Dealing with the Uncertainty of Big Data Security

Reg Harbeck,
Mainframe Analytics
Services

Brian Cummings
Tata Consultancy

Friday, Aug. 8, 2014: 10:00 AM
Session 15667

SHARE in Pittsburgh 2014
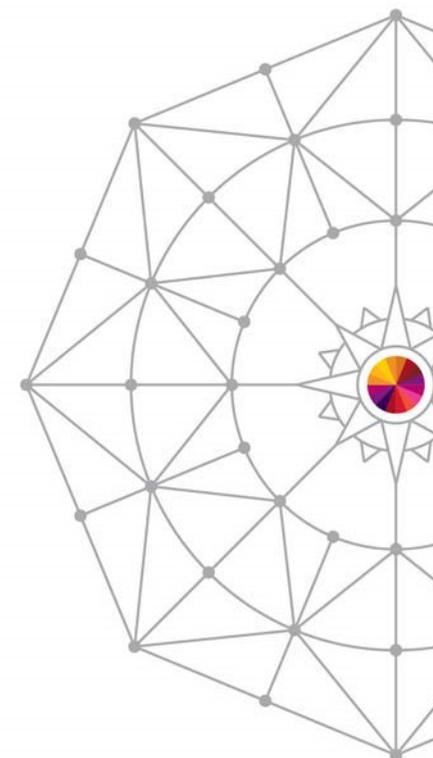
Who We Are

Brian Cummings

- Currently working at Tata Consultancy Services
- Mainframe security expert with 30+ years' experience
- Information Security & Risk Strategist & Program Architect
- GRC and Enterprise Risk Management Evangelist
- Advocate for Risk-Balanced Solutions & Controls
- Focus on RACF, Architecture, GRC, Big Data, Compliance, Privacy

Who We Are

Reg Harbeck

- Currently Chief Strategist at Mainframe Analytics ltd.
- 25+ years in systems, networks, applications, security on mainframe and other platforms
- Expertise in all three z/OS security packages, regulatory compliance

Plan of Attack

- Context and Concepts

- Unmeasurability, incompleteness, uncertainty and other challenges in securing big data

- The Hitch Hiker's Guide to an ongoing response to these issues

Context and Concepts

- Big Data
  - Some relational
  - Some unstructured
  - Some multimedia
  - Varying and increasing confidentiality, sensitivity and validity
  - Few hard edges
  - Consequences of exposure unpredictably severe
    - Mapping regulations and laws non-trivial

The Big Data "Value Chain"

Risk increases as your Big Data gains value

Increasing escalation in data value and corresponding risk

Decisions and Action

Permission of Creative Commons

Reporting and Human Analytics

Data Analytics

Data Aggregation and Organization

Data, Content Input

Security Risks

Confidentiality: Your work can have value to others.

Integrity: Big data integrity is key to appropriate decision making

Availability: Big data project criticality increases as you approach decision making. Timely decisions could be impacted by an untimely interruption.

Baseline Value & Risk

Standard and relatively constant data, privacy, and segregation of duties risk and controls

Big Data Value Chain

6

Context and Concepts

- Heisenberg's Uncertainty Principle
  - Can't measure without changing beyond a certain granularity
  - Relevance to big data:
    - More and more expensive to properly characterize nature and criticality of data
    - More awareness = more responsibility
    - Less awareness = greater exposure
    - Cannot have perfect awareness and control, but regulations and headlines assume it

$$\Delta p \, \Delta x \geq \frac{1}{2} \, \hbar$$

$$\Delta E \, \Delta t \geq \frac{1}{2} \, \hbar$$

P=momentum
X=position

Big Data environments will be highly dynamic…taking in new data, aggregating data, analyzing data, producing predictions and actionable insights. The "C-I-A" of various data elements will be difficult to define and evaluate.

Context and Concepts

- Gödel's Incompleteness Theorem
  - Any sufficiently complex system of expression is capable of containing indeterminably true/false assertions, e.g., "This sentence is false."
  - Implications for big data:
    - Metadata about the criticality and sensitivity of dat[a] becomes a security issue – "this data is not sensitive" can be a sensitive fact
    - Processing requirements for in-depth granularity become a resource issue, but also a hint about the security of a given piece of data.

As you move through the big data value chain, data at one point might be of little value, but may have higher value a step further taken in the context of aggregation or analytics. The challenge is to pinpoint when the data reaches a value (completeness) that would be deemed sensitive and subject to increased security.

A PROBLEM WITH LANGUAGE AND LOGIC

NO # languages capable of making the statement "A" can be internally self-consistent.
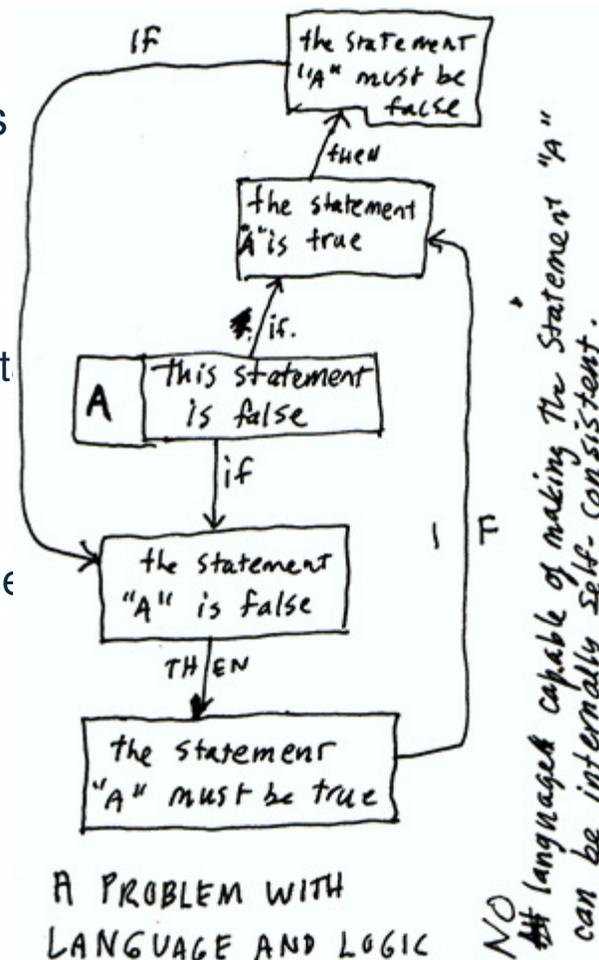
Context and Concepts

- The Halting Problem
  - The only way to know whether an arbitrary simple program ever finishes running is to run it. In running it, if it doesn't stop, you don't know if it ever will (halt)
  - Implications for big data:
    - Can't generically predict for an arbitrary block of data how sensitive it is or how much processing it will take to properly secure it
    - Security becomes a journey without a knowable end
    - Can't accept unlimited processing time to determine how to best secure a given mass of data
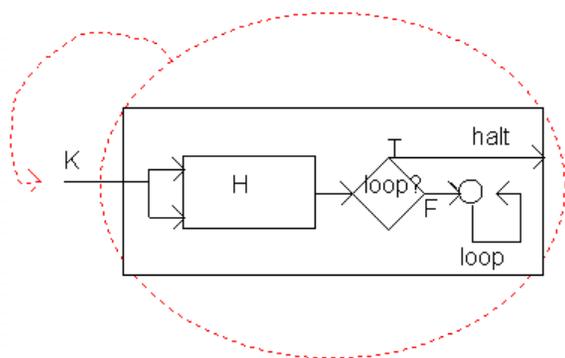
Will there be a steady-state in Big Data environments…a point at which data stops changing? Likely not, and you can't wait to see if it ever reaches that state. Don't let the pursuit of perfection get in the way of doing something.

Context and Concepts

- Schrödinger's cat
  - Mathematically, it's 50% dead until you look inside the box, i.e., both dead and alive
  - Implications for big data:
    - Without measuring/examining a given mass of data, you can only assign probabilities of its sensitivity and the effort to secure it
    - Once the effort and time are taken to measure/determine its actual sensitivity, the odds are no longer relevant for the specific, but improve generic characterization in the future
    - If you don't measure a specific mass of data, hackers and headlines might

We face a conundrum in the transmutation of data as it moves through the Big Data value chain. How to secure something as it continually evolves. Is it or isn't it sensitive?

THE CAT IN THE BOX BY Dr. Schrodinger
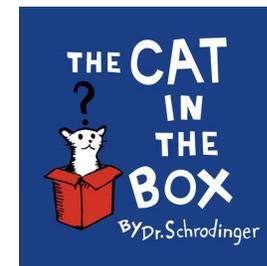
A Hitch Hiker's Guide

- Don't Panic – it won't help
- SHARE your journey
- "Write up" problem – when in doubt secure more
- Once Schrödinger's cat's out of the bag, it's too late
- Security after the fact is a sitting duck
- Security before the fact is somewhere between denial of service/access and a discouragement of future involvement
- Current motivational thinking about Quantum Physics suggests to predict an outcome, choose it

A Hitch Hiker's Guide

- Discussion re Heisenberg:
  - What are your experiences with uncertainty about security and measurement of data sensitivity?
  - Practical approaches to balancing generic/specific in characterizing and securing data?

Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

A Hitch Hiker's Guide

- Discussion re Gödel's Incompleteness Theorem
  - What are your experiences with mobius loops between data, metadata, and sensitivity?
  - Practical approaches to cutting such Gordian knots?
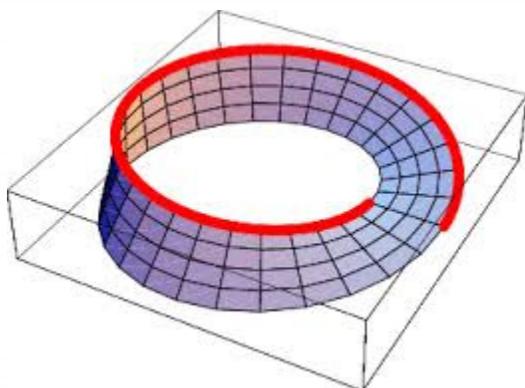
A Hitch Hiker's Guide

- Discussion re Halting Problem
  - What are your experiences with predicting complex sensitivity of data?
  - Practical approaches to quickly determining sensitivity of arbitrary masses of big data?

Social media network connections among Twitter users
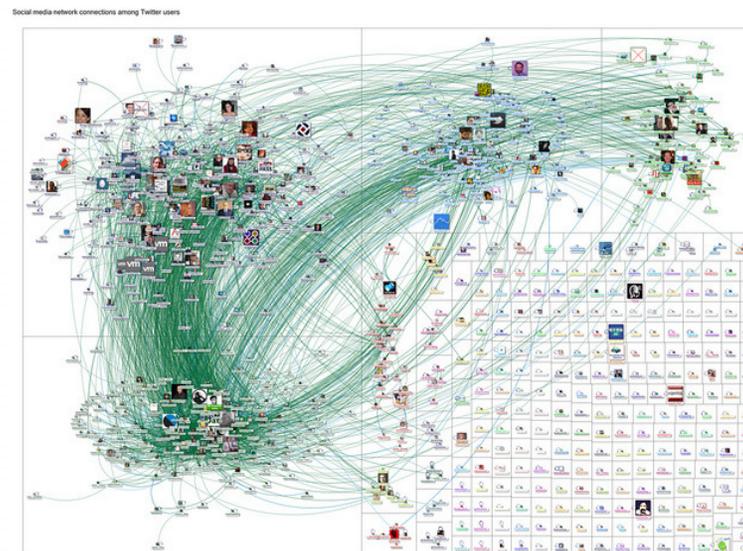
Created with NodeXL (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

Forbes 9/24/2012

A Hitch Hiker's Guide

- Discussion re Schrödinger's cat
  - What are your experiences with characterizing vs measuring the sensitivity of data?
  - Practical approaches to balancing effort to secure arbitrary data with high need to know and high sensitivity without getting stuck in a fractal vortex
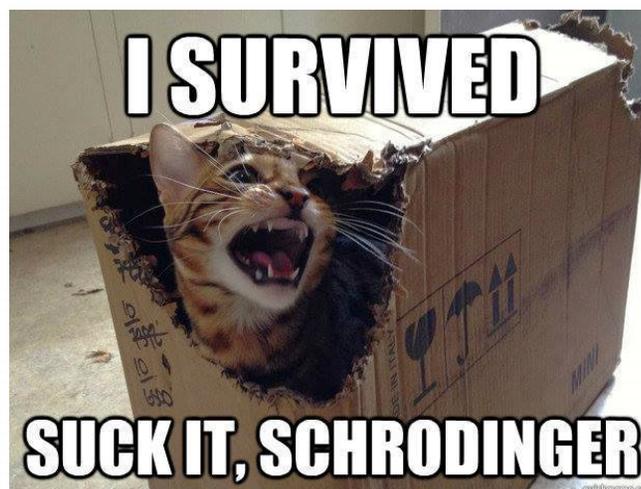
Conclusions and Take-Aways

- Security is a journey, not a destination
- Important to SHARE your insights
- Balance need-to-know, sensitivity, resource-intensivity, timeliness, and potential regulatory, headline and hacker exposure
- Attitude, action, attention, adjustment… repeat

With the uncertainties presented by Big Data environments, can a case be made to tightly restrict those environments? The value of analytical results, predictions, and actionable insights will be quite high. How is aggregated data validated? How is the integrity of analytical results validated? Would it make sense to publish these analytical results and recommendations to a DRM solution? We have much to learn, but we know we need to be concerned and protective.

Details? See Brian's "Big Data – Big Security Risk" SHARE presentation.

Thank You!

Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval