Chaos Engineering

Open Science for Software Engineering

Sylvain Hellegouarch
ChaosIQ CTO

https://chaosiq.io
@lawouach

A talk in three acts

The scene

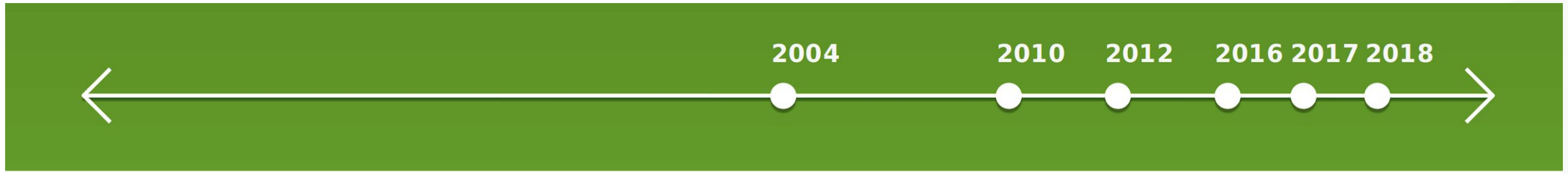
The characters

The plan

Act I -
The one with History

A look at the past?

A worthwhile detour



Chaos engineering

2004 Amazon—Jesse Robbins. Master of disaster

2010 Netflix—Greg Orzell. @chaosimia - First implementation of Chaos Monkey to enforce use of auto-scaled stateless services

2012 NetflixOSS open sources simian army

2016 Gremlin Inc founded

2017 Netflix chaos eng book. Chaos toolkit open source project

2018 Chaos concepts getting adopted widely, and this conference!

Watch Adrian Cockroft's awesome talk at ChaosConf
https://www.youtube.com/watch?v=cefJd2v037U

Let’s illustrate the challenge with a case-study

The Near-Loss and Recovery of America's First Space Station

https://nsc.nasa.gov/resources/case-studies/detail/down-but-not-out

The Near-Loss and Recovery of America's First Space Station

The context

- Skylab first US space station launched in 1973
- Years of design
- Relied on the previous Apollo program

The Near-Loss and Recovery of America's First Space Station

What happened

- Suffered loss of meteoroid shield during launch
- On the face of it, you’d think about the impacts of a meteoroid first, right?
    - Well, the first issue was that temperature went up high in the workshop (up to 200°)
- Engineers worked out ways to reduce the temperature (Recovery first!)
- Next launch was postponed by 10 days

The Near-Loss and Recovery of America's First Space Station

Findings

The overarching ***management system used for Skylab was essentially the same as used for the Apollo program*** *— and was fully operational for Skylab.* ***No inconsistencies or conflicts were found in management records****. What may have affected the oversight of the aerodynamic loads was* ***the view that the shield was a structural component, rather than a complex system*** *involving several distinct technical disciplines.*

In our industry: It worked in the past and it’s a small change. Ring a bell?

Despite six years of design*, review and testing, the project team failed to recognize the shield’s design deficiency because they* ***presumed the shield would be tight to the tank*** *and structurally integrated as set forth in the design criteria.*

Smart and sharp engineers and scientists but previous project may have misled their confidence which wasn’t backed by enough experiments and data.

The Near-Loss and Recovery of America's First Space Station

Findings

Concurrently, the investigation board emphasized that management must always be alert to the potential hazards of its systems and ***take care that an attention to rigor and detail does not inject an undue emphasis on formalism****, documentation and visibility. According to the board,* ***such an emphasis could submerge intuitive thought processes of engineers*** *or analysts.*

Achieving a ***cross-fertilization and engineers’ experience*** *in analysis, design, test or operations* ***will always be important.***

*It's just one of these cases where **Mars is going to give us a new deal**, and we're going to have to **play the cards we get, not the ones we want***

Jim Erickson / Project Manager at Nasa for Mars Rovers missions

Be ready not to be ready

Fast forward to 2018

We learnt, adapted and improved…

Humans are amazing (sometimes)

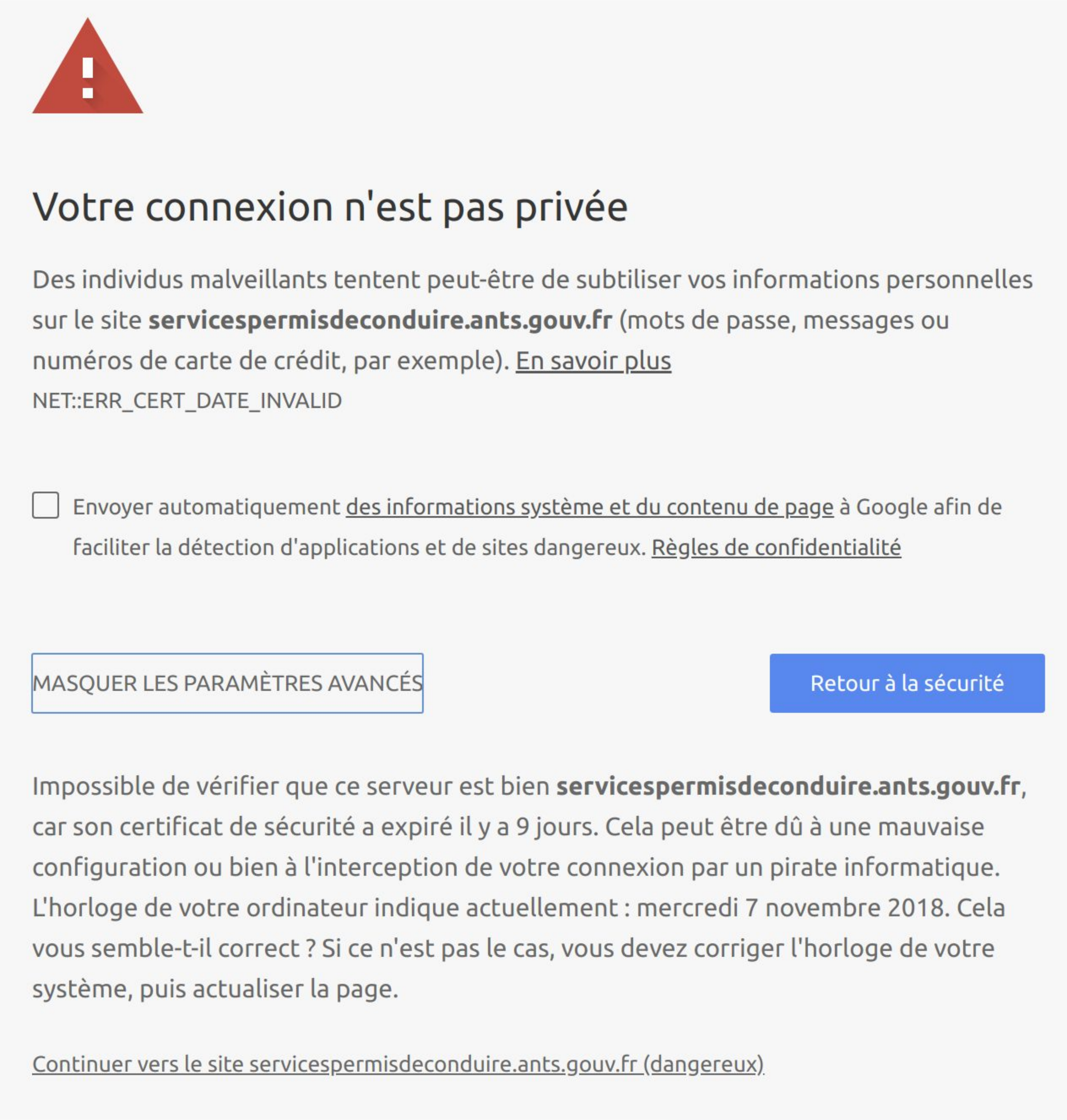
Copyright NASA - Mission InSight

We have learnt indeed. But as systems reliability goes, we could still improve…

A regular certificate warning but in French

French public service for driving license

Certificate had been invalid for about 9 days



Votre connexion n'est pas privée

Des individus malveillants tentent peut-être de subtiliser vos informations personnelles sur le site **servicespermisdeconduire.ants.gouv.fr** (mots de passe, messages ou numéros de carte de crédit, par exemple). En savoir plus

