

Corpus of git commit related data for OpenStack Nova

Copyright Doug Williams - 2014

Updated: 9/3/2014

Sources:

- Launchpad (<https://launchpad.net/nova> (<https://launchpad.net/nova>))
- Gerrit (<https://review.openstack.org/#/q/status:open,n.z> (<https://review.openstack.org/#/q/status:open,n.z>))
- Git (<https://github.com/openstack/nova> (<https://github.com/openstack/nova>))

Structure of Corpus:

- combined_commits.jsonz : Basic record for each commit, integrating data from Launchpad, Gerrit and Git
- all_blame.jsonz : Applies GIT BLAME to the parent commit of every bug fix related commit. Only includes those lines included within the window for diff. Proximity is distance of line within blame for a changed line in a bug fix commit, where proximity=1 for changed lines or lines adjacent to inserts

Note: All .jsonz files are gzipped json, with individual json entries for each record. Apply json.loads() to each individual line.

Examples:

- Sample entry from combined_commits.jsonz
- Sample entry from all_blame.jsonz
- Example of blame normalization

History

- 9/2/2014: Updated corpus, used additional patch information from Jenkins to annotate commits with author jenkins@review.openstack.org.
Updated normalize_blame to handle null entries
- 9/3/2014: Update nova_all_blame, filtering out all entries > 3000 diff lines that were contributing to huge file size

Code to load corpus

```
In [1]: import gzip  
import json  
import pprint as pp
```

```
In [2]: def jload_helper(f):  
    #restore the object  
    result = []  
    dict_result = {}  
    for line in f:  
        v = json.loads(line)  
        if 'json_key' in v:  
            #individual entries are items in a dict:  
            k = v['json_key']  
            del v['json_key']  
            dict_result[k] = v  
            pass  
        else:  
            result.append(v)  
    if dict_result:  
        print 'returning dict'  
        return dict_result  
    elif len(result) > 1:  
        print 'returning list'  
        return result  
    elif len(result) == 1:  
        print 'returning singleton'  
        return result[0]  
    else:  
        return False
```

```
In [3]: def jload(name):  
    if name.endswith('z'):  
        print 'gzipped'  
        with gzip.open(name, 'rb') as f:  
            return jload_helper(f)  
    else:  
        with open(name, 'rb') as f:  
            return jload_helper(f)
```

Sample entry from nova_combined_commits.jsonz

```
In [4]: combined_commits = jload('nova_combined_commits.jsonz')
        gzipped
        returning dict
```

Basic commit entry (without corresponding Launchpad info)

```
In [5]: pp.pprint(combined_commits.items()[0])
(u'0b698186b56af6580633dedd7916df2897945f29',
 {u'author': u'<git.Actor "MORITA Kazutaka <morita.kazutaka@gmail.com>">',
  u'cid': u'0b698186b56af6580633dedd7916df2897945f29',
  u'committer': u'<git.Actor "MORITA Kazutaka <morita.kazutaka@gmail.com>">',
  u'date': 1305808274,
  u'msg': u'Avoid wildcard import.',
  u'parents': [u'5a3bd10a2ada4ab1287b115051578b7b27bba6bc' ]})
```

