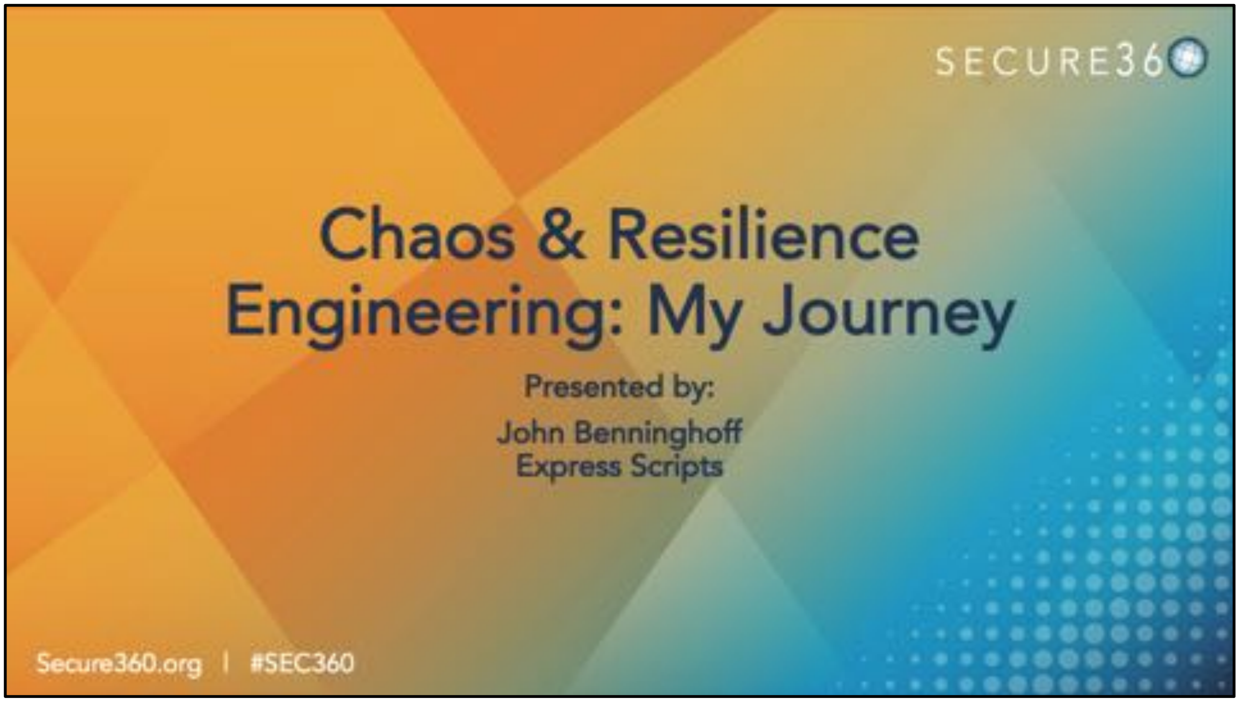


SECURE360

Chaos & Resilience Engineering: My Journey

Presented by:

John Benninghoff
Express Scripts

Secure360.org | #SEC360

My “COVID” photo

About Me: I’ve been in security for 20 years, I’ve spent about half that time focusing on infrastructure, half on software development, with several different orgs. Currently leading the application security practice at Express Scripts (Cigna).

ACT I: My journey to find chaos engineering

ACT II: Chaos engineering and how resilience engineering complements it

ACT III: What I’ve learned so far

END: How to get started

Story begins in 2008...

SECURE360 conference

ACT I: my journey to find chaos engineering
2008, end of a chapter: laid off, new job, Secure360 2009 @ Miles Edmunson's talk on risk homeostasis theory; what is homeostasis? antilock brakes. This insight into people & risk got me thinking about my frustration with the current state of the security industry of 2009...

