Evaluating complex adaptive systems

What can we learn from swans, cats, dragons, toads, bugs and oxymorons?

DEOS, OSD Symposium March 2012, Tokyo

Robin E Bloomfield
Adelard LLP and CSR City University London

reb@adelard.com
reb@csr.city.sc.uk
College Building, City University, London EC1V 0HB
Tel: +44 20 7832 5850 (sec Adelard)
Tel: +44 20 7040 8420 (sec CSR)
Introduction

• Focus on evaluation

• Challenge of complex adaptive systems
  • What is current engineering state of practice
  • What are particularities of complex, adaptive systems

• Future directions
  • Engineering complex adaptive systems
Flash crash - May 2010, ~$1tr

Preliminary Findings Regarding the Market Events of May 6, 2010

Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues
Complex adaptive systems

• Complex adaptive
  • Large scale, complicated
  • System properties and emergent behaviours, tipping points, phase transitions, cascades
  • Adaptive with people and society as part of the system or part of openness
  • Limits of reductionism, potential for surprise
• Finance sector and computer based trading
  • “Meltdown in the markets: a nuclear perspective”
  • Swans, Dragons, Cats, Toads and Bugs
Nuclear reactor protection and control
Protection parameters

- Protection parameters
- Pressure
- Temperature
- Plant damage boundary
- Protection envelope
- Normal operation
Systemic risk definition and evaluation

Intolerable risk level

Risk cannot be justified on any grounds

Tolerable only if risk reduction is impracticable or if its cost is grossly disproportionate to the improvement gained

Tolerable if cost of reduction would exceed the improvement gained

The ALARP region

Broadly acceptable region

Negligible risk level

No need for detailed working to demonstrate ALARP
Safety Assessment Principles (SAPS)

• SAPS cornerstone of UK licensing approach
  • Goal based, based on eight fundamental principles
    • responsibility,
  • leadership,
  • understanding,
  • fairness,
  • inter-generational issues,
  • protection,
  • recovery and mitigation.
Safety cases – regulatory obligation

• Safety cases are required by licence conditions

• The Conditions are non-prescriptive and set goals that the licensee is responsible for meeting, amongst other things by applying detailed safety standards and establishing safe procedures for the facilities.

• A "safety case" is defined as

  • the document or documents produced by the licensee documentation to justify safety during the design, construction, manufacture, commissioning, operation and decommissioning phases of the installation.

• Safety Assessment Principles (SAPs) describe safety case process and principles to be covered

  • “.... the establishment of and compliance with appropriate standards and practices throughout the software development life-cycle should be made, commensurate with the level of reliability required, by a demonstration of ‘production excellence’ and ‘confidence-building’ measures.”
Communication and reasoning

• Structured safety and assurance cases have two essential roles:
  • communication is an essential function of the case, from this we can build confidence
    • boundary objects that record the shared understanding between the different stakeholders
  • a method for reasoning about dependability (safety, security, reliability, resilience ...) properties of the system
  • Both are required to have systems that are trusted and trustworthy
• “a documented body of evidence that provides a convincing and valid argument that a system is adequately safe for a given application in a given environment”
In practice ... the engineering
In practice …
Reactor protection systems

- Primary protection system
- Secondary protection system
- Tertiary segregation

Voting system

Shut down

Plant signals

Diverse signals A, B, C, D redundant channels

Segregation

Communication between channels

Also ESFAS for post trip control and protection
Techniques

- goal-based assessments of attributes (e.g. demonstration of accuracy, reliability, correctness etc.)

- assessment of potential vulnerabilities and formal demonstration of integrity properties (e.g. deadlock, run time exceptions caused by divide by zero)

- static analysis of software source code with mathematically formal demonstration of functional correctness of the code with respect to a specification

- dynamic analysis of the software with test coverage metrics; statistical testing commensurate with the level of reliability required

- demonstration that any software tools do not comprise the integrity of the code.

- a post-installation demonstration, performed independently of the suppliers, that the system performs to requirements

- analysis of experience from previous use of the system (and previous versions of the system) in the nuclear industry or elsewhere
Do everything and do it at least twice.