Entry associated with bug fix. 'g:' prefix is data from gerrit, 'lp:' prefix is data from launchpad. All other fields from git commit

```
In [6]: bug_fix_commits = [k for k,v in combined_commits.items() if 'lp:id' in v]
        print 'commits associated with big fixes:', len(bug_fix_commits)
        print bug_fix_commits[0]
        print
        pp.pprint (combined_commits[bug_fix_commits[0]] )

commits associated with big fixes: 8324
e7068fd481f6cb887da3a8488435d7c911539bb2

{u'author': u'<git.Actor "Dan Prince <dprince@redhat.com>">',
```

```

u'author2': u'<git.Actor "Jenkins <jenkins@review.openstack.org>">',
u'bug': u'962515',
u'change_id': u'Ic1be1459515654d45febd89da58b19e0840aaf9d',
u'cid': u'e7068fd481f6cb887da3a8488435d7c911539bb2',
u'committer': u'<git.Actor "Gerrit Code Review <review@openstack.org>">',
u'date': 1333041489,
u'g:_number': 5957,
u'g:branch': u'master',
u'g:change_id': u'Ic1be1459515654d45febd89da58b19e0840aaf9d',
u'g:created': u'2012-03-29 15:05:29.000000000',
u'g:labels': {u'Code-Review': {u'all': [{u'_account_id': 693,
                                         u'name': u'Christopher MacGown',
                                         u'value': 2},
                                         {u'_account_id': 209,
                                         u'name': u'Josh Kearney',
                                         u'value': 2},
                                         {u'_account_id': 642,
                                         u'name': u'Matt Dietz',
                                         u'value': 2}]},
              u'approved': {u'_account_id': 693,
                             u'name': u'Christopher MacGown'}},
u'Verified': {u'all': [{u'_account_id': 3,
                        u'name': u'Jenkins',
                        u'value': 1}]},
u'recommended': {u'_account_id': 3,
                  u'avatars': [],
                  u'name': u'Jenkins',
                  u'username': u'jenkins'},
                  u'value': 1},
u'Workflow': {u'all': [{u'_account_id': 209,
                        u'name': u'Josh Kearney',
                        u'value': 1}]},
                  u'approved': {u'_account_id': 209,
                                 u'name': u'Josh Kearney'}}},
u'g:messages': [{u'_revision_number': 1,
                  u'author': {u'_account_id': 2166,
                              u'avatars': []},

```

```

        u'name': u'SmokeStack'},
    u'date': u'2012-03-29 15:44:09.000000000',
    u'id': u'AAAFH//9x0=',
    u'message': u'Patch Set 1:\n\nSmokeStack Results (patch set 1):\n\tUnit Success: h
ttp://smokestack.openstack.org/?go=/jobs/16799\n\tLibvirt (Fedora 16) Success: http://smokestack.ope
nstack.org/?go=/jobs/16797\n\tXenServer (Ubuntu 11.10) Success: http://smokestack.openstack.org/?go=
/jobs/16798'},
    {u'_revision_number': 1,
      u'author': {u'_account_id': 693,
        u'avatars': [],
        u'email': u'chris@pistoncloud.com',
        u'name': u'Christopher MacGown',
        u'username': u'0x44'},
      u'date': u'2012-03-29 15:58:44.000000000',
      u'id': u'AAAFH//9xo=',
      u'message': u'Patch Set 1: Looks good to me (core reviewer)\n\n'},
    {u'_revision_number': 1,
      u'author': {u'_account_id': 642,
        u'avatars': [],
        u'email': u'matt.dietz@rackspace.com',
        u'name': u'Matt Dietz',
        u'username': u'cerberus'},
      u'date': u'2012-03-29 16:28:53.000000000',
      u'id': u'AAAFH//9wE=',
      u'message': u'Patch Set 1: Looks good to me (core reviewer)\n\nActual patch looks
good, but discussion on the bug makes me think I need to hold off on pushing this through?'},
    {u'_revision_number': 1,
      u'author': {u'_account_id': 209,
        u'avatars': [],
        u'email': u'josh@jk0.org',
        u'name': u'Josh Kearney'},
      u'date': u'2012-03-29 16:44:14.000000000',
      u'id': u'AAAFH//9vg=',
      u'message': u'Patch Set 1: Looks good to me (core reviewer); Approved\n\n'},
    {u'_revision_number': 1,
      u'author': {u'_account_id': 3,
        u'avatars': [],