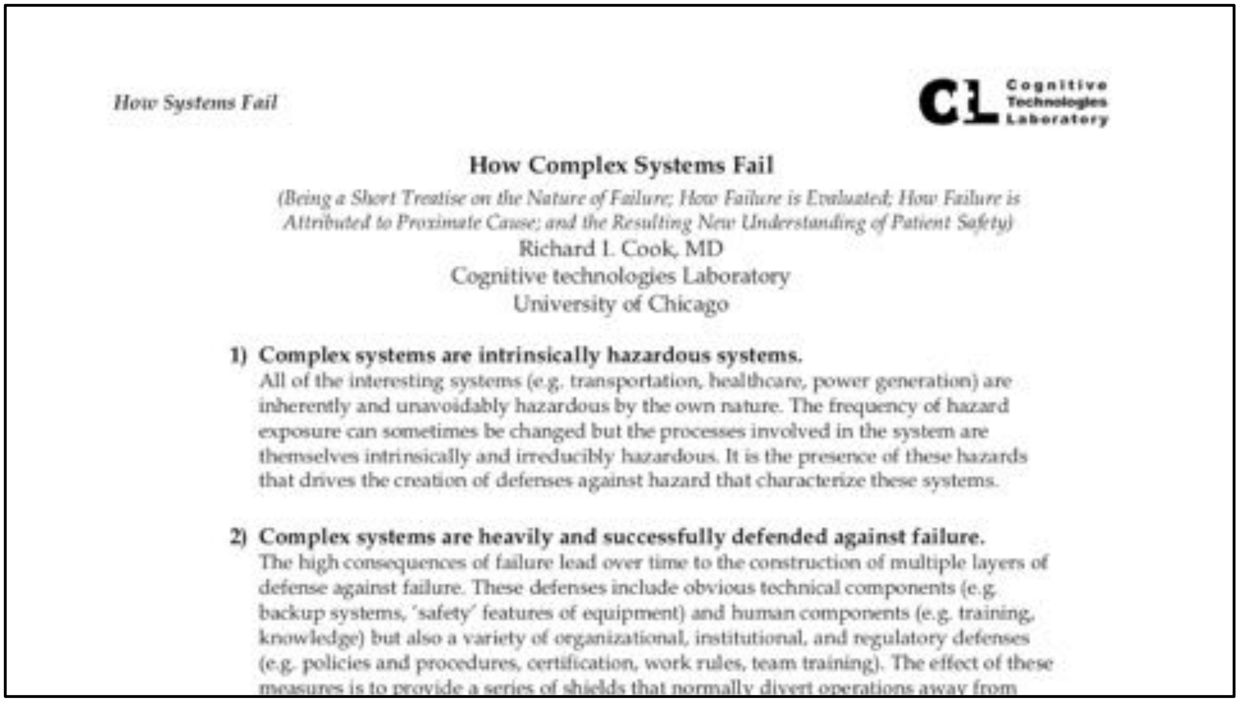
https://en.wikipedia.org/wiki/Risk_compensation#Risk_homeostasis

The 4 pillars of security: physical, technology, policy, people. State of security 2009, “blame the user” Using my interests & training in social science (Economics, behavioral econ) to find a better way. My search for a better way of risk management took me in a different direction, inspired by my grandfather...

Images (L to R):
https://commons.wikimedia.org/wiki/File:Prison_gate_still_life.JPG
https://commons.wikimedia.org/wiki/File:Knams-15-knsq-15.jpg
https://commons.wikimedia.org/wiki/File:Constitution_of_the_United_States,_page_1.jpg
https://commons.wikimedia.org/wiki/File:Software_Developer_at_work_03.jpg

My grandfather: a pilot of ~65 years (15-80), circa 1940, always used his pre-flight checklist, started my interest in aviation safety, *The Checklist Manifesto*. Started reading safety, aviation safety, can we use this for security?, led me to a paper recommended by a colleague…

https://en.wikipedia.org/wiki/The_Checklist_Manifesto



How Systems Fail

CL Cognitive Technologies Laboratory

How Complex Systems Fail

(Being a Short Treatise on the Nature of Failure; How Failure is Evaluated; How Failure is Attributed to Proximate Cause; and the Resulting New Understanding of Patient Safety)

Richard I. Cook, MD
Cognitive technologies Laboratory
University of Chicago

1) **Complex systems are intrinsically hazardous systems.**
All of the interesting systems (e.g. transportation, healthcare, power generation) are inherently and unavoidably hazardous by the own nature. The frequency of hazard exposure can sometimes be changed but the processes involved in the system are themselves intrinsically and irreducibly hazardous. It is the presence of these hazards that drives the creation of defenses against hazard that characterize these systems.

2) **Complex systems are heavily and successfully defended against failure.**
The high consequences of failure lead over time to the construction of multiple layers of defense against failure. These defenses include obvious technical components (e.g. backup systems, 'safety' features of equipment) and human components (e.g. training, knowledge) but also a variety of organizational, institutional, and regulatory defenses (e.g. policies and procedures, certification, work rules, team training). The effect of these measures is to provide a series of shields that normally divert operations away from

"How Complex Systems Fail"; free, 3 pages, changed my thinking, failures are complex and unpredictable, there is no 'root cause', people are both a source of weakness and strength. Looked for related papers on Google...

http://web.mit.edu/2.75/resources/random/How%20Complex%20Systems%20Fail.pdf
video: https://www.youtube.com/watch?v=2S0k12uZR14

Engineering a Safer World, modern view of safety as a sociotechnical system (people & technology, the impact of organizational dynamics on safety & failure), ‘blame is the enemy of safety’, STAMP/STPA Workshop @ MIT, more academic safety science papers, and more searches…

https://mitpress.mit.edu/books/engineering-safer-world
https://psas.scripts.mit.edu/home/

Photo:
https://commons.wikimedia.org/wiki/File:MIT_Building_10_and_the_Great_Dome,_Cambridge_MA.jpg

? Old Library at Trinity College Dublin. Again searching Google: Managing Risk & Systems Change Masters @ TCD; introduction to safety science, risk assessment, managing change, design, leadership, and organizational development – furthered my belief of the need for empathy in understanding failures, complexity of sociotechnical systems, which led me back to…

https://psychology.tcd.ie/postgraduate/msc-riskandchange/

Photo: https://commons.wikimedia.org/wiki/File:Trinity_college_library.jpg

SECURE360 conference

Back at Secure360 2019 @ Matt Schillerstrom & Lenny Sharpe's talk on Chaos engineering (somewhat familiar; but had evolved). Overlap between Chaos Engineering reading list and my masters’ program. Found John Allspaw's journey at Etsy (CTO Etsy, Blameless post-mortem, master’s degree program in Human Factors and Systems Safety at Lund University, Sweden), found the others in technology that are already using safety!

https://secure360.org/
http://www.humanfactors.lth.se

CHAOS MONKEY
NETFLIX
SIMIAN ARMY
O'REILLY
NETFLIX
Chaos Engineering
Building Confidence in System Behavior through Experiments
Casey Rosenthal, Lorin Hochstein, Aaron Blohowiak, Nora Jones & Ali Basiri

ACT II: Chaos engineering and how resilience engineering complements it "so what is Chaos Engineering?"
Chaos Monkey, Simian Army, Chaos Engineering: history of how Netflix CM (2011) evolved into CE. (moving to cloud, "how do we know our self-healing systems work?"), running CM during the day, Simian Army has been retired/moved to other projects, CM is still active, requires spinnaker. CE now a practice! Many tools, including commercial tools available. **New CE book 2020.**
principlesofchaos.org 4 rules on experiments.

https://github.com/Netflix/chaosmonkey
https://github.com/Netflix/SimianArmy
https://www.oreilly.com/library/view/chaos-engineering/9781491988459/
http://shop.oreilly.com/product/0636920203957.do
https://principlesofchaos.org/

Jordan: (on our middleware team, problem: unowned systems) - how takedowns are like chaos experiments – success rate: over a hundred takedowns, one small failure – you may find that you’re already doing chaos engineering, just not formally.

CRC Press
Resilience Engineering
Concepts and Precepts
弾
Edited by
ERIK HOLLNAGEL
DAVID D. WOODS
NANCY LEVESON
O'REILLY
Chaos Engineering
Building Confidence in System Behavior through Experiments
Casey Rosenthal, Lorin Hochstein, Aaron Blohowiak, Nora Jones & Ali Basiri

