

Machine Learning

finding Linux bugs

Stef Walter, Rado Vbrowsky

Red Hat

“Any sufficiently complex system will have bugs. Bugs are entropy, and entropy is fundamental to this universe.”

— Martin Pitt

Any sufficiently complete testing system
will be "plagued" by test flakes

Cockpit Integration Tests

```
Real Failures:  +  42,764
Known Issues:   +  54,371

Test Flakes:    +  25,420
-----
Test Failures:   122,555
-----
Total Runs:      2,466,742
```

Data from last 90 days:

<https://images-cockpit.apps.ci.centos.org/tests-train-1.jsonl.gz>

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG
PILE OF LINEAR ALGEBRA, THEN COLLECT
THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL
THEY START LOOKING RIGHT.



This is a source of training data

```
{
  "pull": "https://api.github.com/repos/cockpit-project/cockpit/pulls/9230",
  "revision": "2e4fd7d309b7f1c591c099e4041f8db20459c5ae",
  "status": "failure",
  "context": "verify/fedora-27",
  "date": "2018-05-24T15:48:53Z",
  "merged": true,
  "test": "testTeamActive (check_networking_team.TestNetworking)",
  "url": "http://fedorapeople.org/groups/cockpit/logs/pull-9230-20180524-150",
  "tracker": "https://api.github.com/repos/cockpit-project/cockpit/issues/89",
  "log": "...Error: timeout\nwait_js_cond(ph_in_text(\"#network-interface .p",
}
```

90+ APIs: File, Command, REST, DBus, Socket

abrt	AppStream	apt-get	atomic	Candlepin /candlepin/
chpasswd	CloudForms	cryptsetup	curl	dbus-daemon
device-mapper	docker-storage-setup	docker	e2fsprogs	etcd
/etc/kdump.conf	/etc/passwd	firewalld	FreeIPA	GnuTLS
GSSAPI	hostnamed	ipa-client	ipa-client	iproute
iptables	iscsi-tools	journalctl	kdump	klist
krb5	Kubernetes /api/	lastlog	libvirt	loginctl
lvm	mdadm	NetworkManager	NetworkManager-te...	oddjob
Openshift /oapi/	Openshift OAuth2	openssl	ostree	oVirt /api/
PackageKit	passwd	PCP	PolicyKit	/proc/meminfo...
/proc/mounts	/proc/net/dev...	procps-ng	/proc/stat	pwquality
qemu	realmd	rm ...	rpmostreed	rpm
selinux-policy-target...	selinux-utils	setenforce	Setroubleshootd	shadow-utils
shutdown	sosreport	ssh-agent	sshd	ssh-keygen
ssh	sssd	storaged	subscription-manager	sudo
/sys/fs/cgroup	/sys/kernel	/sys/power	systemd	timedated
Tuned	udev	UDisks2	/usr/bin/kubectl	/usr/bin/timedatectl
/usr/bin/virt-install	/var/log/wtmp	/var/run/utmp	virsh	who/w
xfspgrog	yum			