```

```

        u'name': u'Jenkins',
        u'username': u'jenkins'},
    u'date': u'2012-03-29 17:18:07.000000000',
    u'id': u'AAAAFH//9uE=',
    u'message': u'Patch Set 1: Verified\n\nBuild Successful \n \nhttps://jenkins.openstack.org/job/gate-integration-tests-devstack-vm/3092/ : SUCCESS \nhttps://jenkins.openstack.org/job/gate-nova-merge/2055/ : SUCCESS \nhttps://jenkins.openstack.org/job/gate-nova-unittests/1711/ : SUCCESS \nhttps://jenkins.openstack.org/job/gate-nova-pep8/3444/ : SUCCESS'},
    {u'author': {u'_account_id': 3,
                  u'avatars': [],
                  u'name': u'Jenkins',
                  u'username': u'jenkins'},
      u'date': u'2012-03-29 17:18:10.000000000',
      u'id': u'AAAAFH//9uA=',
      u'message': u'Change has been successfully merged into the git repository.'}},
    u'g:owner': {u'_account_id': 360, u'name': u'Dan Prince'},
    u'g:project': u'openstack/nova',
    u'g:status': u'MERGED',
    u'g:subject': u'Adds middleware to limit request body sizes.',
    u'g:topic': u'bug/962515',
    u'g:updated': u'2012-03-29 17:18:10.000000000',
    u'lp:activity': [],
    u'lp:attachments': [u'https://api.launchpad.net/1.0/bugs/962515/+attachment/2924517',
                        u'https://api.launchpad.net/1.0/bugs/962515/+attachment/2924518'],
    u'lp:change_id': u'Ifalca56a33ca517679fa80bfd10b962eebe3a76',
    u'lp:cid': u'1ebec5726c7a9db0a6f29fad0ef747b0c087f702',
    u'lp:cves': [],
    u'lp:date_created': u'2012-03-22 20:46:45.304409+00:00',
    u'lp:description': u'Using the following Ruby sample script I can increase the size of the Nova API log file 128M per POST:\n\nrequire \'rubygems\'\nrequire \'openstack/compute\'\n\nUSERNAME=ENV[\'NOVA_USERNAME\']\nAPI_KEY=ENV[\'NOVA_API_KEY\']\nAPI_URL=ENV[\'NOVA_URL\']\n\nbigboy = "0" * 22222222\n\nconn=OpenStack::Compute::Connection.new(:username => USERNAME, :api_key => API_KEY, :auth_url => API_URL, :service_type => \'compute\', :retry_auth => false)\nconn.create_server(:name => bigboy, :imageRef => "8da06b6a-1ddd-4d4f-aa54-579e95b5e8b1", :flavorRef => 1)\n\n-----\n\nSimilarly I can do the same thing with a PUT (server name update):\n\nconn=OpenStack::Compute::Connection.new(:username => USERNAME, :api_key => API_KEY, :auth_url => API_URL, :service_type => \'compute\', :retry_auth => false)\nserver=conn.server("695e8b03-aed1-40ab-81bc-8e7456c36215")\nserver.update(:name => bigboy)

```

\n\n--\n\nEach of these requests will increase the Nova API log file size by 128M per request:\n\n[root@nova1 ~]# du -hs /var/log/nova/api.log\n128M\t/var/log/nova/api.log\n\n--\n\nThe root cause of the issue is that we rely on the Nova database column size to limit the size of the instance name. We should put in an API check on the instance name size/length before sending it off to the database.\n\nAdditionally, as part of this fix I would also like to incorporate a simple request size limiting middleware into our API pipeline so that really large requests are invalid to begin with.\n\n--\n\nThis exploit could allow an authenticated user to run the Nova API server out of disk space.\n\nNova's rate limiting middleware will help guard the number of POST and PUT requests a given user can make. The default POST limit to /servers is 50 per day. The PUT limit is however much higher at 10 per minute. Either of these could provide opportunities to run API servers out of disk space.',

```
u'lp:duplicate_of': u'None',
u'lp:heat': u'264',
u'lp:id': u'962515',
u'lp:information_type': u'Public Security',
u'lp:message_count': u'37',
u'lp:messages': [{u'bug_attachments': [],
u'content': u'Using the following Ruby sample script I can increase the size of the Nova API log file 128M per POST:\n\nrequire \'rubygems\'\nrequire \'openstack/compute\'\n\nUSERNAME=ENV[\'NOVA_USERNAME\']\nAPI_KEY=ENV[\'NOVA_API_KEY\']\nAPI_URL=ENV[\'NOVA_URL\']\n\nbigboy = "0"
* 22222222\n\nconn=OpenStack::Compute::Connection.new(:username => USERNAME, :api_key => API_KEY, :auth_url => API_URL, :service_type => \'compute\', :retry_auth => false)\nconn.create_server(:name => bigboy, :imageRef => "8da06b6a-1ddd-4d4f-aa54-579e95b5e8b1", :flavorRef => 1)\n\n-----\n\nSimilarly I can do the same thing with a PUT (server name update):\n\nconn=OpenStack::Compute::Connection.new(:username => USERNAME, :api_key => API_KEY, :auth_url => API_URL, :service_type => \'compute\', :retry_auth => false)\nserver=conn.server("695e8b03-aed1-40ab-81bc-8e7456c36215")\nserver.update(:name => bigboy)\n\n--\n\nEach of these requests will increase the Nova API log file size by 128M per request:\n\n[root@nova1 ~]# du -hs /var/log/nova/api.log\n128M\t/var/log/nova/api.log\n\n--\n\nThe root cause of the issue is that we rely on the Nova database column size to limit the size of the instance name. We should put in an API check on the instance name size/length before sending it off to the database.\n\nAdditionally, as part of this fix I would also like to incorporate a simple request size limiting middleware into our API pipeline so that really large requests are invalid to begin with.\n\n--\n\nThis exploit could allow an authenticated user to run the Nova API server out of disk space.\n\nNova's rate limiting middleware will help guard the number of POST and PUT requests a given user can make. The default POST limit to /servers is 50 per day. The PUT limit is however much higher at 10 per minute. Either of these could provide opportunities to run API servers out of disk space.',
```

```
u'date_created': u'2012-03-22 20:46:45.304409+00:00',
```



```

    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/0',
    u'subject': u"PUT/POST of large server name's can increase nova API log file size
massively"},
    {u'bug_attachments': [],
     u'content': u'even if the log is not at DEBUG level ?',
     u'date_created': u'2012-03-22 21:06:46.033944+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/1',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
     u'content': u'Adding PTL',
     u'date_created': u'2012-03-22 21:07:46.273112+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/2',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
     u'content': u"tttx: This happens when running without DEBUG (I'm not using the --v
erbose flag etc.).",
     u'date_created': u'2012-03-22 22:17:19.145064+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/3',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
     u'content': u'nasty. Is it also stored in the db that large?',
     u'date_created': u'2012-03-23 01:12:18.462166+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/4',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
     u'content': u"vish: it doesn't get store in the DB that large. Just the nova api.
log file.",
     u'date_created': u'2012-03-23 14:01:55.225738+00:00',