- very careful lifecycle
- static analysis and proof
- statistical testing
- negative testing
- compliance
Some issues and challenges

In engineering trusted systems

• Defence in depth and diversity
• Interdependencies and resilience
• Adaptation and socio-technical systems
• Confidence and doubt
• Learning form experience – difficult and important
• Security and threats
The human element

- people as a source of resilience, as a threat and as victims
- don’t blame the pilot – but the system
- 30 min rule, tempo and expertise
Socio-technical, adaptation

• A socio-technical perspective on assurance cases:
  • In addition to claims that physical hazards, security threats have been addressed
  • Define a range of vulnerabilities (narrow scope, misaligned responsibilities, undifferentiated users, adaptation, automation biases, non-independence of arguments) and develop arguments of how they might be addressed.
  • Develop methods for review wrt socio-technical issues
  • Ideas taken from EPSRC INDEED and DIRC projects
Defence in depth and diversity

- Improvement trend (low pfd subsets)
  Knight and Leveson

Van der Meulen

Benefits of diversity
Conservative approach

- Figure 6a shows a typical ‘real’ distribution, $f(p)$ that satisfies the expert’s belief:

$$P(pfd < y) = 1 - x$$

Of all the many such distributions that satisfy his belief, Figure 6b shows the most conservative: here all the probability mass for the interval $(0, y)$ is concentrated at $y$, and all the mass for the interval $(y, 1)$ is concentrated at 1. Then:

$$P(\text{system fails on randomly selected demand } )$$

$$= \int_0^y p \cdot f(p) + \int_y^1 p \cdot f(p) < y(1 - x) + x$$

$$= x + y - xy = z, \text{ say}$$
Trojan Horse Virus, 28 yrs before Stuxnet

- SCADA Attack In June 1982, malicious code implanted in SCADA software caused a rupture of the trans-Siberian gas pipeline.

- According to Reed, “the pipeline software that ran the pumps, turbines, and valve settings was programmed to produce pressures far beyond those acceptable to the pipeline joints and welds.” This caused the largest non-nuclear explosion on record (3KT).

- Thomas Reed, At the Abyss: An Insider’s History of the Cold War

- This is disputed by other sources, issue of attribution
Critical infrastructure interdependencies
Analysis of field experience

• Use of field experience to complement the justification
  • When information is lacking
  • To gain confidence
  • Identify vulnerabilities

• Review of defects to check impact on application

• Looking for evidence of reliability growth
  • Long-term, conservative models

• Depends on quality of data
  • Normally only part of case
Summary

• Established approaches for evaluating and communicating trust in critical computer based systems

• Plenty of issues
  • Defence in depth and diversity
  • Interdependencies and resilience
  • Adaptation and socio-technical systems
  • Confidence and doubt
  • Learning form experience – difficult and important

• Security and threats
Flash crash - May 2010

Preliminary Findings Regarding
the Market Events of May 6, 2010

Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on
Emerging Regulatory Issues
Black swans - power laws and fat tails

Fig. 4 Distribution of areas of landslides triggered by the Jan. 17, 1994 Northridge earthquake, California, USA. The straight line in log-log scale qualifies a power law distribution with exponent \( \mu \approx \frac{2}{3} \). Reproduced from Turcotte [5].

Fig. 5 Distribution areas of forest fires and wildfires in the United States and Australia. (A) 4284 fires on U.S. Fish and Wildlife Service Lands (1986-1995); (B) 120 fires in the western United States (1150-1960); (C) 164 fires in Alaskan boreal forests (1990-1991); (D) 298 fires in the ACT (1926-1991). The straight line in log-log scale qualifies a power law distribution with exponent \( \mu \approx 0.3-0.5 \). Reproduced from Malamud et al. [6].

Fig. 6 The number density \( N(M) \) of rain events versus the event size \( M \) (open circles) on a double logarithmic scale. Events are collected in bins of exponentially increasing widths. The horizontal position of a data point corresponds to the geometric mean of the beginning and the end of a bin. The vertical position is the number of events in that bin divided by the bin size. To facilitate comparison, the number of events are rescaled to annual values by dividing by the fraction of a whole year during which the data were collected. The straight line in log-log scale qualifies a power law distribution with exponent \( \mu \approx 0.4 \). Reproduced from Peters and Christensen [7].

2.2. Power law distributions in social systems

Fig. 7 Survival distribution of positive (continuous line) and negative daily returns (dotted line) of the Dow Jones Industrial Average index over the time interval from May 27, 1896 to May 31, 2000, which represents a sample size of \( n=28\,415 \) data points. The straight part in the tail in this log-log scale qualifies a power law distribution with exponent \( \mu=3 \). Reproduced from Malevergne et al. [8].

Visually power laws but see critiques
Financial crashes – no precursors?

- Power law distributions embody the notion that extreme events are not exceptional events.
- Instead, extreme events should be considered to be rather frequent and to result from the same organization principle(s) as those generating other events: because they belong to the same statistical distribution, this suggests common generating mechanism(s).
- A great earthquake is just an earthquake that started small ... and did not stop; it is inherently unpredictable due to its sharing of all the properties and characteristics of smaller events (except for its size), so that no genuinely informative precursor can be identified.
- View expounded in formulation of the concept of self-organized criticality and also the espoused by the “Black Swan Theory” which views high-impact rare events as unpredictable.

(adapted from D Sornette, Why stock markets crash)
Dragons

- Even with power laws and other fits get outliers

---

**Fig. 14** A short time series of the DJIA on which the red arrows define two drawdowns, while the green arrow defines a drawup. There are many possible definitions of drawdowns (and their symmetric drawups). Details are given in Johansen and Sornette [29-31].

Since we are interested in characterizing the statistics of extreme events, to illustrate our claim, consider the simplified textbook example of a crash that occurs over a period of three days, as a sequence of three successive drops of 10% each summing up to a total loss of 30%. Now, a 10% market drop, that occurs over one day, can be seen in the data of the Nasdaq composite index to happen on average once every four years. Since there are approximately 250 trading days in a year, a 10% market drop is thus a 10^{-3} probability event. What is the probability for observing three such drops in a row? The answer is (10^{-3})^3 = 10^{-9}. Such one-in-one-billion event has a recurrence time of roughly 4 million years! Thus, it should never be observed in our short available time series. However, many crashes of such sizes or larger have occurred in the last decades all over the world.

What is wrong with the reasoning leading to the exceedingly small 10^{-9} probability for such a crash? It is the assumption of independence between the three losses! In contrast, our claim is that financial crashes are transient bursts of dependence between successive large losses. As such, they are missed by the standard one-point statistics consisting in decomposing the runs of losses into elementary daily returns. With some exaggeration to emphasize my message, I would say that by cutting the mammoth in pieces, we only observe mice.

We thus propose to analyze drawdowns (and their symmetrical drawups), because they are better adapted to capture the risk perception of investors, and therefore better reflect the realized market risks. Indeed, we demonstrate below that the distributions of drawdowns diagnose efficiently financial crashes, which are seen as dragon-kings, i.e., special events associated with specific bubble regimes that precede them.

**Fig. 15** Distribution of drawdowns D for the Nasdaq Composite index, showing several “outliers” in the tail, that qualify as dragon-kings. It turns out that these anomalous events are precisely associated with documented crashes, such as the crash of the ITC bubble in April 2000 or the Oct. 1987 World stock market crash. The continuous line is the fit to the stretched exponential distribution ~ e^{-D/D_0} c with an exponent c ≈ 0.8. Reproduced from Johansen and Sornette [30].

**Fig. 15** shows the distribution of drawdowns obtained from the Nasdaq composite index over a 20-year period, which includes several great crashes shown by the arrow. As analyzed carefully by Johansen and Sornette [29-32], about 99% of the drawdowns can be represented nicely by a common distribution, which turns out to have a tail slightly fatter than an exponential. And the remaining few events have been found to be statistically different: the hypothesis that they belong to the same distribution as 99% of the population of the other drawdowns is rejected at the 99.9% confidence level [32,33]. Fig. 16 presents further evidence for 30 individual US companies of the existence of dragon-kings in the distribution of drawdowns (run of losses) and drawups (runs of gains).

**Fig. 16** Same as Fig 15 for 30 major US companies. This figure shows in addition the distribution of drawups, i.e., runs of gains, depicted on the right side of the peak. The drawdown/drawup distribution of each company has been scaled by its corresponding scale factor D_0 defined in the stretched exponential fit mentioned in Fig.15, so as to be able to superimpose the 30 distribution functions. The collapse of the 30 curves is good for the bulk but fails in the tails, where dragon-kings can be observed.

The evidence of dragon-kings reported by Johansen and Sornette [29-33] encompasses exchange markets (US dollar against the Deutsch Mark and against the French agglomerations: stretched exponential and "King effect"

\[ y = -3.263x + 13.682 \]

\[ R^2 = 0.9944 \]

Paris = King

**Fig. 15**: Distribution of drawdowns D for the Nasdaq Composite index, showing several “outliers” in the tail, that qualify as dragon-kings. It turns out that these anomalous events are...
Predicting bubbles

\[
\log p(t) = A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \ln(t_c - t) - \phi)
\]
Pencil analogy

Long-term collaborative or herding behavior produces unstable system
Can measure potential energy changes
Common toads appear to be able to sense an impending earthquake and will flee their colony days before the seismic activity strikes.

The evidence from a population of toads which left their breeding colony three days before an earthquake that struck L’Aquila in Italy in 2009.

How toads sensed the quake is unclear, but most breeding pairs and males fled.

They reacted despite the colony being 74km from the quake’s epicentre, say biologists in the Journal of Zoology.

http://news.bbc.co.uk/earth/hi/earth_news/newsid_8593000/8593396.stm

Abstraction can remove key indicators
Performative models

- In the past, in the engineering domain, the models used to design and assess the risks do not affect the threats or challenges that the system faces.

- Modelling severe weather does not change the wind speed in London (except perhaps via a slow political process and peoples’ behaviour).

- In the financial area this is not the case: models can be what is termed performative, having a direct and unforeseen impact on the markets and how it fails.

- Engineered systems and security risks.

- Knowledge and access to design models may inform an adversary and hence have a potential impact on the threats a system faces.

Trust in computer-based systems (and fpgas)

- Recent work by Johnson
  - analysed a set of 18,520 ultrafast black swan events in stock-price movements between 2006 and 2011
  - empirical evidence for an abrupt transition from a mixed human-machine to a all-machine phase (<650ms for crashes, <950ms for spikes)

- How much do computers need to be trusted – the problem of “bugs”
  - Impact on themselves,
  - On platform they are trading in
  - On wider system
    - issues of risk apportionment, of composition and confidentiality
Engineering complex adaptive systems

- Can simplified envelopes of operation be defined? Can we define control strategy for unstable system?
- What would be the parameters that need to be measured and what are we trying to infer?
- What would the availability and reliability requirement be for such a system. E.g. the probability of failure on demand, the frequency of spurious activation?
- What does “safety” and resilience mean in the HFT context. Would need to consider both the platform and the application?
- What is the balance between automation and operator recovery. Who would design this?
- If crashes are hard to anticipate should the focus not shift to recovery and resilience?
Engineering complex adaptive systems

- swans, cats, dragons, toads and bugs

Developing understanding to engineer complex adaptive systems will change how we currently engineer systems.
Some basic questions

• In engineering complex adaptive systems we would like to answer:
  • What is the system, what are the risks, who has them, and when, are they tolerable?
    • What is the probability of a crisis and what would the consequences be?
    • What are the risks from the complex adaptive system to the economy and society as a whole?
    • What risk levels are tolerable and who/how has this been decided?
Engineering complex adaptive systems

- **Developing intervention strategies for different time bands**
  - Circuit breakers and trip protection
  - Forced diversity in ecology, disrupt corellation
  - Alignment of economic incentives

- **New approaches to risk assessment**
  - Investigating safe envelopes and recovery
  - Specification for single, group and collective
  - Deriving risk targets for properties - failure and success
  - Separation of conventional and complex risks (Fukushima vs flash crash)
  - Emphasis on stability as a dependability attribute
  - The risk of change or not – focus?

Understanding complex systems - dynamics of change, signals
Conclusions - challenges and opportunities

- Finance sector and computer-based trading just one example of complex adaptive systems (automotive, air traffic, medical…)

- Challenges come from the complex adaptive nature of these systems and their risks and benefits
  - do we need a fundamental change to risk assessments, assess “small” changes, stability and cliff edge identification

- Opportunities from deploying and adapting our current engineering approaches to systemic risks

- Work in progress….

  - Hope to publish public draft in May 2012 “Meltdown in the markets: a nuclear perspective”
Adelard Centre for Software Reliability

- Safety and assurance cases and safety management systems
- Independent safety assessment
- Software assurance, including formal methods and static analysis
- Development, interpretation and application of standards and guidelines
- Applied research in safety, security, critical infrastructure interdependencies
- Policy to technology
- ASCE – the Assurance and Safety Case Environment
- Clients in nuclear, defence, financial, transport sectors
- Evaluation of socio-technical systems
- Technical, interdisciplinary
- Research
- With international community and users
- Education
- Placements, internships, scholarships, courses, MSc and CPD
- Innovation
- Director, Dr Peter Popov
- DivSQL, PIA-FARA