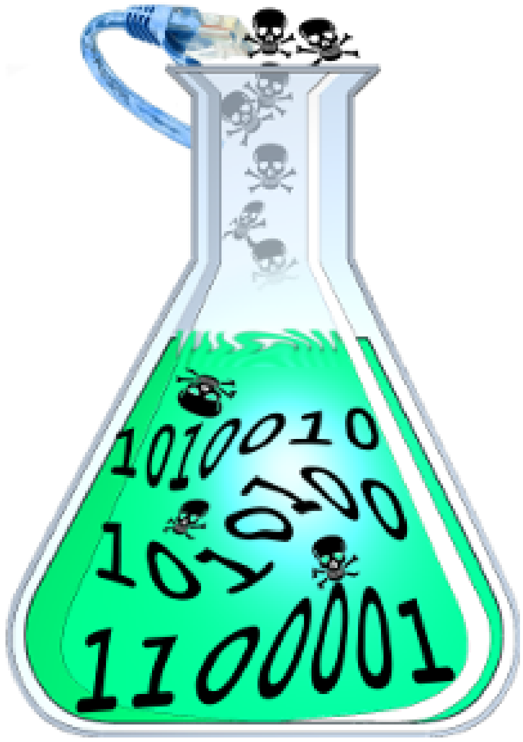
Labtainer Framework Development Guide

Fully provisioned cybersecurity labs

April 14, 2020



Contents

1 Introduction

This document is intended for use by developers who maintain the Labtainer framework. It is also applicable to lab designers who wish to follow Labtainers configuration management and testing conventions for their labs. It does not address the mechanics of lab creation, which are covered in the *Labtainers Lab Designer User Guide*.

Note: The Labtainer framework is developed within and for Linux environments using the command line.

The procedures described herein assume development occurs on a Linux VM that itself is hosted on a Linux platform using VirtualBox. That underlying Linux platform also hosts *test VMs* that will run regression tests. Other configurations are certainly possible, but they would require the developer to potentially alter procedures and/or scripts.

1.1 Linux host installation

The host platform should include VirtualBox (to host the Development VM and test VMs), and Docker, (to host a set of test registries). The host platform should have a directory named SEED that will be shared by each of the VMs. If the host is to publish distributions to the NPS website, then it should have an ability to transfer files to

`davs://nps.edu/webdav/c3o-staging/document_library/labtainers`

2 Development VM Installation

This section describes installation of software on the development VM, which should have at least 150 GB of disk.

2.1 Developer Software Prerequisites

Labtainers is primarily implemented using python3. The containers within a lab include python2 scripts that are part of the framework, e.g., functions that collect student artifacts. The following packages are required on a Linux distribution to support Labtainer framework development.

- **git**
- **make**
- **g++**
- **Latex** (texlive-extra)
- **Docker (Community Edition)** [See **Docker Installation** Section 2.4]
- **pip3** `apt-get install python3-pip`
- **dateutil** `pip3 install py-dateutil`
- **xdotool**
- **VirtualBox**

The VirtualBox product is used to to run Labtainer VMs for testing. Currently, tests are performed on Ubuntu16 and Ubuntu18 VMs, the former tests backwards compatibility of the frameworks python3 support.

2.2 Getting Labtainers from Github

In a Linux terminal, change the working directory into the directory you want to store Labtainers.

Run this in the terminal:

```
git clone https://github.com/mfthomps/Labtainers.git
```

2.3 Setting up the Development Environment

- Disable any auto-updates on your machine as this may interfere with 'apt-get' requests you may have during development.
- Modify your /.bashrc file.
  1. Add **LABTAINER_DIR** as an environment variable and set its value as the path to the `/Labtainers` directory.
  2. Modify the $PATH to include `./bin` and `$LABTAINER_DIR/scripts/designer/bin`.
  3. In summary, your /.bashrc should include something like this:

```
export LABTAINER_DIR=$HOME/Labtainers
export TEST_REGISTRY="YES"
if [[ ":$PATH:" != *":./bin:"* ]]; then
  export PATH="${PATH}:./bin:$LABTAINER_DIR/scripts/designer/bin"
fi
```

- cd into $LABTAINER_DIR/setup_scripts:
  - Run pull-all.py to get all base docker images.
  - Run build-docs.sh to build the lab manuals for all labs.
- cd into $LABTAINER_DIR/tool-src/capinout and run `./mkit.sh`
- Add the vbox share group using `setup_scripts/vbox-share.sh`
- Map the SEED directory on the Linux host as a shared folder. This directory is used to share distribution files between the development system and the test VMs. Accept defaults so its name is

```
/media/sf_SEED
```

2.4 Docker Installation

For full and convenient installation of Docker and setting of Docker privileges, run 'install-docker-ubuntu.sh' in 'setup_scripts', assuming you are developing in Ubuntu. **Note:** Make sure to run the script as user (not sudo), so that your user can be added to the Docker group.