```

```

    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/5',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u'Using MySQL...',
      u'date_created': u'2012-03-23 14:02:57.613291+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/6',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u"I suspect this affects Diablo as well ?\nLike for the Keystone issu
e, we don't really want the coordinated disclosure to interfere with our RC publication or final rel
ease...\n\nIf the request size limiting middleware works around the issue completely, would merging
it innocently in Nova RC2 and working on disclosure asynchronously be an option ?",
      u'date_created': u'2012-03-23 15:01:01.505346+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/7',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u'So...\n\nThe fix for the instance name issue is pretty simple:\n\n+
++ b/nova/api/openstack/compute/servers.py\n@@ -505,6 +505,10 @@ class Controller(wsgi.Controller):\n
n      msg = _("Server name is an empty string")\n      raise exc.HTTPBadRequest(expla
nation=msg)\n\n+      if not len(value) < 256:\n+          msg = _("Server name must be less th
an 255 characters.")\n+          raise exc.HTTPBadRequest(explanation=msg)\n\n----\n\nI am however
concerned that this same problem could occur in other places as well. We don't appear to validate
the size of image snapshot names either for example. I haven't tested to see what Glance actually d
oes with a 128M image name but it could cause problems...\n\n---\n\nPutting in middleware to limit t
he request size is fairly straightforward and shouldn't affect anything that I know of in the OSAPI
(it would potentially affect the EC2 API but we can ignore it for this anyway). And it would global
ly guard against really large requests attacks.\n\nI'm not sure it really belongs in nova's code b
ase long term however. Both Apache and Nginx are quite capable of doing this already and most produc
tion deployments are probably using something similar for SSL or load balancing already.\n\nIn summa
ry maybe it is a good short term solution... and long term we can move towards better validation in
the OSAPI. We recently added some Validator middleware to EC2's API. Perhaps something similar woul

```

```

d help with the OSAPI requests as well.\n',
    u'date_created': u'2012-03-23 15:50:40.052342+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/8',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u'We could push a public fix for an innocent-looking "long server nam
es trigger 500 error instead of 400" bug that would also silently fix this.\n\nThe issue being, as y
ou mentioned, that it doesn\'t really address potential other affected areas.',
      u'date_created': u'2012-03-23 16:29:13.749353+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/9',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u'(fixing it silently in the RC allows us to embargo/coordinate with
downstream on the diablo fix)',
      u'date_created': u'2012-03-23 16:30:03.508407+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/10',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u'Are we really in that much of a time crunch to get this into an RC?
Pushing out security patches disguised as something else is a bad habit for us to get into ...',
      u'date_created': u'2012-03-23 17:03:58.865907+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/11',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u'ttx: So... what I told you on IRC about 257 char server names was i
ncorrect. It appears that a request to set the server name to a value larger than 255 will silently
truncate it... that is until it gets really large (Megs) and that is when SQL exceptions start getti
ng thrown thus causing this issue.\n\nBottom line is to expose this the requests need to be really l
arge.'},

```

```

    u'date_created': u'2012-03-23 17:19:11.209652+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/12',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
     u'content': u"@Russell: I don't like it either. The trick is that RC2 for Nova is
likely to be ready on Thursday next week... and I would hate to delay it or re-roll it. I just don'
t see how we can make the coordinated release date align on such short notice. Or can we ?",
     u'date_created': u'2012-03-23 17:44:14.431248+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/13',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
     u'content': u"If the notification went out today, Thursday seems ok. Otherwise,
I think that's cutting it too close.",
     u'date_created': u'2012-03-23 17:56:52.064477+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/14',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
     u'content': u"... or maybe Monday. I guess that's still a few days notice.",
     u'date_created': u'2012-03-23 18:04:57.106692+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/15',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [u'https://api.launchpad.net/1.0/nova/+bug/962515/+attachment
/2924517'],
     u'content': u'Patch that adds a validation for OSAPI server name length.',
     u'date_created': u'2012-03-23 20:01:19.010636+00:00',
     u'resource_type': u'#message',
     u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/16',
     u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},

```

```

{u'bug_attachments': [],
 u'content': u'Request body size limiting middleware patch.',
 u'date_created': u'2012-03-23 20:02:18.149098+00:00',
 u'resource_type': u'#message',
 u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/17',
 u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
{u'bug_attachments': [],
 u'content': u"Attached are patches that should fix this issue. I'll wait to hear
back from Thierry about how to proceed.",
 u'date_created': u'2012-03-23 20:04:24.244637+00:00',
 u'resource_type': u'#message',
 u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/18',
 u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
{u'bug_attachments': [],
 u'content': u'We have three options:\n1- File a public "server names over size sh
ould return 400" bug and fix it openly in RC2, then calmly do a CRD for stable/diablo\n2- Propose th
e OSAPI server name length patch for coordinated disclosure and complete it just in time for RC2\n3-
Propose the server name patch AND the sizelimit middleware for coordinated disclosure and complete
them just in time for RC2\n\nMy issue with solution (3) is that we would push all requests through t
he new sizelimiting middleware as the very last thing before RC2 publication... which sounds extreme
ly risky to me, that late in the cycle, for the potential benefit of catching unknown other ways of
abusing large requests. I'll let Vish decide, but with my release manager hat on, I'd -1 this.\n\n
If we don't push the sizelimit middleware in Essex anyway, then I'd rather do option (1) than opti
on (2).\nOther opinions ?',
 u'date_created': u'2012-03-26 10:00:01.344721+00:00',
 u'resource_type': u'#message',
 u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/19',
 u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
{u'bug_attachments': [],
 u'content': u'Proposed impact description:\n\nTitle: Long server names grow nova-
api log files significantly\nImpact: High\nReporter: Dan Prince\nProducts: Nova\nAffects: All versio
ns\n\nDescription:\nDan Prince reported a vulnerability in OpenStack Compute (Nova) API servers. By
PUTing or POSTing extremely long server names to the OpenStack API, any authenticated user may grow
nova-api log files significantly, potentially resulting in disk space exhaustion and denial of servi

```

```

ce to the affected nova-api nodes. Only setups running the OpenStack API are affected.',
    u'date_created': u'2012-03-26 10:09:41.385455+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/20',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u"I vote for option 2, just because I think this practice of openly f
ixing security issues as something else less serious is poor practice. I'll defer to ttx for the fi
nal decision, though.",
      u'date_created': u'2012-03-26 13:54:34.260250+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/21',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u"I'm OK with option 2, if the stakeholders announcement is sent toda
y for a CRD at the end of the week (Thursday or Friday). Would like to have Vish's blessing of the p
atch and the strategy first.",
      u'date_created': u'2012-03-26 14:58:05.539436+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/22',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u"2 is ok. I still don't mind putting in the size middleware as an o
ption so that deployers can turn it on if needed. But I'm fine doing it with a backport to stable/e
sssex later.",
      u'date_created': u'2012-03-26 17:26:56.808024+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/23',
      u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u"I'm about to send out the advance notification ...\\n\\nI noticed one
typo in the patch. The log message says the name must be less than 255 characters, but the code en
forces less than 256. I'm just changing the message to say 256 instead of 255.",