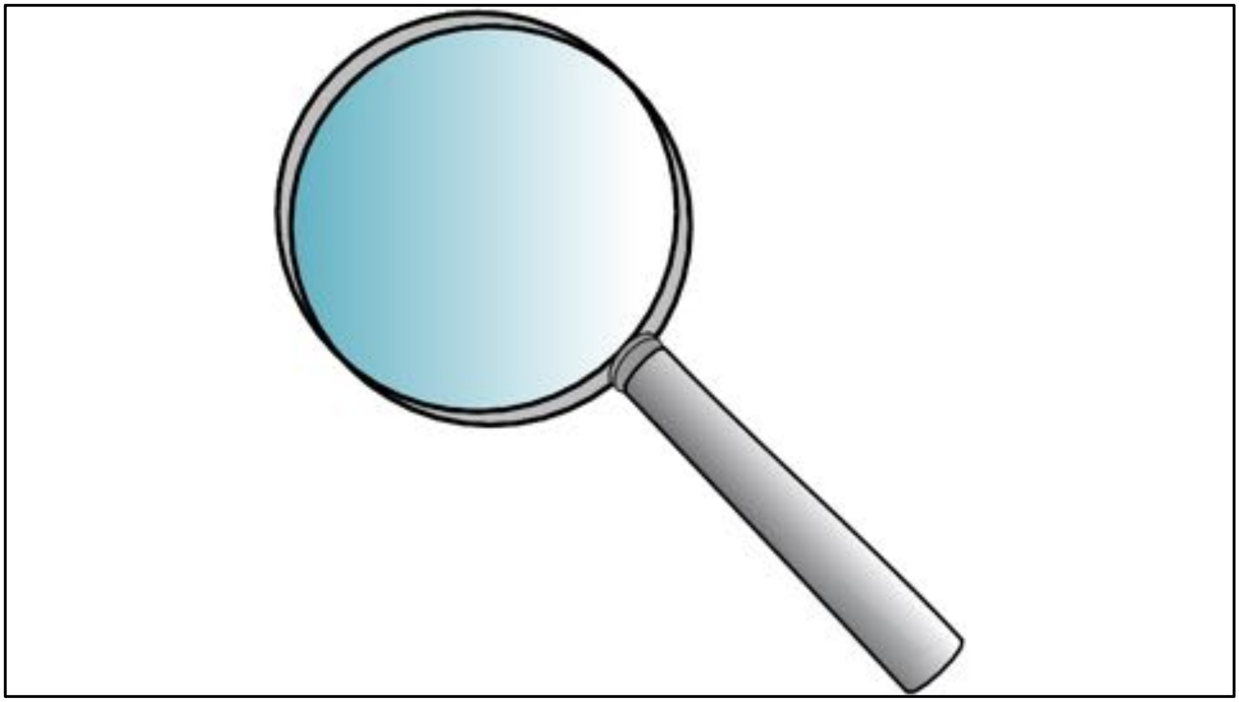
So why are we doing these experiments? Chaos & Resilience engineering complement each other: with CE, you break things, with RE, you understand why they break & how to make the system better able to withstand breakage. RE: Book 2006 - conf 2005 – what about Resilience?...

https://www.crcpress.com/Resilience-Engineering-Concepts-and-Precepts/Woods-Hollnagel/p/book/9780754649045

Image: https://commons.wikimedia.org/wiki/File:Yin_yang.svg

Image? Erik Hollnagel. Definition of Resilience (ecology). Hollnagel's four potentials – (respond, monitor, learn, anticipate) chaos experiments are for learning & anticipating – we’re all using these 4 in our daily work

http://erikhollnagel.com/ideas/resilience%20assessment%20grid.html



So, how do you set up the experiment? Before the experiment, the first 2 potentials, respond & monitor: ie observability, so that we can abort if things go wrong. Target's experience with teams volunteering for chaos, need: logging, monitoring, playbooks/runbooks.