Reboot the system, so that user receives Docker privileges.

- If on a different Linux distribution look in the same folder for your corresponding distribution (CentOS, Debian, Fedora).

- If your Linux distribution is none of these, please view Docker's webpage documentation to learn how to install Docker and set Docker privileges on your machine.
- **NOTE:** These install-docker scripts include the installation of other packages outside Docker that are necessary for building labs.

3 Framework implementation overview

3.1 Implementation elements

The Labtainer framework implementation is primarily python scripts. A number of the top level scripts share functions found in scripts/labtainer-student/**bin**/labutils.py. The top level scripts are organized as follows:

- **Student**
  - `labtainer` (start) and `stoplab` – In the labtainers-student/bin directory, these run on the Linux host and manage the pulling, starting and stopping of containers. They also coordinate collection of student artifacts.
  - Container scripts – In the labtainers-student/**lab_bin** directory, these execute on containers, e.g., to hook bash and parameterize containers.
- **Instructor**
  - `gradelab` and `stopgrader`– Push student artifacts onto grader container and get assessment results.
  - Container scripts – perform grading functions.
- **Lab designer**
  - Building – rebuild in labtainers-student/bin
  - Publishing labs – labtainers/distrib/publish.py
  - Base Labtainer images – scripts/designer/bin, create and publish the base images.
- **Other**
  - VM appliances – //host_scripts, update and publish VM appliances as OVA files for VirtualBox and VMWare.
  - Regression testing of grading functions is performed by labtainer-instructor/regress.py. Expected results are stored in the labtainer/testsets directory.
  - Regression testing of labs and grading combined: scripts in testsets/bin; data sets are not distributed, they are in `labtainer/simlab/<labname>` Get simlab data sets using

    ```
    git clone https://gitlab.nps.edu/mfthomps/Labtainers-simlab.git/
    ```

3.2 Control flow

Student scripts, e.g., `labtainer`, run from the scripts/labtainer-student directory. That directory also contains the bin/labutils.py, which contains most of the framework functions.

The first time a given lab is run, the `docker create` function is used to create containers. The `docker start` function is then used to start the container, and is used for subsequent starts of the same lab.

When a student container is first started "docker exec" is used to run parameterize.sh on the container.

That script also invokes hookBash.sh, which adds the bash sdtin/stout capturing hook, and adds the startup.sh call into the .profile.

The startup.sh uses a lock to control which terminal displays the instructions. In practice most instructions are now pdf files. The startup.sh invoked by student will source a student_startup.sh if present.

The Student.py script runs when a lab is stopped to collect artifacts and kill lingering monitored processes.

Grading is performed on a separate container built for each lab, derived from the labtainer.grader image.

The `checkwork` function forces a collection of artifacts, and a grader container is then run to perform grading.

3.3 mynotify

The mynotify runs as a service. It is installed from the labtainer-student/lab_bin directory. It will exit silently if the lab has no notify file in .local/bin. See its log on each container within /tmp/mynotify.log The service uses the Linux inotify service to detect and record access to files.

4 Distribution publishing

The Labtainer framework is distributed via the c3o website as a tar file, or, optionally a VM applicance (both VMWare and VirtualBox). The Docker images are distributed via the Docker Hub.

The labtainer/distrib/mkdist.sh script runs on a Linux VM hosted on windows or Linux, and creates the distribution tar and copies it into a shared folder. The mk-devel.sh script makes the developers version of the tar. From that shared folder, the two tar files are copied to the

```
\\my.nps.edu@SSL\DavWWWRoot\webdav\c30-staging\document\_library"
```

and then "Publish to Live" is performed on the Liferay site.

The distributions are created from a git repos, as described in section 5.

