USING CHAOS TO BUILD RESILIENT SYSTEMS

@tammybuatow, Gremlin
What’s the scale of your infra?
How many services do you have running in production?
How many engineers do you have at your company?
A Common Chaos Engineering Journey

@tammybutow
tammybutow

#QCONNYC
TOP 5 MOST POPULAR WAYS TO USE CHAOS ENGINEERING IN 2018

- Kubernetes
- Kafka
- AWS ECS
- Cassandra
- Elasticsearch

@tammybutow

#QCONNYC
ADVANCED USES OF CHAOS ENGINEERING

Source → Build → Test → Production

- Continuous Integration
- Continuous Delivery
- Continuous Deployment
What happened this week: June 2018 Slack Outage

@londons_explore 2 days ago

Google is expert at designing services which you won't notice when there is downtime.

Take Google Search for example. When there is downtime, results might be slightly less accurate, or the parcel tracking box might not appear, or the page won't say the "last visited" time beside search results.

The SRE's are running around fixing whatever subsystem is down or broken, but you the user probably don't notice.

kolton 12:04 PM

And testing those graceful degradations ... is where chaos eng comes in. :)

@tammybutow #QCONNYC
@tammybütow

#QCONNYC
TAMMY BÜTOW
Principal SRE, Gremlin
Causing chaos in prod since 2009.

Previously SRE Manager @ Dropbox leading Databases, Block Storage and Code Workflows for 500 million users and 800 engineers.

@tammybütow
GREMLIN

• We are practitioners of Chaos Engineering

• We build software that helps engineers build resilient systems in a safe, secure and simple way.

• We offer 11 ways to inject chaos for your Chaos Engineering experiments (e.g. host/container packet loss and shutdown)

@tammybütow  #QCONNYC
PART 1: LAYING THE FOUNDATION
Let’s Define A Resilient System:

- A resilient system is a highly available and durable system.

- A resilient system can maintain an acceptable level of service in the face of failure.

- A resilient system can weather the storm (a misconfiguration, a large scale natural disaster or controlled chaos engineering).
It would be silly to give an Olympic pole-vaulter a broom and ban them from practicing!
“Thoughtful planned experiments designed to reveal the weaknesses in our systems”

- Kolton Andrus, Gremlin CEO
Think of it like a vaccination:
Inject something harmful in order to build an immunity.
Eventually systems will break in many undesired ways.

Break them first on purpose with controlled chaos! 💥
DOGFOODING

• Using your own product.

• For us that means using Gremlin for our Chaos Engineering experiments.

• Failure Fridays
Failure Fridays are dedicated time for teams to collaboratively focus on using Chaos Engineering practices to reveal weaknesses in your services.
WHY DO DISTRIBUTED SYSTEMS NEED CHAOS?

• Unusual hard to debug failures are common

• Systems & companies scale rapidly and Chaos Engineering helps you learn along the way

@tammybütow

#QCONNYC
FULL-STACK CHAOS ENGINEERING

• You can inject chaos at any layer.

• API, App, Cache, Database, OS, Host, Network, Power & more.
WHY RUN CHAOS ENGINEERING EXPERIMENTS?
Are you confident that your metrics and alerting are as good as they should be?

#pagerpain 📢

@tammybütow #QCONNYC
Are you confident your customers are getting as good an experience as they should be?

#customerpain 😞
Are you losing money due to downtime and broken features?

#businesspain 💸

@tammybutow

#QCONNYC
HOW DO YOU RUN CHAOS ENGINEERING EXPERIMENTS?

@tammybutow  #QCONNYC
HOW TO RUN A CHAOS ENGINEERING EXPERIMENT

• Form a hypothesis
• Consider blast radius
• Run experiment
• Measure results
• Find & fix issues or scale

@tammybülow  #QCONNYC
Don’t run before you can walk
The 3 Prerequisites for Chaos Engineering

1. Monitoring & Observability
2. On-Call & Incident Management
3. Know Your Cost of Downtime Per Hour
What Do I Use For Monitoring & Observability?

DATADOG

SENTRY

WIRESHARK

@tammybutow

#QCONNYC
We All Need To Know The Cost Of Downtime

British Airways Owner Says Power Outage Cost 80 Million Pounds

By Maria Tadeo and Christopher Jasper
15 June 2017, 00:44 GMT-4

British Airways owner said a power outage that led to the cancellation of hundreds of flights last month probably cost it about 80 million pounds ($107 million), but said it was not the airline's fault.
We All Need Incident Management
HOW TO CHOOSE A CHAOS EXPERIMENT

- Identify top 5 critical systems
- Choose 1 system
- Whiteboard the system
- Select attack: resource/state/network
- Determine scope
WHAT SHOULD WE MEASURE?

- Availability — 500s
- Service specific KPIs
- System metrics: CPU, IO, Disk
- Customer complaints
# HOW TO RUN YOUR OWN GAME DAY!

## RESOURCE DOWNLOADS

<table>
<thead>
<tr>
<th>Resource</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>GameDay Agenda</td>
<td><a href="#">Download</a></td>
</tr>
<tr>
<td>GameDay Execution Template</td>
<td><a href="#">Download</a></td>
</tr>
<tr>
<td>GameDay Recording Template</td>
<td><a href="#">Download</a></td>
</tr>
<tr>
<td>Objective Communication Template</td>
<td><a href="#">Download</a></td>
</tr>
<tr>
<td>Placeholder Invite for Participants</td>
<td><a href="#">Download</a></td>
</tr>
<tr>
<td>Pre GameDay Checklist</td>
<td><a href="#">Download</a></td>
</tr>
</tbody>
</table>

[gremlin.com/gameday](#)
HOW TO RUN YOUR OWN GAMEDAY!

gremlin.com/gameday

@tammybutow #QCONNYC
EXAMPLE SYSTEM: KUBERNETES RETAIL STORE

User

Primary: kube-01

Node: kube-02
Node: kube-03
Node: kube-04

@tammybutow #QCONNYC
PART 2: RESOURCE CHAOS ENGINEERING
We can increase CPU, Disk, IO & Memory consumption to ensure monitoring is setup to catch problems.

Important to catch issues before they turn into high severity incidents (unable to purchase new product!) and downtime for customers.
CPU CHAOS
LET’S CREATE A “KNOWN-KNOWN” EXPERIMENT

su - chaos
cd scripts
ls
cat burncpu.sh
./burncpu.sh

https://github.com/tammybutow/chaosengineeringbootcamp

@tammybutow  #QCONNYC
CHAOS IN TOP

@tammybutow  #QCONNYC
LET’S KILL THE CHAOS NOW

`pkll -u chaos`

@tammybutow #QCONNYC
NO MORE CHAOS IN TOP

@tammybuatow  #QCONNYYC
DISK CHAOS
**DISK CHAOS**

```bash
root@ubuntu-s-1vcpu-1gb-sfo2-01:~# df -h
Filesystem  Size  Used  Avail  Use%  Mounted on
udev       490M   0   490M    0%  /dev
tmpfs      100M   3.1M   97M    4%  /run
/dev/vda1  25G   1.1G   23G    5%  /
  tmpfs    497M   0   497M    0%  /dev/shm
tmpfs      5.0M   0   5.0M    0%  /run/lock
tmpfs     497M   0   497M    0%  /sys/fs/cgroup
/dev/vda15 105M   3.4M  101M    4%  /boot/efi
tmpfs     100M   0   100M    0%  /run/user/0
root@ubuntu-s-1vcpu-1gb-sfo2-01:~#  
```

@tammybütow  #QCONNYC
MEMORY CHAOS

free -m

NAME
free - Display amount of free and used memory in the system

SYNOPSIS
free [options]

description
free displays the total amount of free and used physical and swap memory in the system, as well as the buffers and caches used by the kernel. The information is gathered by parsing /proc/meminfo. The displayed columns are:

- total: Total installed memory (MemTotal and SwapTotal in /proc/meminfo)
- used: Used memory (calculated as total - free - buffers - cache)
- free: Unused memory (MemFree and SwapFree in /proc/meminfo)
- shared: Memory used (mostly) by tmpfs (Shmem in /proc/meminfo, available on kernels 2.6.32, displayed as zero if not available)

buffers
PART 3: STATE
CHAOS ENGINEERING
PROCESS CHAOS
PROCESS CHAOS

Ways to create process chaos on purpose:

- Kill one process
- Loop kill a process
- Spawn new processes
- Fork bomb
PROCESS CHAOS

PGREP(1) User Commands PGREP(1)

NAME
pgrep, pkill - look up or signal processes based on name and other attributes

SYNOPSIS
pgrep [options] pattern
pkill [options] pattern

DESCRIPTION
pgrep looks through the currently running processes and lists the process IDs which match the selection criteria to stdout. All the criteria have to match. For example,

$ pgrep -u root sshd

will only list the processes called sshd AND owned by root. On the other hand,

$ pgrep -u root,daemon

will list the processes owned by root OR daemon.

@tammybutow

#QCONNYC
SHUTDOWN CHAOS
SHUTDOWN CHAOS

NAME
shutdown - Halt, power-off or reboot the machine

SYNOPSIS
shutdown [OPTIONS...] [TIME] [WALL...]

DESCRIPTION
shutdown may be used to halt, power-off or reboot the machine.

The first argument may be a time string (which is usually "now"). Optionally, this may be followed by a wall message to be sent to all logged-in users before going down.

The time string may either be in the format "hh:mm" for hour/minutes specifying the time to execute the shutdown at, specified in 24h clock format. Alternatively it may be in the syntax "+m" referring to the specified number of minutes m from now. "now" is an alias for "+0", i.e. for triggering an immediate shutdown. If no time argument is specified, "+1" is implied.

Note that to specify a wall message you must specify a time argument, too.

@tammybutow #QCONNYC

shutdown -h
WHAT ARE OTHER WAYS YOU CAN TURN OFF A SERVER?

WHAT IF YOU WANT TO TURN OFF EVERY SERVER WHEN IT’S ONE WEEK OLD?
HALT, REBOOT & POWEROFF CHAOS

HALT(8)       halt     HALT(8)

NAME
halt, poweroff, reboot - Halt, power-off or reboot the machine

SYNOPSIS
halt [OPTIONS...]
poweroff [OPTIONS...]
reboot [OPTIONS...]

DESCRIPTION
halt, poweroff, reboot may be used to halt, power-off or reboot the machine.

OPTIONS
The following options are understood:

--help
   Print a short help text and exit.

--halt
   Halt the machine, regardless of which one of the three commands is
WHAT ABOUT SHUTTING DOWN CONTAINERS AND K8’S PODS?
THE MANY WAYS TO KILL CONTAINERS

- Kill self
- Kill a container from the host
- Use one container to kill another
- Use one container to kills several containers
- Use several containers to kill several

@tammybutow
#QCONNYC
The average lifespan of a container is 2.5 days
And they fail in many unexpected ways.
TIME TRAVEL CHAOS
TIME TRAVEL CHAOS AKA CLOCK SKEW

Clock skew on my VPS server

I am having a problem with the time on my VPS hosted at Rackspace. It's running Red Hat
Enterprise Linux 5.4. Date commands gives a time which is 11 minutes ahead of the actual time.

I am running `ntpd` and my `/etc/ntp.conf` is as per the 2nd option given here

I also modified `/etc/sysconfig/ntpd` and changed `SYNC_WI_CLOCKkeys` and rebooted the server.

None of this has helped so far. I contacted Rackspace and their answer was the time on the main
host is OK.

Output of `ntpq -p` is:

```
root@webserver [etc/sysconfig]# ntpq -p
remote refid st when poll reach delay offset jitter
*181.183.19.174 2 u 7 64 377 20.666 -673871 19.246
mail.your-site.com 72.6.6.6.6 3 u 24 64 377 20.666 -673874 24.355
linode.apax.net 10.0.0.250 3 u 10 64 377 54.088 -673838 23.812
LOCAL(3) . . . .
root@webserver [etc/sysconfig]# ntpq -p
```

I would appreciate any assistance on this.
PART 4: NETWORK CHAOS ENGINEERING
BLACKHOLE CHAOS
BLACKHOLE CHAOS

ROUTE(8) Linux Programmer's Manual ROUTE(8)

NAME
route - show / manipulate the IP routing table

SYNOPSIS
route [-CfVnee]

route [-v] [-A family] add [-net|--host] target [netmask Nm] [gw Gw] [metric N] [mss M] [window W] [irtt I] [reject] [mod] [dyn] [reinstall] [[dev] If]

route [-v] [-A family] del [-net|--host] target [gw Gw] [netmask Nm] [metric N] [[dev] If]

route [-V] [--version] [-h] [--help]

DESCRIPTION
Route manipulates the kernel's IP routing tables. Its primary use is to set up static routes to specific hosts or networks via an interface after it has been configured with the ifconfig(8) program.

When the add or del options are used, route modifies the routing tables. Without these options, route displays the current contents of the routing

ip route show

@tammybutow

#QCONNYC
DNS CHAOS
DNS CHAOS

2016 Dyn cyberattack

From Wikipedia, the free encyclopedia

The 2016 Dyn cyberattack took place on October 21, 2016, and involved multiple distributed denial-of-service attacks (DDoS attacks) targeting systems operated by Domain Name System (DNS) provider Dyn, which caused major Internet platforms and services to be unavailable to large swathes of users in Europe and North America. The groups Anonymous and New World Hackers claimed responsibility for the attack, but scant evidence was provided.

As a DNS provider, Dyn provides to end-users the service of mapping an Internet domain name—when, for instance, entered into a web browser—to its corresponding IP address. The distributed denial-of-service (DDoS) attack was accomplished through a large number of DNS lookup requests from tens of millions of IP addresses. The activities are believed to have been executed through a botnet consisting of a large number of Internet-connected devices—such as printers, IP cameras, residential gateways and baby monitors—that had been infected with the Mirai malware.

Contents

1 Timeline and impact
   1.1 Affected services
2 Investigation
3 Perpetrators
4 See also
5 References

Dyn cyberattack

Map of areas most affected by attack,
16:45 UTC, 21 October 2016.¹

Date  October 21, 2016
Time  12:10 – 14:20 UTC
      16:30 – 18:11 UTC
      21:00 – 23:11 UTC
      [citation needed] [needs update]
Location  Europe and North America,
         especially the Eastern United States
Type  Distributed denial-of-service
Participants  Unknown
DNS CHAOS

@tammybütow

#QCONNYC
LATENCY CHAOS
## LATENCY CHAOS

```bash
# My traceroute [v0.86]

ubuntu-s-1vcpu-1gb-sfo2-01 (0.0.0.0)  Sat Jun  9 06:46:26 2018
resolv: Received error response 2. (server failure)er of fields quit

<table>
<thead>
<tr>
<th>Host</th>
<th>Loss%</th>
<th>Snt</th>
<th>Last</th>
<th>Avg</th>
<th>Best</th>
<th>Wrst</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 206.189.208.253</td>
<td>0.0%</td>
<td>25</td>
<td>0.3</td>
<td>1.7</td>
<td>0.3</td>
<td>18.7</td>
<td>3.7</td>
</tr>
<tr>
<td>2. 138.197.249.152</td>
<td>0.0%</td>
<td>25</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>1.3</td>
<td>0.0</td>
</tr>
<tr>
<td>3. 206.41.106.63</td>
<td>0.0%</td>
<td>25</td>
<td>1.9</td>
<td>1.2</td>
<td>1.0</td>
<td>2.1</td>
<td>0.0</td>
</tr>
<tr>
<td>4. 108.170.242.241</td>
<td>0.0%</td>
<td>24</td>
<td>2.6</td>
<td>2.6</td>
<td>2.4</td>
<td>3.9</td>
<td>0.2</td>
</tr>
<tr>
<td>5. 108.170.237.147</td>
<td>0.0%</td>
<td>24</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>6. nuq04s29-in-f14.1e100.net</td>
<td>0.0%</td>
<td>24</td>
<td>1.6</td>
<td>1.5</td>
<td>1.4</td>
<td>1.7</td>
<td>0.0</td>
</tr>
</tbody>
</table>
```

mtr google.com

@tammybütow  #QCONNYC
PACKET LOSS CHAOS
Packet Loss Chaos

Package: wireshark (2.4.5-1) [universe]

- network traffic analyzer - meta-package

Other Packages Related to wireshark
- depends
- recommends
- suggests
- enhances

- wireshark
  network traffic analyzer
- or
  wireshark-gtk
  network traffic analyzer - GTK+ version

Download wireshark

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Package Size</th>
<th>Installed Size</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>amd64</td>
<td>4.4 kB</td>
<td>54.0 kB</td>
<td>[list of files]</td>
</tr>
<tr>
<td>arm64</td>
<td>4.4 kB</td>
<td>54.0 kB</td>
<td>[list of files]</td>
</tr>
<tr>
<td>armhf</td>
<td>4.4 kB</td>
<td>54.0 kB</td>
<td>[list of files]</td>
</tr>
<tr>
<td>i386</td>
<td>4.4 kB</td>
<td>54.0 kB</td>
<td>[list of files]</td>
</tr>
<tr>
<td>ppc64le</td>
<td>4.4 kB</td>
<td>54.0 kB</td>
<td>[list of files]</td>
</tr>
<tr>
<td>s390x</td>
<td>4.4 kB</td>
<td>54.0 kB</td>
<td>[list of files]</td>
</tr>
</tbody>
</table>

Links for wireshark

- Ubuntu Resources:
  - Bug Reports
  - Ubuntu Changelog
  - Copyright File

Download Source Package wireshark:
- [wireshark_2.4.5-1.deb]
- [wireshark_2.4.5.orig.tar.gz]
- [wireshark_2.4.5-1.debian.tar.xz]

Maintainer:
- Ubuntu MOTU Developers (Mail Archive)
PART 5: COMPLEX OUTAGES
We can combine different types of chaos engineering experiments to reproduce complicated outages.

Reproducing outages gives you confidence you can handle it if/when it happens again.

@tammybutow

#QCONNYC
Let’s go back in time to look at some of the worst outage stories that kicked off the introduction of chaos engineering.
DROPBOX’S WORST OUTAGE EVER

Some master-replica pairs were impacted which resulted in the site going down.


@tammybutow  #QCONNYC
UBER’S DATABASE OUTAGE

1. Master log replication to S3 failed
2. Logs backed up on the primary
3. Alerts fired to engineer but they are ignored
4. Disk fills up on database primary
5. Engineer deletes unarchived WAL files
6. Error in config prevents promotion

— Matt Ranney, Uber, 2015
OUTAGES HAPPEN.
THERE ARE MANY MORE OUTAGES YOU CAN READ ABOUT HERE:
https://github.com/danluuu/post-mortems
HOW CAN YOU CONTINUE YOUR CHAOS ENGINEERING JOURNEY?
JOIN THE CHAOS SLACK

GREMLIN.COM/SLACK

@tammybutow

#QCONNYC
LEARN WITH THE GREMLIN COMMUNITY

Break things together

The Gremlin community offers resources and relationships to help you build, maintain, and scale applications.

Explore the tutorials.

Learn how to build resilient applications through failure-focused engineering.

How to install and use Gremlin on Ubuntu 18.04
By Tommy LaBrie

How to install and use Gremlin on Amazon Web Services.
By Tommy LaBrie

How to install and use Gremlin on Gremlin v7
By Tommy LaBrie

GREMLIN.COM/COMMUNITY

@tammybutow #QCONNYC
THE FIRST CHAOS ENGINEERING CONFERENCE!
THANK YOU
QCON NYC
@tammybütow

#CHAOSENGINEERING