```

```

    u'date_created': u'2012-03-26 19:30:33.637901+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/24',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
    u'content': u'CRD is Thursday, March 29th, 1500 UTC',
    u'date_created': u'2012-03-26 19:54:50.446270+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/25',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
    u'content': u'Added markmc and Daviey for stable/diablo',
    u'date_created': u'2012-03-29 14:19:39.647006+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/26',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
    u'content': u'CRD reached and patch pushed to gerrit, opening the issue',
    u'date_created': u'2012-03-29 15:09:05.976067+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/27',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
    u'content': u'Added blamar for pre-review',
    u'date_created': u'2012-03-29 15:10:09.611595+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/28',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
    u'content': u'Reviewed: https://review.openstack.org/5955\nCommitted: http://git
hub.com/openstack/nova/commit/c7f526fae6062e9ab51f65474af71d496aa66554\nSubmitter: Jenkins\nBranch:
master\n\ncommit c7f526fae6062e9ab51f65474af71d496aa66554\nAuthor: Dan Prince <email address hidd

```

```

en>\nDate:   Fri Mar 23 15:03:19 2012 -0400\n\n    Add validation for OSAPI server name length.\n\n    Fixes LP Bug #962515.\n\n    Change-Id: Iee895604f8e9101a341a5909fc5ba2dd8e708b4b\n',
    u'date_created': u'2012-03-29 15:32:43.912876+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/29',
    u'subject': u'Fix merged to nova (master)'}},
    {u'bug_attachments': [],
    u'content': u'Reviewed: https://review.openstack.org/5957\nCommitted: http://git
hub.com/openstack/nova/commit/0fa7d12dbfb7ae016657dd91034b4c0781ea43de\nSubmitter: Jenkins\nBranch:
master\n\ncommit 0fa7d12dbfb7ae016657dd91034b4c0781ea43de\nAuthor: Dan Prince <email address hidd
en>\nDate:   Fri Mar 23 15:40:57 2012 -0400\n\n    Adds middleware to limit request body sizes.\n\n    Fixes LP Bug #962515.\n\n    Change-Id: Iclbe1459515654d45febd89da58b19e0840aaf9d\n',
    u'date_created': u'2012-03-29 17:18:17.601917+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/30',
    u'subject': u'Fix merged to nova (master)'}},
    {u'bug_attachments': [],
    u'content': u'Why is the limit 256 characters?',
    u'date_created': u'2012-03-29 19:04:49.091296+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/31',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
    u'content': u'That is the limit of the field in the instances table:\n\n| server_
name | varchar(255) | YES | | NULL | | \n',
    u'date_created': u'2012-03-29 19:47:49.544222+00:00',
    u'resource_type': u'#message',
    u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/32',
    u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
    u'content': u'How many other places in nova/elsewhere have this issue :-(\nSeems
like any buffer being read from input needs to be size limited and/or type checked before it hits th
e DB (which will also either chop off part of the string or fail at insertion time).',
    u'date_created': u'2012-03-29 20:48:51.822796+00:00',
    u'resource_type': u'#message',

