

the *Availability Digest*

www.availabilitydigest.com
[@availabilitydig](https://twitter.com/availabilitydig)

Saving Data Centers by Creating Chaos February 2018

Mission-critical application outages have been plaguing companies since the dawn of the Internet. So much of our economy is driven by online ecommerce that any system outage often exacts a major cost to the company suffering the outage. Even moving critical applications to the cloud does not solve the problem. Clouds fail also.



When an outage occurs, customers can't buy products or services from the affected company. Until recently, the response to a system outage has been reactive. Call in as many people as necessary to find and fix the problem.

Creating Chaos to Minimize Outages

However, large enterprises now have started to break systems on purpose to smoke out weaknesses in their systems. The intent is to determine the impact of a particular outage and how to harden the systems before failures catch them unprepared. This is called 'chaos engineering.'

Netflix and Amazon have been pursuing chaos engineering for almost a decade. Several years ago, Netflix introduced Chaos Monkey to test and enhance the resilience of compute instances.

Chaos Monkey

Chaos Monkey is a software development tool designed to test the resiliency and recoverability of the Amazon Web Services (AWS) used by Netflix. The software simulates failures by shutting down one or more Netflix virtual machines running in AWS. According to its developers, Chaos Monkey was named for the way it wreaks havoc like a wild and armed monkey set loose in a data center.



Chaos Monkey works on the principle that the best way to avoid major failures is to fail constantly. However, unlike unexpected failures, which seem to occur at the worst possible times, Chaos Monkey can be turned off during peak operations. It has a configurable schedule that allows simulated failures to occur at times when they can be closely monitored. In this way, Netflix can address unexpected errors rather than just wait for a catastrophe to happen.

Chaos Monkey is the original member of Netflix's Simian Army, a collection of software tools designed to test the AWS infrastructure. Simian Army, including Chaos Monkey, is open source and is available for other cloud services to use.

Failure Injection Training

Netflix also built a platform called Failure Injection Training (FIT). FIT injects failures into production systems. These are the same kind of failures that are expected to happen during operations. The goal is to understand how best to handle these failures by hardening systems.

Early on, it became obvious that Netflix had to contain the 'blast radius' of FIT to limit the effect on its customers. The initial failure injections affected all of their millions of customers. FIT now has the facilities to allow engineers to be precise on the breadth of a chaos experiment.

Through the use of FIT, Netflix improved the reliability of its sites from three nines to four nines.

Gremlin

In days gone by, software ran in a controlled, bare-metal environment. It had total control of how it ran. Now, software is running in an infrastructure using services that are outside of its control. Therefore, there is a need for enterprises to run their own chaos engineering experiments to determine modes of failure and how to handle them.

To fill this need, Netflix created Gremlin. Gremlin provides cloud providers and enterprises the tools to run their own chaos engineering experiments. Gremlin can run on bare metal, on cloud hosts, or in containers. Gremlin includes multi-factor authentication, secure single sign-on, and role-based control so that it can run experiments in a production environment.

Gremlin includes several capabilities:

- Resource gremlins shows how a service begins to degrade when something goes wrong with the CPU, memory, disk, or I/O.
- Network Gremlins show the impact on the application due to lost or delayed traffic.
- State Gremlins reboot the operating system, change the system time, and attack specific processes.

Summary

Chaos engineering is a technique for injecting failures into a system to determine their impact and to help determine the best way to handle them. It is very useful as a means to harden a system against unexpected failures. Its success is demonstrated by Netflix's ability to pick up an extra nine in the reliability of their sites (i.e., site reliability was improved by a factor of ten).

Netflix has released as open source its own chaos engineering tool, Gremlin. With Gremlin, enterprises can inject their own chaos into their systems to determine failure modes so that the system can be hardened.

Acknowledgements

Information for this article was taken from the following sources:

[Creating Chaos to Save the Datacenter, Next Platform](#); December 12, 2017.
[WhatIs.com](#)