NET::ERR_CERT_DATE_INVALID

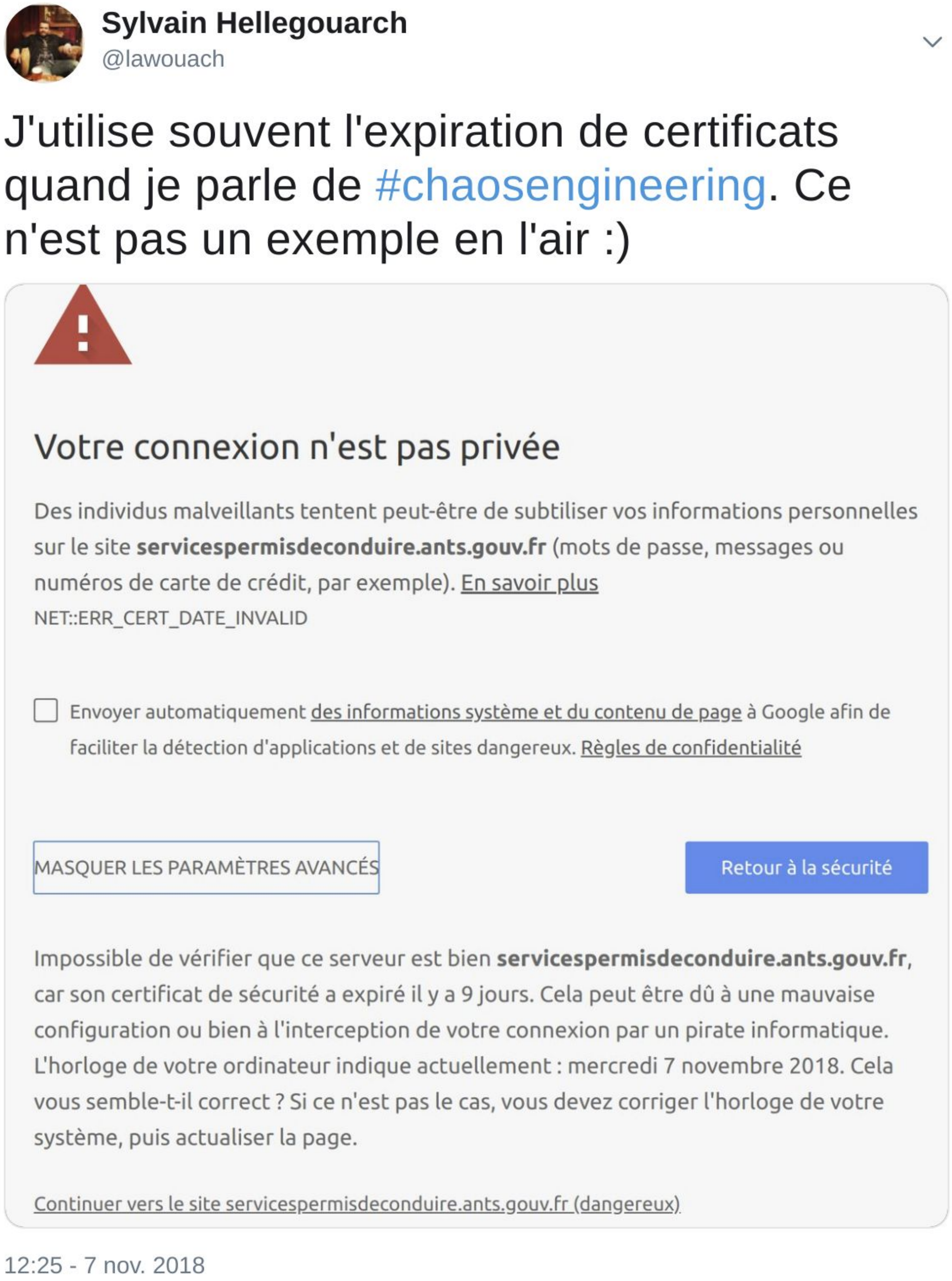
□ Envoyer automatiquement des informations système et du contenu de page à Google afin de faciliter la détection d'applications et de sites dangereux. Règles de confidentialité

Impossible de vérifier que ce serveur est bien **servicespermisdeconduire.ants.gouv.fr**, car son certificat de sécurité a expiré il y a 9 jours. Cela peut être dû à une mauvaise configuration ou bien à l'interception de votre connexion par un pirate informatique. L'horloge de votre ordinateur indique actuellement : mercredi 7 novembre 2018. Cela vous semble-t-il correct ? Si ce n'est pas le cas, vous devez corriger l'horloge de votre système, puis actualiser la page.

Continuer vers le site servicespermisdeconduire.ants.gouv.fr (dangereux)

Twitter seems to be your best alerting platform sometimes

Sent that message at 12:25pm (not just me but a few others too)



Sylvain Hellegouarch
@lawouach

J'utilise souvent l'expiration de certificats quand je parle de #chaosengineering. Ce n'est pas un exemple en l'air :)

Votre connexion n'est pas privée

Des individus malveillants tentent peut-être de subtiliser vos informations personnelles sur le site **servicespermisdeconduire.ants.gouv.fr** (mots de passe, messages ou numéros de carte de crédit, par exemple). En savoir plus

NET::ERR_CERT_DATE_INVALID

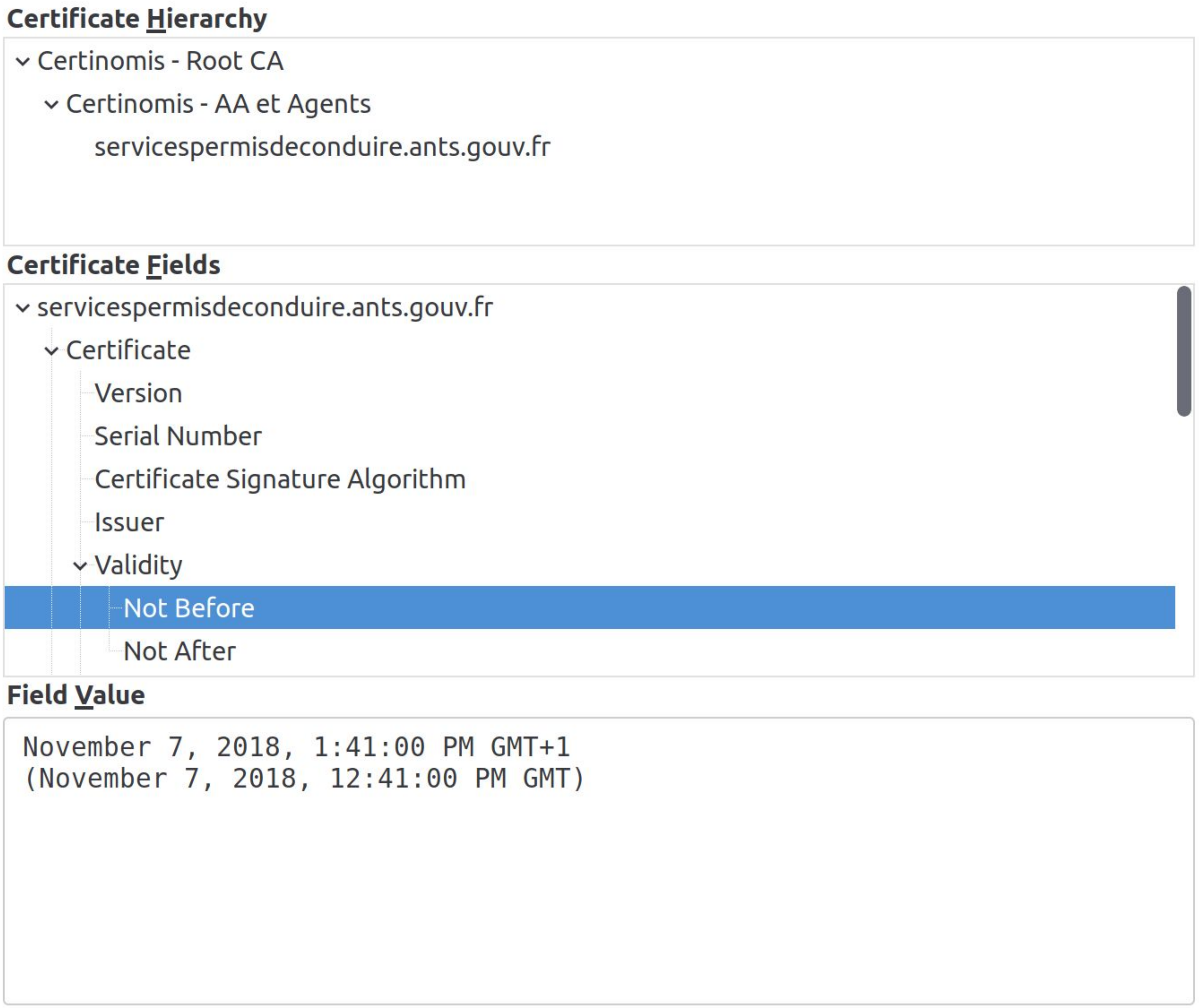
□ Envoyer automatiquement des informations système et du contenu de page à Google afin de faciliter la détection d'applications et de sites dangereux. Règles de confidentialité

Impossible de vérifier que ce serveur est bien **servicespermisdeconduire.ants.gouv.fr**, car son certificat de sécurité a expiré il y a 9 jours. Cela peut être dû à une mauvaise configuration ou bien à l'interception de votre connexion par un pirate informatique. L'horloge de votre ordinateur indique actuellement : mercredi 7 novembre 2018. Cela vous semble-t-il correct ? Si ce n'est pas le cas, vous devez corriger l'horloge de votre système, puis actualiser la page.

Continuer vers le site servicespermisdeconduire.ants.gouv.fr (dangereux)

12:25 - 7 nov. 2018

Updated at 1:41 pm
that same day



Certificate Hierarchy

- Certinomis - Root CA
  - Certinomis - AA et Agents
    - servicespermisdeconduire.ants.gouv.fr

Certificate Fields

- servicespermisdeconduire.ants.gouv.fr
  - Certificate
    - Version
    - Serial Number
    - Certificate Signature Algorithm
    - Issuer
    - Validity
      - Not Before
      - Not After