Image: https://commons.wikimedia.org/wiki/File:Magnifying_glass_01.svg

Debriefing

After the experiment, second two potentials, Learn & anticipate. “Blameless” post-mortem. Focus is on learning, not blame. (hindsight bias; Just Culture).

https://codeascraft.com/2016/11/17/debriefing-facilitation-guide/

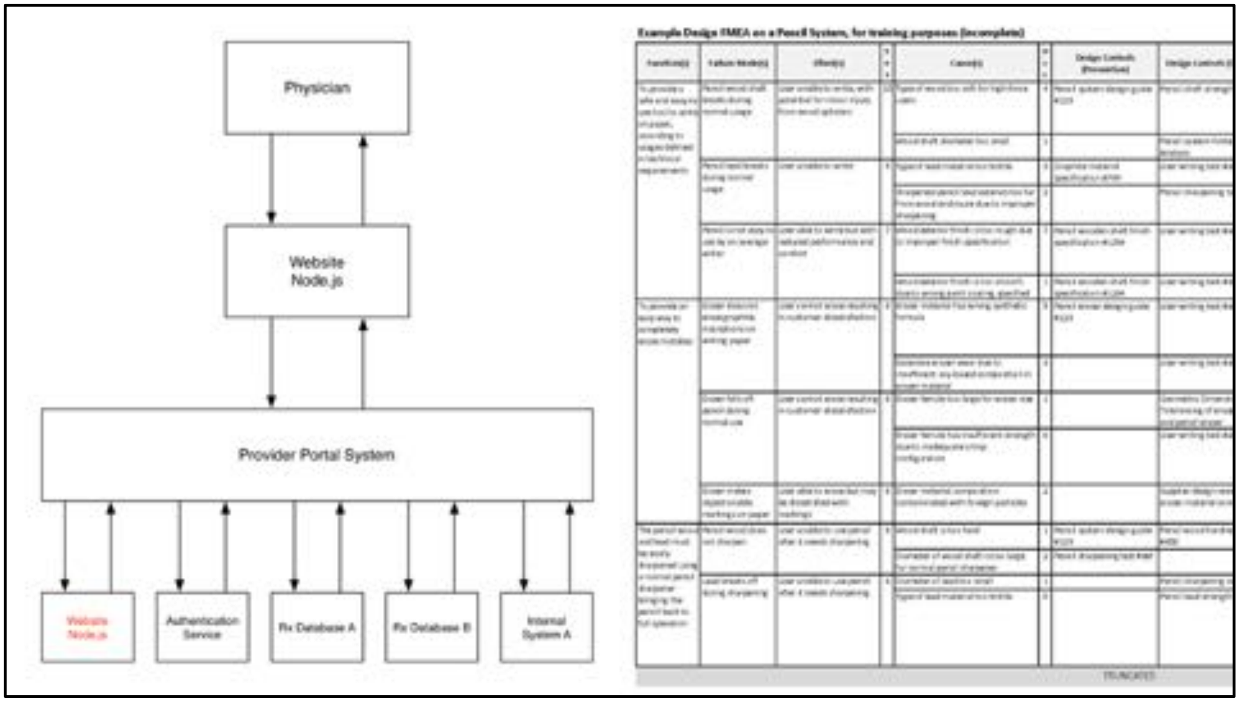
After learning, how do we change the *system* to improve resilience - not a list of remediation items to fix. Solution is to find how to change, not do more work. (Even in aviation safety, it's hard to change after an accident, biggest changes happen after biggest accidents)

Images:
https://commons.wikimedia.org/wiki/File:Checklist.svg
https://commons.wikimedia.org/wiki/File:ProhibitionSign2.svg

ACT III: What I’ve learned so far
Lesson 1: **“Incident Management Teams in Technology are similar to those in Oil & Gas”**: did a case study of our Technology IMT. ITIL defined goals, process, but not sensemaking or social aspects: Importance of Situation Awareness, behaviors/skills for successful IMT, communication structure.
"Incident Command Skills in the Management of an Oil Industry Drilling Incident: a Case Study" (Crichton paper) identified similar skills for Oil & Gas (situation assessment, Decision making, Teamwork, and Leadership), a similar tiered command/communication structure.
Takeaway: Crichton paper suggests training plan for IMT skills, communication structure

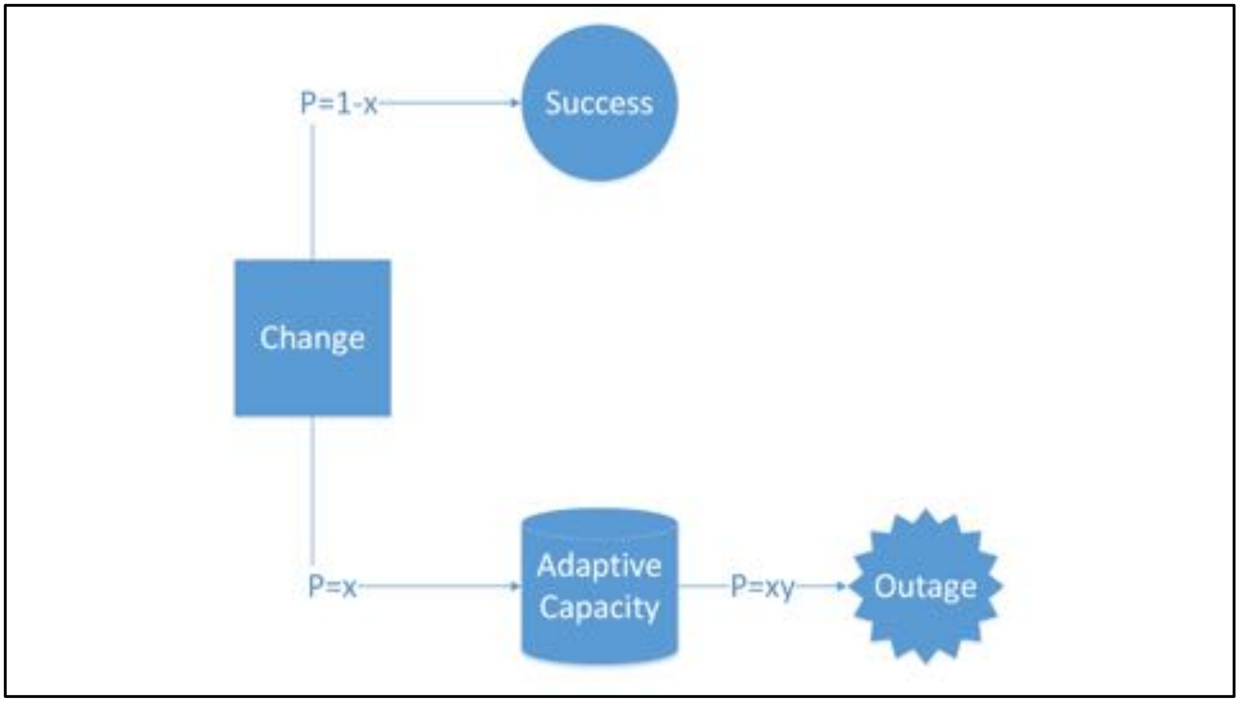
Crichton: ref: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-5973.2005.00466.x PDF: http://www.academia.edu/download/61613058/Crichton_Lauche_Flin_JCCM_200520191226-4428-hgnlq1.pdf

L Photo: https://commons.wikimedia.org/wiki/File:NOC-IUPUI.jpg
R Photo: https://commons.wikimedia.org/wiki/File:Gulf_Offshore_Platform.jpg



Lesson 2: **“Safety has risk assessment methods that can be applied to computer systems”**: Did a comparative review of 2 risk analysis methods, NIST 800-30 and STPA. (Considered but rejected COBIT 5 due to lack of completeness & public availability). Trade-offs with NIST (list approach) vs STPA (more complete, steeper learning curve, longer). NIST focused on IT risks, (DR plan) STPA engagement of non-technical stakeholders correctly identified data integrity as biggest risk. Use of FMEA for Gamedays. [started @ Amazon early 2000s by Jesse Robbins]
Takeaway: use FMEA, other safety methods like STPA as part of CE/RE

NIST 800-30: https://csrc.nist.gov/publications/detail/sp/800-30/rev-1/final
STPA Handbook: http://psas.scripts.mit.edu/home/materials/
FMEA: https://en.wikipedia.org/wiki/Failure_mode_and_effects_analysis
GameDay: https://queue.acm.org/detail.cfm?id=2371297



Lesson 3: **“Changes cause outages”** Wikipedia: “One of the biggest causes of downtime is misconfiguration, where a planned change goes wrong.” Change failure model.
Takeaway: Getting better: lower change failure rate x,y. Recover faster! (it’s not a failure if nobody notices) Spread out changes over time.