4.1 VM Appliances

Two prepackaged VM appliances are maintained: one for VirtualBox, and one for VMWare. Each include their respective guest additions. The VMs are maintained on a native Linux system using command line utilities, e.g., VBoxManage. The VMs are rigged to update Labtainers, including a pull of baseline images, on each boot until the first lab is commenced. Scripts named "export*" are used to created the appliance files. The scripts re-import into test images, which must be manually tested. The WinSCP script pushes new applicance images to the CyberCIEGE download directory on the C3O web server. (Wine and WinSCP must be installed on the Linux host that manages the VMs.

The VM appliances should be updated or recreated whenever changes are made to Labtairer base images, otherwise, they are not expected to be changed. To revise the VM appliances, use the scripts from host_scripts on on the Linux system that hosts VirtualBox and VMWare to update the VM appliances so they contain the latest baseline images. After the VM starts and updates the baseline images, use:

```
sudo dd if=/dev/zero of=/emptyfile bs=1M
sudo rm -fr /emptyfile
```

to zero unused space and then run

```
./poweroffVB.sh
./compact.sh
```

to compact the VM image. Then export it:

```
./exportVB.sh
```

This will create the appliance OVA image, and will create a test VM from that appliance. The test VM will start. Use that to run ad-hoc tests.

Do the same for vmware.

Then push the images to the web server, in our case this is the nps.box.com account pointed to by the Labtainers web server.

The appliances automatically update the baselines and the Labtainer scripts on boot, so there is only really advantage to doing this for baseline changes, since they take a while to download.

4.1.1 Installation sizes

An initial install, including the base images, requires about 4GB. Installing a larger lab, e.g., snort, requires an additional 1GB. Running bufoverflow added 22M.

5 Source control and Configuration Management

This section describes Labtainers source control and mechanisms to support continuous integration. Labtainers is managed using git, Docker registries, and a set of custom scripts that control rebuilding and publishing of artifacts. Artifacts are published to test environments associated with each development branch of the product. Publishing releases for public distribution occurs after development branches are tested and merged into the master git branch.

5.1 Build artifacts

Labtainers development creates the following artifacts:

- The distribution tar file for students
- A distribution tar file for lab designers (which is a superset of that for students).
- A test script tar distribution containing SimLab scripts. (These come from a separately managed repo.)
- The Docker container images for each lab.

- The Docker container image for the grader.

The tar distributions are created using scripts from the `distrib` directory. The Docker images are built and published to a Docker registry using the publish.py script, which includes file dependency logic to only rebuild images when one of their sources change. By default, the publish.py script pushes to a local registry rather than to the DockerHub. Updated images are pushed to the DockerHub as part of publishing a new revision of Labtainers.

Currently, there is no attempt to archive artifacts, i.e., only the latest versions are available on publishing sites.

5.1.1 Build steps

The following steps must be performed for each build to ensure distributions are based on the latest file versions. These steps are implemented with in the `full_build.sh` script.

1. Pull the latest git version of the current branch with `git pull`
2. Refresh branch registry from the premaster (unless building premaster) using `refresh_branch.py`
3. Rebuild and publish Labtainer base images using `scripts/labdesigner/bin/mkbases.py`
4. Rebuild and publish labs using `publish.py`
5. Create distribution tar files with `mkall.sh`
6. Run smoke test on Ubuntu16 and Ubuntu18 machines using scripts in `testsets/bin`

5.1.2 Base images

Changes to base docker images referenced by the lab containers will trigger rebuilds. Base docker images are extended by creating new dockerfiles with ".xtra" file extensions. This lets us add features to a base without rebuilding all previous labs that use that base. While these ".xtra" images are built with docker files managed within the designer/base_dockerfiles directory they are not true base images. Only the true base images are included in the initial distribution. In general, avoid changes to a base docker image because doing so could lead some installations to include two copies of the base image, which are very much larger than most other Labtainer images. Modifcations to an xtra extension image will not affect existing installations that have run some labs. Whenever a new lab is started, if it relies on a newer version of the xtra extension, that will be pulled as needed for the lab container images.

When a lab container image is created, it is labeled with the base image name and its image ID (a checksum generated by Docker). This label is generated by a dockerfile that provides labeling veneer on top of newly created images (see the `relabel` function of the `publish.py` script.

When a lab is started, the framework confirms it contains the appropriate base image. If not, the user is prompted to download it.

5.1.3 Framework versions

The "framework version" is a mechanism for providing compatability between new labs and the framework. This value is indpendent of release identifiers. As a Labtainers lab evolves, it may require additional support from the framework. If a new lab image requires an updated Labtainers framework, then the "framework_version" must be incremented within the bin/labutils.py script **before** the image is built and published. This will prompt users to run update-labtainer.sh prior to running any newer lab image. Also insure that these lines are present in the container dockerfile:

```
ARG version
LABEL version=$version
```

And, be sure to publish the revised framework before publishing the revised lab(s).

5.2 Releases and Container Images

A Labtainers *release* contains the set of artifacts described above. File versions within the tar files of a release are all pulled from the git *master* git branch on the development system. Docker container images within a release are built from a *premaster* git branch as described below, and then pushed to DockerHub via the `refresh_mirror.py` script.

5.2.1 The premaster branch

Labtainers source control management includes a *premaster* branch which shall always be on the workflow of creating new releases. All merges on the path to a release go through the premaster branch. No changes are made to the master branch. The only way the master branch ever is updated is via a merge with the premaster branch, after all of its testing is complete. This approach has two goals: 1) ensure that results of merge conflict resolution are tested prior to inclusion within the master, and allow us to test container images before they are published in a new version.

Container images on DockerHub are pushed from a registry containing images build from the premaster branch. The push occurs during the final merge from the premaster branch into the master branch during a release step. The images within the premaster registry are updated only through a rebuild, i.e., `full_build.sh`. Images are not not pushed from development registries directly to the premaster registry.

It is intended that no changes be made directly to the premaster branch, rather, changes are merged into the premaster from other development branches. Once a merge into the premaster commences, no hotfixes affecting build images should occur until the merge completes and the premaster is merged into the master.

5.3 Development branches

Development of new features and fixes occur within development branches. New branches are made off of the premaster branch, but not during a premaster merge.

5.4 Test registries

The test registries are used to test the premaster and development branches of Labtainers.

Test registries are named by their port numbers (currently, all test registries must reside on the same host). These port numbers are mapped to git branch names. This mapping occurs in the config/registry.config file. The rebuild.py command pulls from the registry associated with the current branch.

All development systems are intended to have the `TEST_REGISTRY` environment variable set to YES so that images are pulled and pushed to the appropriate test registry.

Within the test systems and the development host , i.e., the computer that builds distributions and docker images, update the /etc/docker/daemon.json file to reflect new registries as "insecure".

```
    "insecure-registries": [
        "testregistry:5000",
        "testregistry:5001",
```

```
        "testregistry:5002",
        "testregistry:5003"
        "testregistry:5004"
    ]
```

On the Linux system that hosts the development VMs, create the test registries using `host_scripts/sta`

5.5 Testing

Regression testing occurs within testing VMs that are provisioned from the Labtainer VBox appliance as follows:

- clone (as linked) a smoketest box
- remove /.doupdate
- echo "frank@beans.com" > /.local/share/labtainers/email.txt
- add $HOME/labtainer/trunk/testsets/bin to path in bashrc
- visudo and change sudo etnry to: ALL=(ALL) NOPASSWD:ALL
- apt-get install xdotool
- apt-get install vim
- setup-scripts/prep-testregistry.sh
- touch /labtainer/.dosmoke
- run the `setup_scripts/smoke-profile-add.sh`
- Add the vbox share group using `setup_scripts/vbox-share.sh`
- Map the SEED directory on the Linux host as a shared folder. Accept defaults so its name is

  ```
  /media/sf_SEED
  ```

- Edit the /etc/hosts and /etc/hostname to define a distinctive hostname, e.g., `ubuntu16smoke`.
- Create a directory at `SEED/test_vms/<hostname>`.
- On the development VM, create a script in `testsets/bin` modeled after `test-ubuntu18.sh`
- Modify the `setup_scripts/full_build.sh` script to invoke the new test script.

The setup_scripts/smoketest.py scripts represents the test procedures for Labtainers. It is expected that local repo development branches will be tested prior to pushing them to GitHub. Similarly, the results of premaster branch merges are expected to be tested locally before it is pushed to GitHub. A test run from a fresh pull from GitHub premaster branch is a prerequisite to publishing a new release.

Local bench testing, e.g., using `rebuild` for a small set of labs, depends on the git workspace and the test registry for the current branch. Local branch testing, i.e., use of full_build.sh, uses the local repo. It is up to the developer to ensure that is up to date.

Integration testing pulls from the GitHub repo for the desired branch.

5.6 Merging

Development branches are merged into the premaster branch as part of creating a new release.

- Be sure that any and all new and changed files are commited in the development branch, and these have been tested.
- Refresh the premaster registry to ensure it matches the DockerHub images: `./refresh_mirror.py -r`
- `git checkout premaster`
- `git merge <dev branch>`
- Fix any confilicts
- Rebuild images using the premaster branch source: `./full_rebuild.py`
- Run smoketest.
- Push premaster to GitHub:

```
git push --set-upstream origin premaster
```

Revert to premaster in case of merge issues or other failures using `revert_premaster.sh`.

5.7 Publish new release

Before publishing a new release, one more round of testing is required to ensure the framwork properly pulls from the remote regisry.

- Merge premaster into master
- Created tar files with mkall
- Run smoke tests, the test vm scripts will set the branch to `master`, which should cause them to pull from DockerHub.
- Copy labtainer.tar and labtainer-developer.tar to Liferay.
- Publish Liferay to live
- Push premaster registry to DockerHub: `refresh_mirror.py`

Note there is no synchronization scheme to coordinate between the publishing of tar files and updating the DockerHub. In some cases, there may be a window in which a Labtainer installation gets a distribution that names a lab that is not yet on DockerHub.

5.8 Remove unused branches

Remove local branches with:

```
git branch -d <branch>
```

Or use the `-D` option to force deletion. But, that not needed if the branch was properly merged. Remove remote branches with

```
git push origin --delete <branch>
```

6 Continuous integration with Jenkins

A Jenkins pipeline automates periodic testing of `premaster` branch of Labtainers. The pipeline script is backed up in `tesetsets/bin/jenkins_pipeline.txt`. The pipeline pulls from the premaster branch of the GitHub repo. It builds any changed lab images (**TBD flag those to remind to merge the premaster registry into the master). It then generates the student and designer distributions and uses those to run the smoketest VMs.

6.1 Jenkins installation

The stock Jenkins is installed on the development VM. After installing Jenkins, add the jenkins user to the vboxfs and the docker groups

```
sudo usermod -a -G vboxfs jenkins
sudo usermod -a -G docker jenkins
```

7 Developer guidelines

7.1 Testing and Running Existing Labs

When running labs, the goal is to force ourselves to run the distributed labs unless we have specific reasons to do otherwise. Labtainers will use locally-created container images if they are present – and these may be stale.

A) To ensure that you are running the latest version of the published lab (or version assocated with your current git branch), first delete the current version of the lab using setup_scripts/removelab.sh.

B) If you find the lab to be broken, e.g., missing a file, please attempt to run ”rebuild.py” on the lab. rebuild.py will ouput a log of issues. Report these findings to the lab author.

C) Always run removelab.sh after you have run an existing lab via rebuild.py.

D) Please review the lab's manual very closely. This is so that both the lab itself and the lab's manual can receive feedback for improvement.

8 GNS3 Support

Please refer to the guide in `docs/gns3` for information on integrating Labtainers with GNS3.

9 Notes

9.1 Race condition on checklocal.sh output

If an mynotify.py event causes an output from checklocal.py, that may conflict with concurrent output from checklocal.py resulting from some program/script running. In theory, the program/script should complete its run of checklocal before the program/script actually gets to access the file that triggers a mynotify watch. So, the latter's output to the timestamped file is appended. Further, the mynotify.py looks for an existing timestamped file, and if not found, looks for one from the previous second. This hack is an attempt to keep the outputs merged. It will fail if the access does not happen within a second of the program start. See the acl lab.

9.2 temporal logic considerations

When evaluating results from logfiles containing timestamps use FILE_TS or FILE_TS_REGEX to ensure you get timestamped values for only matching records. Reliance on goals.config to matchany can result in timestamped results that don't corrolate to the desired record.

9.3 parameterizing the start.config

A copy of the parameterized version of start.config is placed into labtainer-student/.tmp/¡lab¿/. This ensures that subsequent runs of the lab always have the same psuedo random values.

9.4 Packaging

The framework has not yet been adapted to use Linux package managers. Currently, scripts are run from a workspace directory and python paths are managed relatively between scripts.

9.5 Todo

Change smoke test to look for email in expected results and set that as the email before starting a lab. Validation should catch results.config naming of non-existent container.

The backups2 lab creates a loopback volume named myfs.img. The lab does not dismount it. This device will go away on a reboot.

Add latex template and makefile when new_lab_setup is run.

9.5.1 Docker problems

The check_nets.py tests for problems that sometimes crop up in Docker. These include Linux routes defined on the host for container networks that no longer exist. And, loopback devices that are not properly deleted? The file-deletion lab fails in a full smoketest, perhaps due to a lingering loopback device? Lab must be completed prior to reboot of the host VM. Reflect that in Lab Manual.

The backups2 lab consumed a loopback device, leaving it define (as seen when running check_nets.sh). This led the file-deletion lab to fail, being unable to get a loopback device. Altered file-deletion to create the "next" loopback device if it does not exist.

Metasploit lab now crashes the VM. g_array_unref: assertion 'array' failed. Leads to X server crash, loss of desktop. Perhaps only occurs after reboot, once, then works ok? Created both containers with NO_PRIVILEGE attribute in start.config, seems to fix it? NO: that breaks it by keeping services from running. Disabled postgresql service in attacker seems to keep the crash from happening? Also happened in a hackazon container derived from a dockerfile that included multiple CMD entries (one in the latest file, one in the parent file).

9.5.2 Grader updates

Automate detection of need to update a local grader image, e.g., in response to a fix to the grader.

9.6 ongoing

Updated framework and grader to use python3. Intent is to not affect existing labs. Need to publish centos-log2 and backups2 due to changes in centos packages. Changed grader and centos.xtra base dockerfiles. Publish along with new update? Will centos-log2 run with old framework? This is begin done in the python3 branch of git.

Python3 changed semantics of randint. Also changed random.seed to take a version number for compatable seeding. Except version 3.5.2 is broken in that a string given as the seed causes a non-deterministic (time?) seed to be used. This bug is fixed in 3.6. Our grader container naturally installs 3.5.2, so we also install 3.6 from dead.snakes ppa and change the links in /usr/bin/python3. The broken 3.5.2 version is also what comes with the Ubuntu used in the original Labtainer VMs. So, we will maintain support for python2 in the framework, and fall back to python2 if we detect 3.5.2.

Some html, e.g., for the softplc, want to visit fonts.googleapis.com. If no gateway/dns is available, there is a long timeout. add ADD-HOST fonts.googleapis.com:127.0.0.1 to start.config to shut it up.

In the VM .profile, move the terminal creation functions to a seperate script run in background – seems .profile must finish or VMWare Horizen borks the Terminal Server startup.

Smoketests still sometimes fail with bad routing tables and/or iptables. Run check_nets.py to test.

Use of Docker cache leads to build dependency errors in which a source file changes but has not actual effect on the image. If the image checksum matches the cache, the date is not altered, and thus the next build will see a false dependency since the source will continue to be newer than the image. Modified building of base images to default to –no-cache; modified building lab images to allow supressing use of cache, though still defaults to use of cache.

Dependence on LABTAINER_DIR introduced with imodules, and will spread and backfill from there. However, current installations have an invalid value for that env variable. An updated update-labtainer.sh fixes that. But update must run twice for it to take effect within the bashrc. After the first run of update-labtainer, the imodule function will be available, but without the value being set. Add diagnostic to imodule to prompt for re-run of update-labtainer. Or just hack around it?

9.7 Build dependence

The home_tar and sys_tar files as considered for build dependence. But the home.tar and sys.tar are skipped because they are remade from a fresh pull. This is a problem when we add dated archives. Must rebuild with a -f.

9.7.1 IModules

The DoRebuild function defines a container_registry, which is then used to query info about this image, but is also used to define the registry within the Dockerfile, where the base is pulled, i.e.,the FROM statement. It seems there needs to be a BASE_REGISTRY as well as a REGISTRY, with the former defaulting to the default registry per LabtainerConfig.

How do we manage name conflicts between labs? That works natrually by precidence. But between base images? Where a designer wants base images from multiple registries, how are they named within the test registry since that collapses all registry designators into the one test?

9.7.2 Base images IDs

The base image information placed in labels of lab container images include the registry name for which the image was built. Since we push premaster registry images to DockerHub, the public lab images have a label reflecting the premaster test registry. This is worked around with in the InspectRemoteReg.py.

9.8 Other bread crumbs

If you get the dreaded "docker.service: Start request repeated too quickly.", then: sudo systemctl daemon-reload sudo systemctl restart docker