Field Value

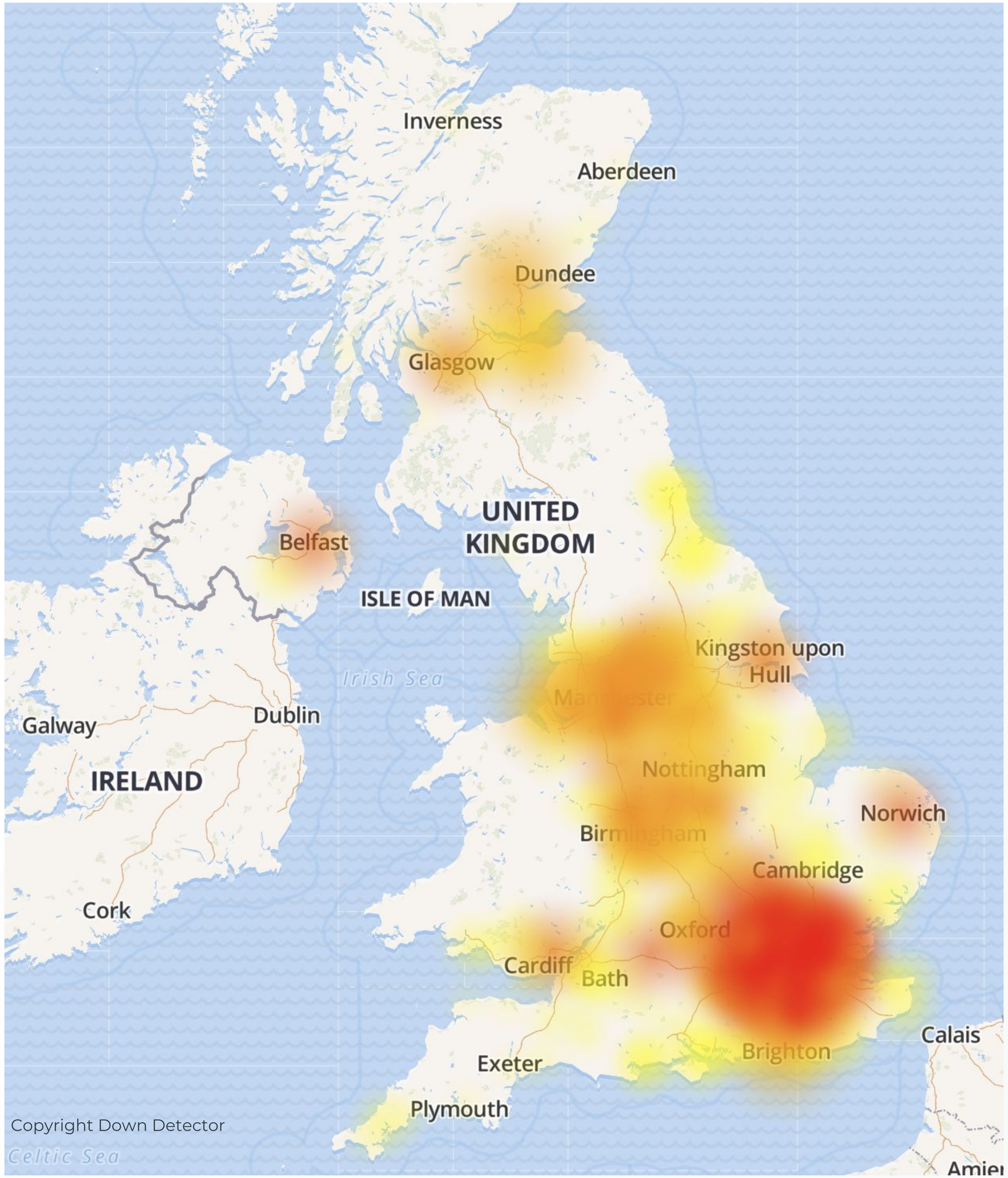
```
November 7, 2018, 1:41:00 PM GMT+1
(November 7, 2018, 12:41:00 PM GMT)
```

Mild impacts but sometimes…

Certificate expiring can cause bigger troubles

02 mobile network outage on December 6th 2018

Earlier Ericsson president Börje Ekholm said "an initial root cause analysis" had indicated that the "main issue was an expired certificate in the software versions installed with these customers".



Copyright Down Detector

Everyone needs more reliable systems and ways to get there!

End of Act I

Act II -
The one with a community

You are not alone

Chaos Engineer, SRE, DevOps, SysAdmin....

Any engineer

In fact any stakeholder

Not just Feature Velocity
But also System Reliability

CNCF Working Group

Proposal

https://github.com/chaoseng/wg-chaoseng

Strong signal that reliability matters to the Cloud Native ecosystem

Deliverables and challenges?

Deliverable 1: Whitepaper

CNCF WG Whitepaper

What it is not:

- Not a specification/standard
- Not dogmatic
- Not a HOWTO

CNCF WG Whitepaper

So, what is it?

- Shared understanding
- Product/Solution Agnostic
- A starting line for users’ journey into Chaos Engineering
- An industry effort to refine the practice
- It’s not about giving solutions but expressing how Chaos Engineering is one tool to reliability problems!

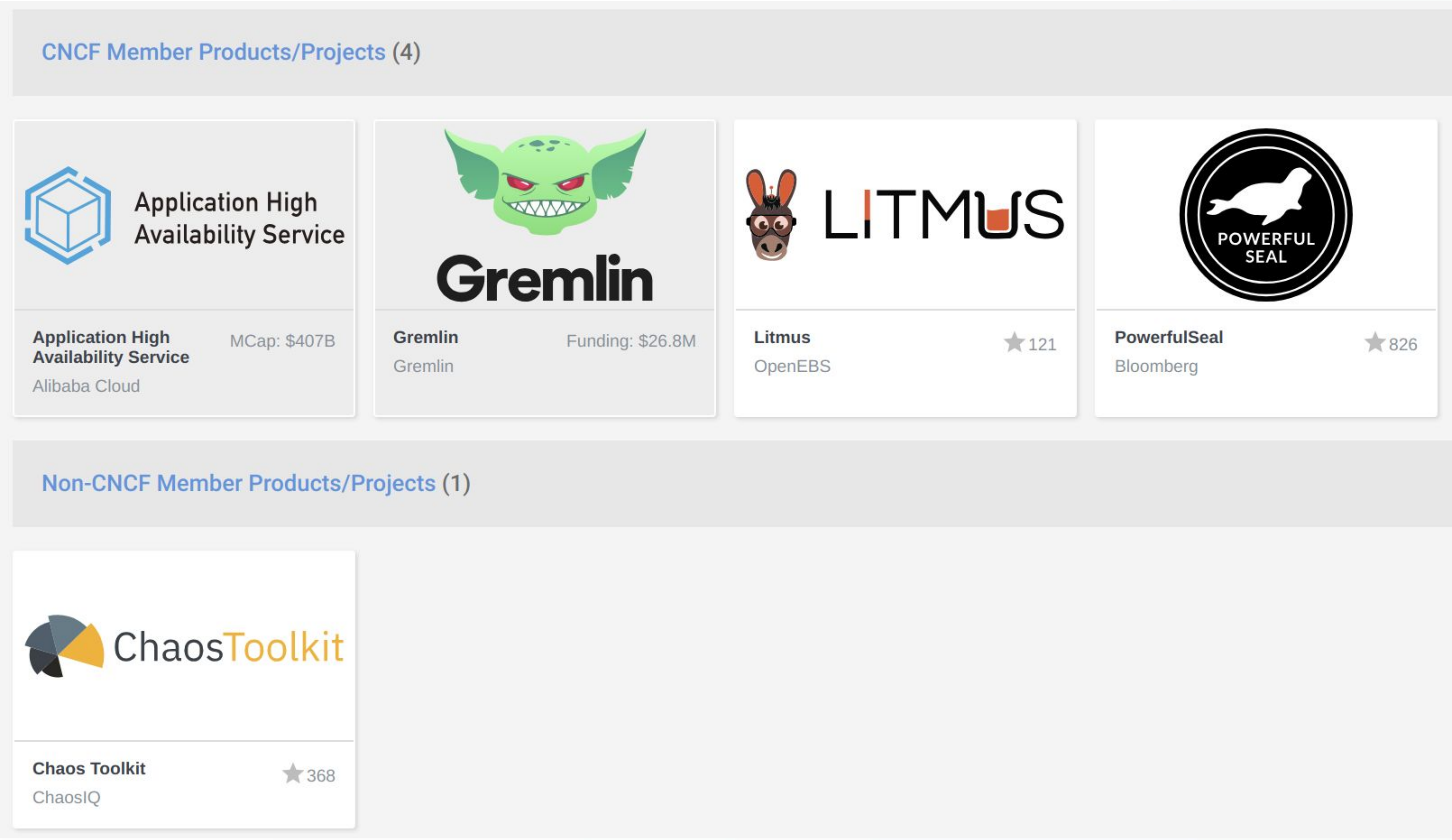
CNCF WG Whitepaper

Improvements

- A new practice so where to draw a line?
- How to better engage with the community?
- Everyone has failures and recovery stories to share!

Deliverable 2: Landscape

CNCF WG Landscape



CNCF Member Products/Projects (4)

Application High Availability Service MCap: $407B
Alibaba Cloud

Gremlin Funding: $26.8M
Gremlin

Litmus ★ 121
OpenEBS

PowerfulSeal ★ 826
Bloomberg

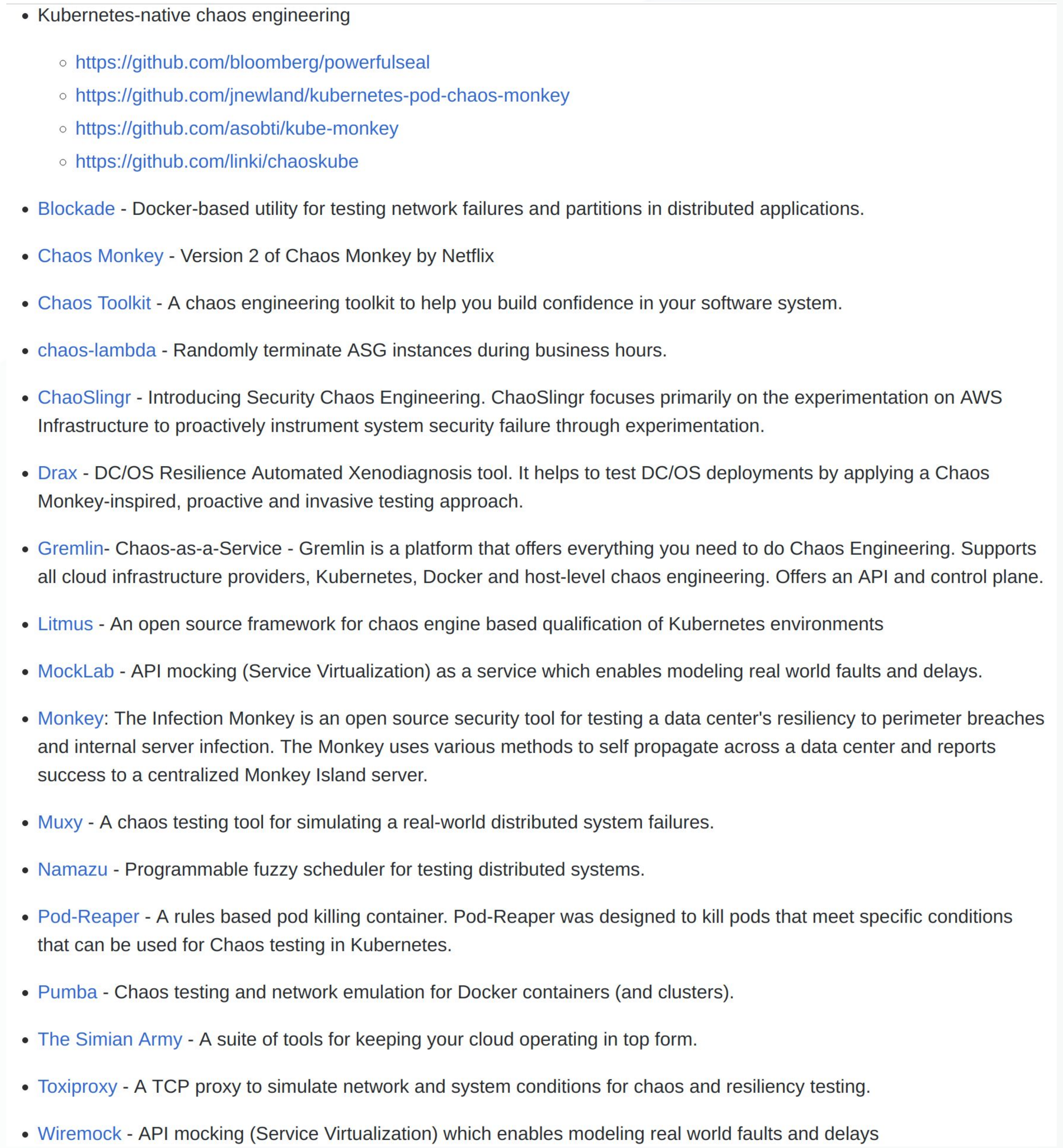
Non-CNCF Member Products/Projects (1)

Chaos Toolkit ★ 368
ChaosIQ

CNCF Landscape

Some awesome tools

But segmented and sparse



- Kubernetes-native chaos engineering
  - https://github.com/bloomberg/powerfulseal
  - https://github.com/jnewland/kubernetes-pod-chaos-monkey
  - https://github.com/asobti/kube-monkey
  - https://github.com/linki/chaoskube
- Blockade - Docker-based utility for testing network failures and partitions in distributed applications.
- Chaos Monkey - Version 2 of Chaos Monkey by Netflix
- Chaos Toolkit - A chaos engineering toolkit to help you build confidence in your software system.
- chaos-lambda - Randomly terminate ASG instances during business hours.
- ChaoSlingr - Introducing Security Chaos Engineering. ChaoSlingr focuses primarily on the experimentation on AWS Infrastructure to proactively instrument system security failure through experimentation.
- Drax - DC/OS Resilience Automated Xenodiagnosis tool. It helps to test DC/OS deployments by applying a Chaos Monkey-inspired, proactive and invasive testing approach.
- Gremlin- Chaos-as-a-Service - Gremlin is a platform that offers everything you need to do Chaos Engineering. Supports all cloud infrastructure providers, Kubernetes, Docker and host-level chaos engineering. Offers an API and control plane.
- Litmus - An open source framework for chaos engine based qualification of Kubernetes environments
- MockLab - API mocking (Service Virtualization) as a service which enables modeling real world faults and delays.
- Monkey: The Infection Monkey is an open source security tool for testing a data center's resiliency to perimeter breaches and internal server infection. The Monkey uses various methods to self propagate across a data center and reports success to a centralized Monkey Island server.
- Muxy - A chaos testing tool for simulating a real-world distributed system failures.
- Namazu - Programmable fuzzy scheduler for testing distributed systems.
- Pod-Reaper - A rules based pod killing container. Pod-Reaper was designed to kill pods that meet specific conditions that can be used for Chaos testing in Kubernetes.
- Pumba - Chaos testing and network emulation for Docker containers (and clusters).
- The Simian Army - A suite of tools for keeping your cloud operating in top form.
- Toxiproxy - A TCP proxy to simulate network and system conditions for chaos and resiliency testing.
- Wiremock - API mocking (Service Virtualization) which enables modeling real world faults and delays

CNCF WG Landscape

Challenge - What are meaningful categories?

- Fault Injection, Orchestration
- Layer: infrastructure, platform, application
- Target: network, cpu…

Many dimensions!

Need community feedback to find the right approach for users to sense which tools to try and how they can complement each other

Short-term Milestone

Complete WHITEPAPER

Respond to Landscape challenges

Submit WG to CNCF TOC

The community needs to make a stand about reliability!

End of Act II

Act III -
The ones with a plan

Chaos Engineering must not be reduced to its tooling or definition

Chaos Engineering is a deliberate practice to explore the unknown to surface new knowledge

But why Chaos Engineering?

Because Reliability - in all its facets - is strategic to everyone

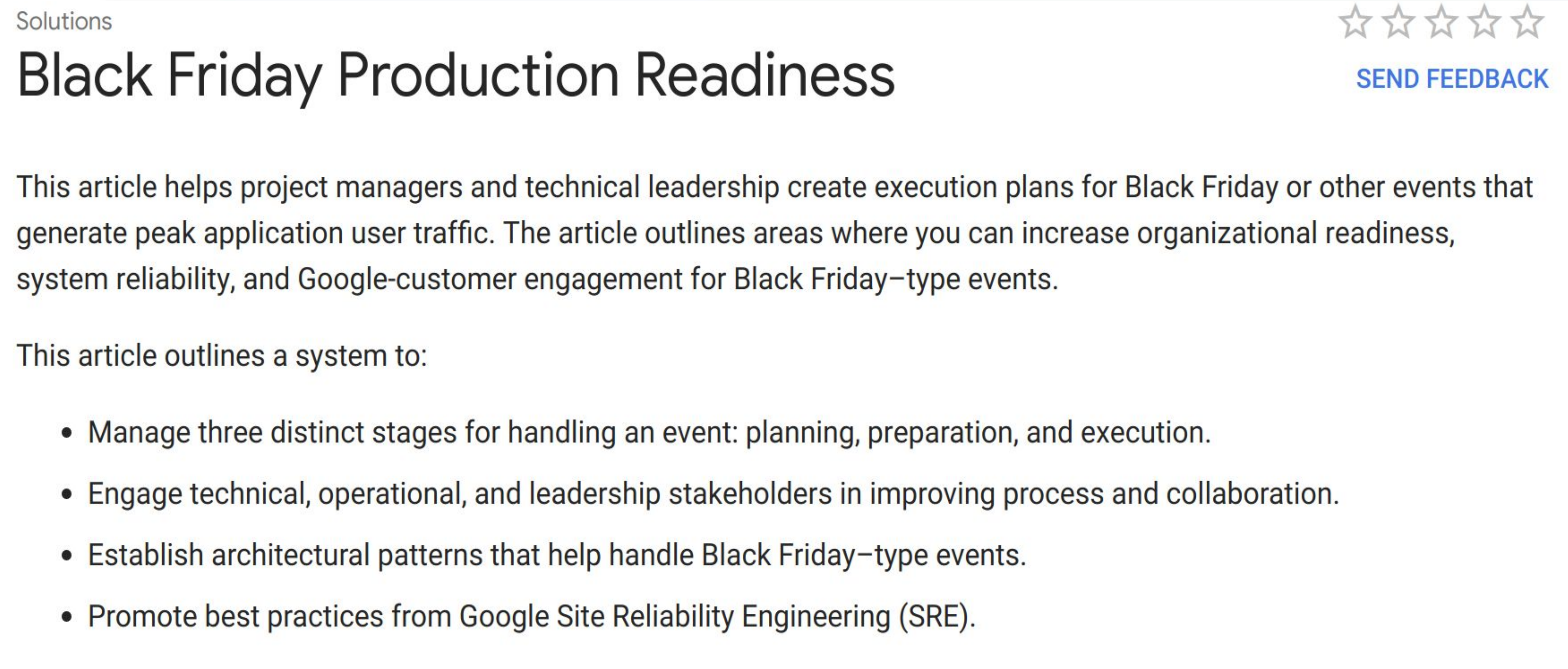
To Collaborate, on that Crucial Requirement for Reliability, we need a Platform to Share our Knowledge

A short detour...

Google Cloud Recommendations for your Black Friday

- Awesome read
- Full of tips (planning, playbooks, postmortems…)
- Mention Disaster Recovery and Chaos Monkey

BUT wouldn't it be better if it offered runnable experiments?



Solutions

Black Friday Production Readiness

This article helps project managers and technical leadership create execution plans for Black Friday or other events that generate peak application user traffic. The article outlines areas where you can increase organizational readiness, system reliability, and Google-customer engagement for Black Friday–type events.

This article outlines a system to:

- Manage three distinct stages for handling an event: planning, preparation, and execution.
- Engage technical, operational, and leadership stakeholders in improving process and collaboration.
- Establish architectural patterns that help handle Black Friday–type events.
- Promote best practices from Google Site Reliability Engineering (SRE).

The Chaos Engineering Principles have given us the vocabulary

http://principlesofchaos.org/

Hypothesis

Ask a question

Experiment

Procedure to operate the question

Observation

Collect of data for drawing a conclusion

Finding

Statement about the hypothesis validity

Chaos Engineering is **Science** and brings you a **Protocol** for exploring your system’s reliability

Chaos Engineering is **Open Science** For Software/System Engineering

We Must Strive to
Share Experiments
and **Findings** to **Help**
Everyone Building
More Reliable Systems

To Unlock that Potential, the Industry must work towards Open Standards and API

Kubernetes has paved the way

Federated across the industry

Serverless WG is a good example

See Cloud Events

https://cloudevents.io/

Open Chaos Initiative

Share experiments as articles of interest across teams, across organisations and even between organisations.

Share experimental findings such that others can peer review and even suggest improvements and comparisons with their own findings based on similar experiments.

Share, collaborate and enable collective learning on how to improve the resilience and technical robustness of systems.

https://openchaos.io/

Let’s recall what Nasa discovered...

Achieving a ***cross-fertilization and engineers’ experience*** *in analysis, design, test or operations* ***will always be important****.*

End of Act III

But beginning of this Movement

Thank You

Sylvain Hellegouarch
ChaosIQ CTO

https://chaosiq.io
@lawouach

Explore further…

- Principles of Chaos Engineering http://principlesofchaos.org/

- Open Chaos Initiative https://openchaos.io/

- CNCF Chaos Engineering WG Whitepaper https://github.com/chaoseng/wg-chaoseng/blob/master/WHITEPAPER.md

- Experiment/Journal Open API https://docs.chaostoolkit.org/reference/concepts/

- How complex systems fail https://www.researchgate.net/publication/228797158_How_complex_systems_fail

- NASA Failures Case Studies https://nsc.nasa.gov/resources/case-studies