```



```

        u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/33',
        u'subject': u"Re: PUT/POST of large server name's can increase nova API log file
size massively"},
    {u'bug_attachments': [],
      u'content': u'Fix proposed to branch: milestone-proposed\nReview: https://review.
openstack.org/6018',
      u'date_created': u'2012-03-30 23:43:03.292125+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/34',
      u'subject': u'Fix proposed to nova (milestone-proposed)'},
    {u'bug_attachments': [],
      u'content': u'Reviewed: https://review.openstack.org/6018\nCommitted: http://git
hub.com/openstack/nova/commit/c869a41951b77c6930bf4fb4734f05cd3d6ac4b1\nSubmitter: Jenkins\nBranch:
milestone-proposed\n\ncommit c869a41951b77c6930bf4fb4734f05cd3d6ac4b1\nAuthor: Dan Prince <email
address hidden>\nDate: Fri Mar 23 15:03:19 2012 -0400\n\n Add validation for OSAPI server name
length.\n\n Fixes LP Bug #962515.\n\n Change-Id: Iee895604f8e9101a341a5909fc5ba2dd8e70
8b4b\n',
      u'date_created': u'2012-03-31 23:42:00.937281+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/35',
      u'subject': u'Fix merged to nova (milestone-proposed)'},
    {u'bug_attachments': [],
      u'content': u'Reviewed: https://review.openstack.org/5956\nCommitted: http://git
hub.com/openstack/nova/commit/1ebec5726c7a9db0a6f29fad0ef747b0c087f702\nSubmitter: Jenkins\nBranch:
stable/diablo\n\ncommit 1ebec5726c7a9db0a6f29fad0ef747b0c087f702\nAuthor: Dan Prince <email addre
ss hidden>\nDate: Thu Mar 29 10:46:59 2012 -0400\n\n Add validation for OSAPI server name lengt
h.\n\n Diablo Fix for LP Bug #962515.\n\n Change-Id: Ifalca56a33ca517679fa80bfd10b962e
eebe3a76\n',
      u'date_created': u'2012-04-11 17:06:19.158719+00:00',
      u'resource_type': u'#message',
      u'self_link': u'https://api.launchpad.net/1.0/nova/+bug/962515/comments/36',
      u'subject': u'Fix merged to nova (stable/diablo)'},
    u'lp:name': u'None',
    u'lp:resource_type': u'#bug',
    u'lp:review': u'5956',
    u'lp:security_related': u'True',
    u'lp:self_link': u'https://api.launchpad.net/1.0/bugs/962515',

```

```
u'lp:tags': [u'in-stable-diablo'],
u'lp:title': u"PUT/POST of large server name's can increase nova API log file size massively",
u'msg': u'Merge "Adds middleware to limit request body sizes."',
u'parents': [u'484f0a42be3c36aaf508264e1109e0c9b1abc07e',
             u'0fa7d12dbfb7ae016657dd91034b4c0781ea43de' ]}]}
```

Sample entry from all_blame.jsonz

```
In [7]: all_blame = jload('nova_all_blame.jsonz')
```

```
gzipped
returning list
```

```
In [13]: pp.pprint(all_blame[1])
```

```
{u'blame': {u'nova/db/sqlalchemy/session.py': [{u'commit': u'7756a1d269946f72e76bae7a8015c3d72063b2c
6',
                                                u'lineno': 15,
                                                u'proximity': 3},
{u'commit': u'7756a1d269946f72e76bae7a8015c3d72063b2c
6',
                                                u'lineno': 16,
                                                u'proximity': 2},
{u'commit': u'7756a1d269946f72e76bae7a8015c3d72063b2c
6',
                                                u'lineno': 17,
                                                u'proximity': 1},
{u'commit': u'16f4faf4039ecab8119a31d77eb197a1928639e
c',
                                                u'lineno': 18,
                                                u'proximity': 1},
{u'commit': u'16f4faf4039ecab8119a31d77eb197a1928639e
c',
                                                u'lineno': 19,
```

```
u'proximity': 1},
{u'commit': u'16f4faf4039ecab8119a31d77eb197a1928639e
```

```
u'lineno': 20,  
u'proximity': 1},  
{u'commit': u'7756a1d269946f72e76bae7a8015c3d72063b2c
```

```
u'lineno': 21,  
u'proximity': 1},  
{u'commit': u'a46964ad11a85effa833dec81384b478a0cf75
```

```
u'lineno': 22,  
u'proximity': 1},  
{u'commit': u'a46964ad11a85effa833decb81384b478a0cf75
```

```
u'lineno': 23,  
u'proximity': 2},  
{u'commit': u'7756a1d269946f72e76bae7a8015c3d72063b2c
```

```
u'lineno': 24,  
u'proximity': 1},  
{u'commit': u'a46964ad11a85effa833dec81384b478a0cf75
```

```
u'lineno': 25,  
u'proximity': 1},  
{u'commit': u'a46964ad11a85effa833dec81384b478a0cf75
```

```
u'lineno': 26,  
u'proximity': 2},  
{u'commit': u'a46964ad11a85effa833dec81384b478a0cf75
```

```
u'lineno': 27,  
u'proximity': 1},  
{u'commit': u'a46964ad11a85effa833dec81384b478a0cf75
```

```
u'lineno': 28,
u'proximity': 1},
```

```

d',
    {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75',
     u'lineno': 29,
     u'proximity': 2},
d',
    {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75',
     u'lineno': 30,
     u'proximity': 3},
d',
    {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75',
     u'lineno': 57,
     u'proximity': 3},
d',
    {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75',
     u'lineno': 58,
     u'proximity': 2},
3',
    {u'commit': u'01a041dd732ae9c56533f6eac25f08c34917d73',
     u'lineno': 59,
     u'proximity': 1},
d',
    {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75',
     u'lineno': 60,
     u'proximity': 1},
d',
    {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75',
     u'lineno': 61,
     u'proximity': 2},
d',
    {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75',
     u'lineno': 62,
     u'proximity': 3}],
u'nova/flags.py': [{u'commit': u'96d0edff2348040362b77491892e525217a17562',
                    u'lineno': 353,
                    u'proximity': 3},
                  {u'commit': u'4a26a6f2b9ac39d560046d0d9cb82f6ce4554e56',
                    u'lineno': 354,