https://en.wikipedia.org/wiki/Downtime

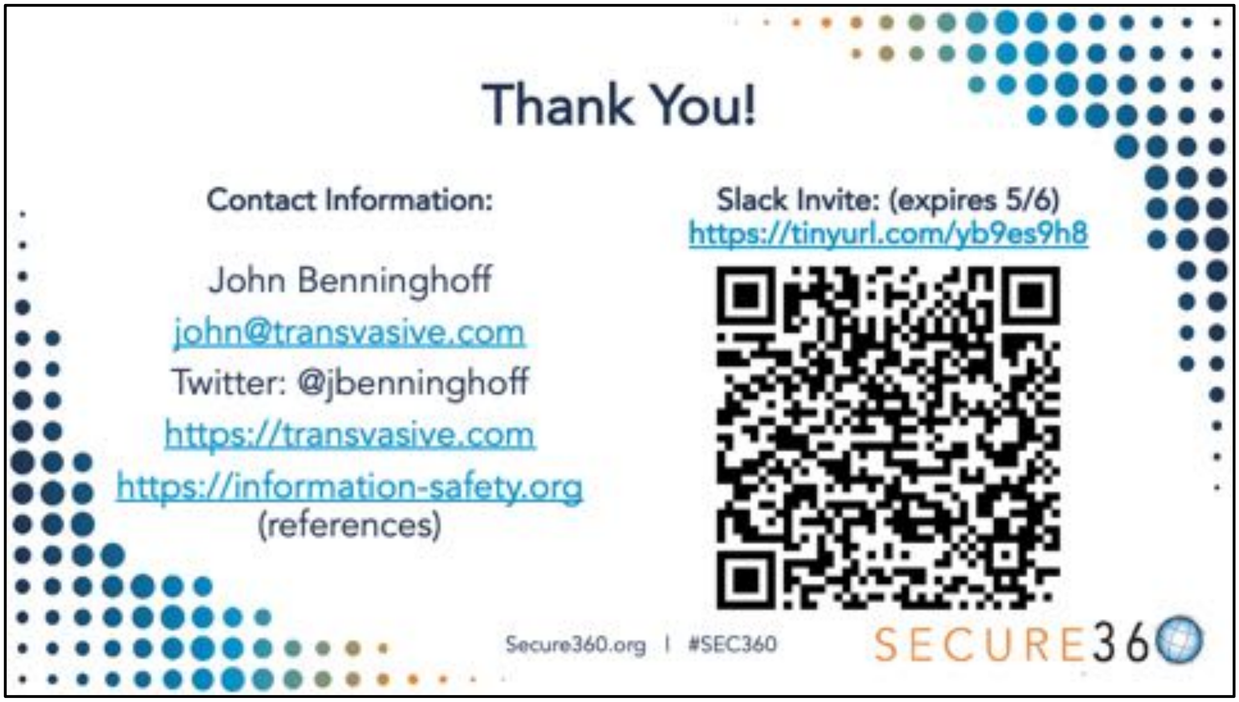
How to get started

- Chaos Engineering – break stuff!
  - Twin Cities CE (meetup.com)
  - dastergon/awesome-chaos-engineering
  - Gremlin
- Resilience Engineering – fix stuff
  - lorin/resilience-engineering
  - learningfromincidents.io
- **information-safety.org!**

END: How to get started

https://www.meetup.com/Twin-Cities-Chaos-Engineering-Community/
https://github.com/dastergon/awesome-chaos-engineering
https://www.gremlin.com
https://github.com/lorin/resilience-engineering (Where do I start? page)
https://www.learningfromincidents.io
https://cloud.google.com/devops
https://itrevolution.com/book/accelerate/
https://www.information-safety.org

Photo:
https://commons.wikimedia.org/wiki/File:Repair_Cafe_by_Ilvy_Njiokiktjien.jpg



Thank You!

Contact Information:

John Benninghoff

john@transvasive.com

Twitter: @jbenninghoff

https://transvasive.com

https://information-safety.org
(references)

Slack Invite: (expires 5/6)
https://tinyurl.com/yb9es9h8

Secure360.org | #SEC360

SECURE360

Thank you to those who helped me in my journey to Chaos & Resilience Engineering
References for this talk @ information-safety.org
Questions?

https://www.information-safety.org
https://transvasive.com