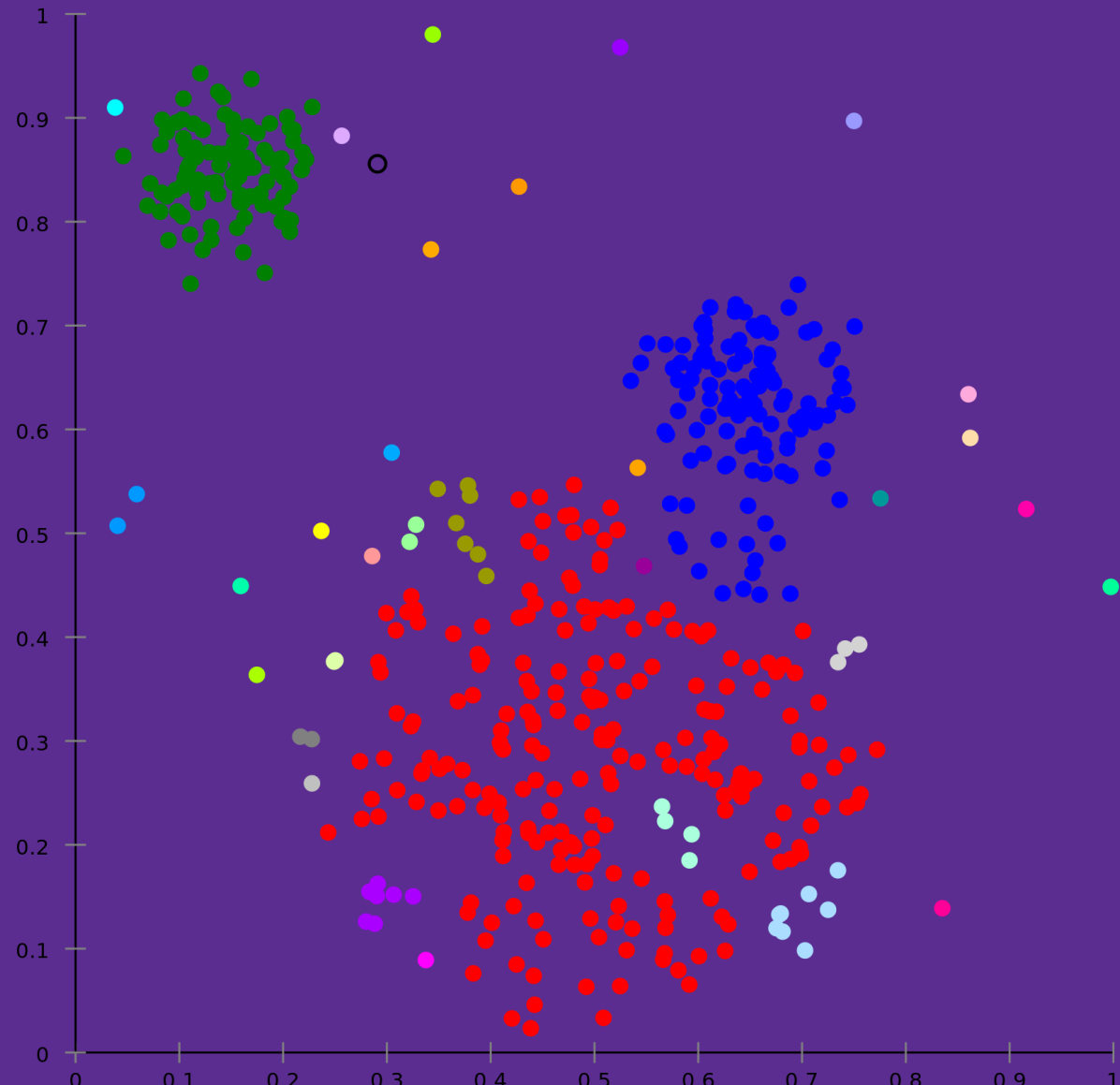
Testing flakes are those bugs. That entropy.

The tests are essentially fuzzing Linux.



Machine Learning

Cluster the test failures



ML techniques in use

- Preprocessing
- Term Frequency - Inverse Document Frequency
- Normalized Compression Distance
- DBSCAN unsupervised clustering
- Multi-Dimensional Scaling
- K-nearest Neighbors classification

Raw log input

```
# -----
# testTeam (check_networking_team.TestNetworking)
#
NAME          UUID                               TYPE          DEVICE
System eth0   5fb06bd0-0bb0-7fffb-45f1-d6edd65f3e03  802-3-ethernet eth0
virbr0        82404faf-cb6b-4f4c-8eb6-65fc1a2ff8da  bridge         virbr0
System eth1   9c92fad9-6ecb-3e6c-eb4d-8a47c6f50c04  802-3-ethernet --

52:54:01:00:00:03 -> eth2
52:54:01:00:00:03 -> eth2
not ok 139 testTeam (check_networking_team.TestNetworking) duration: 106s
Traceback (most recent call last):
  File "/build/cockpit/bots/./test/verify/check-networking-team", line 81, in
    b.wait_present("#network-interface-slaves tr[data-interface='%s']" % iface)
  File "/build/cockpit/test/common/testlib.py", line 230, in wait_present
    return self.wait_js_func('ph_is_present', selector)
  File "/build/cockpit/test/common/testlib.py", line 224, in wait_js_func
    return self.phantom.wait("%s(%s)" % (func, ','.join(map(jsquote, args))))
  File "/build/cockpit/test/common/testlib.py", line 821, in
    return lambda *args: self._invoke(name, *args)
  File "/build/cockpit/test/common/testlib.py", line 847, in _invoke
    raise Error(res['error'])
Error: timeout

Wrote TestNetworking-testTeam-rhel-7-127.0.0.2-2601-FAIL.png
Wrote TestNetworking-testTeam-rhel-7-127.0.0.2-2601-FAIL.html
Journal extracted to TestNetworking-testTeam-rhel-7-127.0.0.2-2601-FAIL.log
Journal extracted to TestNetworking-testTeam-fedora-26-127.0.0.2-2602-FAIL.log
```

Extracting the logs

```
# -----
# testTeam (check_networking_team.TestNetworking)
#
NAME          UUID          TYPE          DEVICE
System eth000  xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxx  000-000-ethernet  eth000
virbr000      xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxx  bridge           virbr000
System eth000  xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxx  000-000-ethernet  --

000:000:000:000:000:000 -> eth000
000:000:000:000:000:000 -> eth000
not ok
Traceback (most recent call last):
File "check-networking-team", line 000, in testTeam
b.wait_present("#network-interface-slaves tr[data-interface='%s']" % iface000)
File "testlib.py", line 000, in wait_present
return self.wait_js_func('ph_is_present', selector)
File "testlib.py", line 000, in wait_js_func
return self.phantom.wait("%s(%s)" % (func, ','.join(map(jsquote, args))))
File "testlib.py", line 000, in
return lambda *args: self._invoke(name, *args)
File "testlib.py", line 000, in _invoke
raise Error(res['error'])
Error: timeout

Wrote png
Wrote html
Journal extracted to log
Journal extracted to log
```

Term Frequency - Inverse Doc Frequency

```
# testTeam (check_networking_team.TestNetworking)
virbr000      xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxxxx  bridge      virbr000
File "check-networking-team", line 000, in testTeam
b.wait_present("#network-interface-slaves tr[data-interface='%s']" % iface000)
```

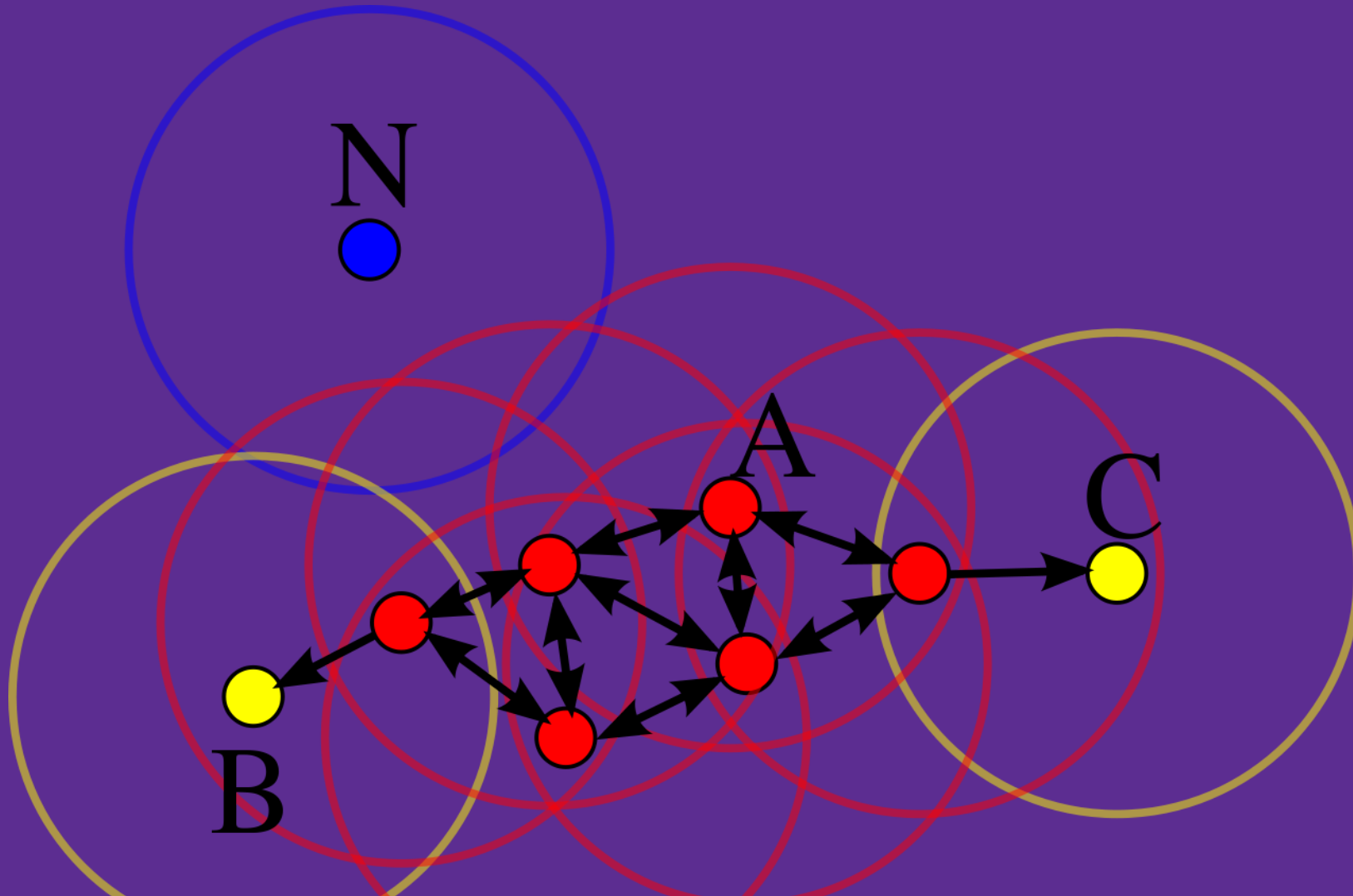
Normalized Compression Distance

$$NCD_Z(x, y) = \frac{Z(xy) - \min\{Z(x), Z(y)\}}{\max\{Z(x), Z(y)\}}.$$

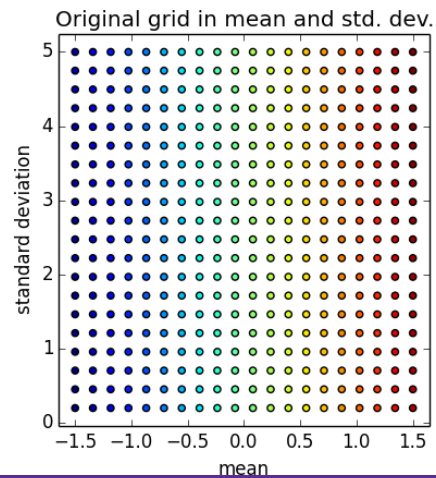
```
Z = lambda v: len(lz4.frame.compress(v))
```

DBSCAN

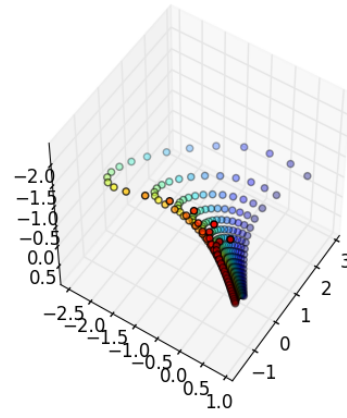
Density-based spatial clustering of applications with noise



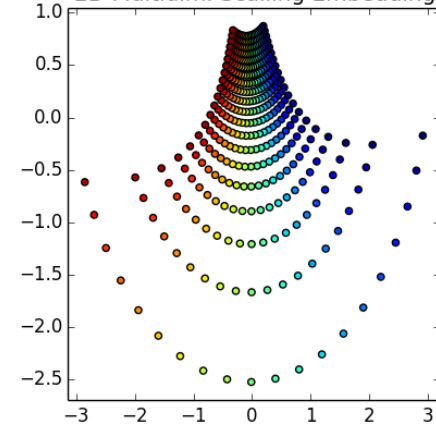
Multi-Dimensional Scaling



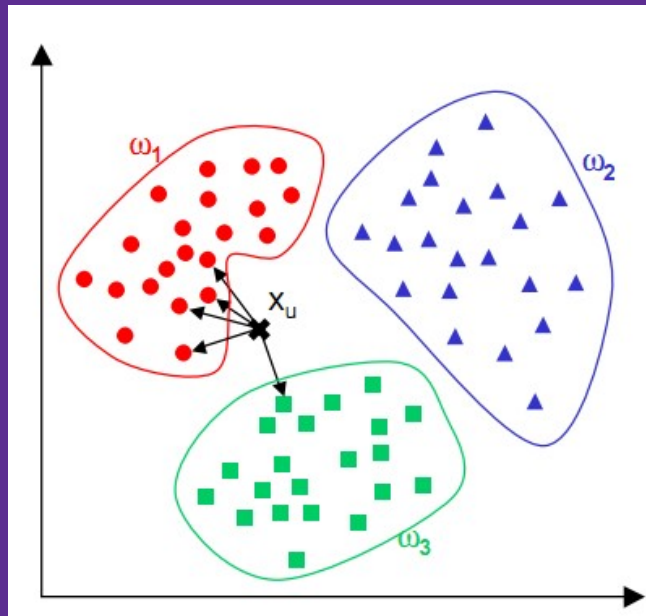
3D Multidim. Scailing Embedding



2D Multidim. Scailing Embedding



k-nearest neighbor





Make it do something

Auto-retrying flakey tests

<http://fedorapeople.org/groups/cockpit/logs/pull-10178-20180928-100000-fae2889d-verify-fedora-29/log.html>

[Result directory](#)

[Raw log](#)

Done: 1 tests failed.

224

22

247 tests, 224 passed, 1 failed, 22 skipped, -3 to go (3 retries).

Failed and skipped tests:

- 13: testPcp (check_metrics.TestMetrics) # duration: 47s # RETRY due to flakiness 📸 screenshot 📖 journal
- 23: testBondingMain (check_networking_bond.TestNetworking) # duration: 89s # RETRY due to flakiness 📸 screenshot 📖 journal 📖 journal
- 23: testBondingMain (check_networking_bond.TestNetworking) # duration: 88s # RETRY due to flakiness 📸 screenshot 📖 journal 📖 journal
- 234: testProjectGroups (check_openshift.TestRegistryPrerelease) # duration: 299s 📸 screenshot 📖 journal 📖 journal
- 1: testBasic (check_kdump.TestKdump) 📸 screenshot 📖 journal -- skipped: Known Issue #9993
- 3: testDashboard (check_kubernetes.TestKubernetes) 📸 screenshot 📖 journal -- skipped: Known Issue #10011
- 7: testTopology (check_kubernetes.TestKubernetes) 📸 screenshot 📖 journal -- skipped: Known Issue #10011
- 9: testDeployDialog (check_kubernetes.TestNuclecule) -- skipped: Nuclecule deploys temporarily removed.
- 39: testTeam (check_networking_team.TestNetworking) 📸 screenshot 📖 journal 📖 journal 📖 journal 📖 journal -- skipped: Known Issue #6108
- 64: testVdoMissingPackages (check_storage_vdo.TestStoragePackages) -- skipped: No vdo available
- 80: testRegister (check_subscriptions.TestSubscriptions) -- skipped: fedora-29: No subscriptions
- 86: testTimeServers (check_system_info.TestSystemInfo) 📸 screenshot 📖 journal -- skipped: Known Issue #9994
- 94: testExternalConsole (check_machines_dbus.TestMachinesDBus) -- skipped: Broken with current chromium, see PR #9229
- 125: testExternalConsole (check_machines_virsh.TestMachinesVirsh) -- skipped: Broken with current chromium, see PR #9229
- 140: testInfoSecurity (check_packagekit.TestUpdates) 📸 screenshot 📖 journal -- skipped: Known Issue #9996
- 141: testInfoTruncation (check_packagekit.TestUpdates) 📸 screenshot 📖 journal -- skipped: Known Issue #9996
- 145: testSecurityOnly (check_packagekit.TestUpdates) 📸 screenshot 📖 journal -- skipped: Known Issue #9996
- 147: testAvailableUpdates (check_packagekit.TestUpdatesSubscriptions) -- skipped: fedora-29: No subscriptions

Annotate test logs

! 85: testSuper (check_reauthorize.TestReauthorize) # duration: 25s 📷 screenshot 📖 journal

```
# testSuper (check_reauthorize.TestReauthorize)
#
ftruncate() failed: Permission denied

DevTools listening on ws://127.0.0.1:9832/devtools/browser/28132994-84e8-492a-bda1-07059595e814
[0123/152927.820842:ERROR:zygote_host_impl_linux.cc(268)] Failed to adjust OOM score of renderer with pid 123004: Permission denied (13)
[0123/152928.864472:ERROR:zygote_host_impl_linux.cc(268)] Failed to adjust OOM score of renderer with pid 123081: Permission denied (13)
> log: done
> log: fail
Traceback (most recent call last):
  File "/build/cockpit/bots/./test/verify/check-reauthorize", line 84, in testSuper
    self.assertEqual(b.text(".super-channel span"), 'result: access-denied')
AssertionError: u'result: disconnected' != 'result: access-denied'

not ok 85 testSuper (check_reauthorize.TestReauthorize) # duration: 25s
Wrote screenshot to TestReauthorize-testSuper-fedora-i386-127.0.0.2-2601-FAIL.png
Wrote HTML dump to TestReauthorize-testSuper-fedora-i386-127.0.0.2-2601-FAIL.html
Wrote JS log to TestReauthorize-testSuper-fedora-i386-127.0.0.2-2601-FAIL.js.log
Journal extracted to TestReauthorize-testSuper-fedora-i386-127.0.0.2-2601-FAIL.log

# Flake probability: 72.5% (neural network)

# Flake likely 66.2% (clustering)
```

Curated set of non-flakey tests

```
0      verify/fedora-27 testBasic (check_ovirt.TestOVirtMachines)
0      verify/fedora-27 testBond (check_networking_bond.TestNetworking)
0      verify/fedora-27 testDeployDialog (check_kubernetes.TestNulecule)
0      verify/fedora-27 testDeployDialog (check_openshift.TestOpenshift)
0      verify/fedora-27 testDeployDialog (check_openshift.TestOpenshiftPrer
0      verify/fedora-27 testDockerCommandInfo (check_openshift.TestRegistry
0      verify/fedora-27 testExternalConsole (check_machines.TestMachines)
0      verify/fedora-27 testIpa (check_realms.TestRealms)
0      verify/fedora-27 testKubevirtMachinesCreate (check_openshift.TestOpe

0      verify/fedora-27 testLuks (check_storage_luks.TestStorage)
0      verify/fedora-27 testPendingClaim (check_kubernetes.TestKubernetes)
0      verify/fedora-27 testRaid (check_storage_mdraid.TestStorage)
50     verify/fedora-27 testBasic (check_apps.TestApps)
50     verify/fedora-27 testBasic (check_networking_basic.TestNetworking)
50     verify/fedora-27 testHostToMaintenance (check_ovirt.TestOVirtMachine
50     verify/fedora-27 testKubevirtMachinesList (check_openshift.TestOpens
50     verify/fedora-27 testProjectGroups (check_openshift.TestRegistryPrer
60     verify/fedora-27 testBasic (check_kdump.TestKdump)
100    verify/fedora-27 testCheckpoint (check_networking_checkpoints.TestNe
100    verify/fedora-27 testDosParts (check_storage_msdos.TestStorage)
100    verify/fedora-27 testISCSI (check_storage_iscsi.TestStorage)
100    verify/fedora-27 testKubevirtMachinesCreate (check_openshift.TestOpe
100    verify/fedora-27 testKubevirtMachinesList (check_openshift.TestOpens
100    verify/fedora-27 testKubevirtVmInstance (check_openshift.TestOpenshi
100    verify/fedora-27 testPendingClaim (check_openshift.TestOpenshift)
100    verify/fedora-27 testPendingClaim (check_openshift.TestOpenshiftPrer
100    verify/fedora-27 testProjectPolicy (check_openshift.TestRegistry)
100    verify/fedora-27 testUndeployKubevirt (check_openshift.TestOpenshift
100    verify/fedora-27 testVSPW (check_openshift.TestOVirtMachines)
```

File issues for "big" bugs

testDockerCommandInfo (check_openshift.TestRegistry) is flaky #10845

Open cockpituous opened this issue on Dec 17, 2018 · 0 comments



cockpituous commented on Dec 17, 2018

Member + 🗨️ ⋮

86% on verify/fedora-29

- pull-10702
- pull-10638

▶ Source material

🔖 cockpituous added the **flake** label on Dec 17, 2018

Assignees

No one—assign yourself

Labels

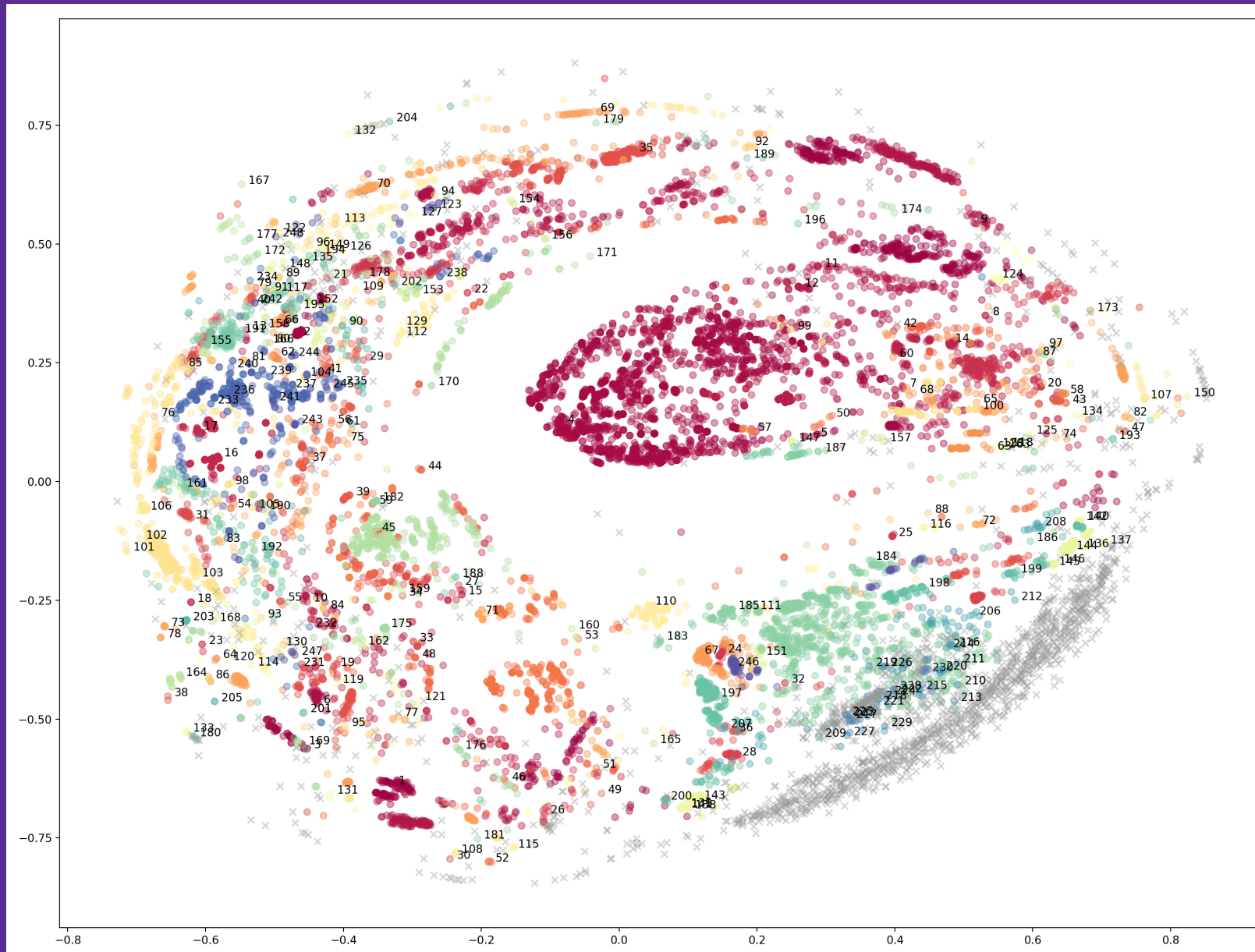
flake

Projects

None yet

Milestone

Visualization



Real bugs

- **Example: Shadow utils usage races**

cluster-7-112.log

<https://github.com/cockpit-project/cockpit/pull/9979>

- **Example: Networkd + SELinux breaks the boot**

cluster-124-4.log

- **Example: Race invalid update of UI state**

cluster-95-28.log

<https://github.com/cockpit-project/cockpit/pull/10127>

- **Example: PackageKit crashing intermittently**



I want!

[https://github.com/cockpit-
project/cockpituous/tree/master/learn](https://github.com/cockpit-project/cockpituous/tree/master/learn)

Use it yourself

```
$ git clone git@github.com:cockpit-project/cockpituous
$ cd cockpituous
```

```
$ ./train-tests -v test-example.jsonl.gz
Loading existing tests data
14208: Items to train
...
100933632: Computed distances in 1181 seconds on 32 cores
332: Clusters (13962 items, 246 noise)
$ ./predict-tests -v test-predict.jsonl
{
  "test": [ [ "testTeamActive", 2 ], [ "testTeamBond", 5 ] ],
  "status": [ [ "failure", 7 ] ],
  "context": [ [ "verify/fedora-27", 3 ], [ "verify/fedora-27", 4 ] ],
  "date": [ "2018-05-24T15:48:53Z", "2018-05-30T01:01:30Z" ],
  "merged": [ [ true, 5 ], [ false, 1 ], [ null, 1 ], ],
  "tracker": [ [ "https://api.github.com/repos/cockpit-project/cockpit/issue"
}
```

Use it with Kubernetes

```
$ kubectl create -f learn/cockpit-learn.yaml
```

```
$ curl --progress-bar --fail --upload-file learn/test-example.jsonl.gz \  
  http://cockpit-learn.project.srv.cluster.local/train/test-example.jsonl.gz
```

```
$ curl -d '@learn/test-predict.jsonl' \  
  http://cockpit-learn.project.srv.cluster.local/predict
```

What's next

- Retrieving data automatically
- Tracking known issues automatically



The Kernel Oraculum

Rado Vbrovsky, Red Hat





Cluster diff 1 (match 41.90%)

ID	Data Dir	Case no.	Tracker no.
1			
Provided calltrace		Suggested calltrace	
f	1dump_stack 2_report_bad_irq 3note_interrupt 4handle_irq_event_percpu 5handle_irq_event 6handle_fasteoi_irq 7handle_irq 8irqtime_account_irq	f	1dump_stack 2_report_bad_irq 3note_interrupt 4handle_irq_event_percpu 5handle_irq_event 6handle_fasteoi_irq 7handle_irq 8sched_clock_idle_wakeup_event 9tick_check_idle
n	9do_IRQ 10common_interrupt 11cpuidle_enter_state 12cpuidle_enter 13call_cpuidle 14do_idle 15cpu_startup_entry 16rest_init 17start_kernel 18set_init_arg	n	10do_IRQ 11common_interrupt 12cpuidle_text_start 13native_safe_halt 14default_idle 15arch_cpu_idle 16cpu_startup_entry 17rest_init 18start_kernel 19repair_env_string

Ludicrous Claim: Any testing system that does not have enough flakes, should introduce more tests or more mutations.



Questions?

cockpit-project.org

#cockpit on FreeNode

Credits:

Clusters: Chire on Wikipedia

CC BY-NC 2.5: Comic by XKCD.com