```

```
        u'proximity': 2},
        {u'commit': u'3c09d486f862de7069b848e8124787cfbf4247f8',
         u'lineno': 355,
         u'proximity': 1},
        {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75d',
         u'lineno': 356,
         u'proximity': 1},
        {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75d',
         u'lineno': 357,
         u'proximity': 2},
        {u'commit': u'a46964ad11a85effa833dec81384b478a0cf75d',
         u'lineno': 358,
         u'proximity': 3}]],
    u'cid': u'f3dd56e916232e38e74d9e2f24ce9a738cac63cf'}
```

Normalize Blame Entry

Weight commits based on proximity to change. Normalize to that overall weight is 0. Normalization can either be reciprocal of distance or exponential

```
In [9]: from collections import defaultdict
```

```
In [10]: def normalize_blame(blameset, exp_weighting=True):
        """returns list of commits with weighted blame"""
        result = defaultdict(float)
        total = 0
        for per_file in blameset['blame'].values():
            if per_file:          #validate not null entry
                for per_line in per_file:
                    total += 1
                    #print per_line
                    if exp_weighting:
                        result[per_line['commit']] += 1.0/(2.0**((per_line['proximity']-1))
                    else:
                        result[per_line['commit']] += 1.0/float(per_line['proximity'])

        return dict([[k, v/total] for k, v in result.items()])
```

```
In [14]: normalize_blame(all_blame[1], exp_weighting=True)
```

```
Out[14]: {u'01a041dd732ae9c56533f6eac25f08c34917d733': 0.03571428571428571,
u'16f4faf4039ecab8119a31d77eb197a1928639ec': 0.10714285714285714,
u'3c09d486f862de7069b848e8124787cfbf4247f8': 0.03571428571428571,
u'4a26a6f2b9ac39d560046d0d9cb82f6ce4554e56': 0.017857142857142856,
u'7756a1d269946f72e76bae7a8015c3d72063b2c6': 0.13392857142857142,
u'96d0edff2348040362b77491892e525217a17562': 0.008928571428571428,
u'a46964ad11a85effa833decb81384b478a0cf75d': 0.35714285714285715}
```

```
In [15]: normalize_blame(all_blame[1], exp_weighting=False)
```

```
Out[15]: {u'01a041dd732ae9c56533f6eac25f08c34917d733': 0.03571428571428571,
u'16f4faf4039ecab8119a31d77eb197a1928639ec': 0.10714285714285714,
u'3c09d486f862de7069b848e8124787cfbf4247f8': 0.03571428571428571,
u'4a26a6f2b9ac39d560046d0d9cb82f6ce4554e56': 0.017857142857142856,
u'7756a1d269946f72e76bae7a8015c3d72063b2c6': 0.1369047619047619,
u'96d0edff2348040362b77491892e525217a17562': 0.011904761904761904,
u'a46964ad11a85effa833decb81384b478a0cf75d': 0.36904761904761907}
```

In [12]: