Using Social Engineering Tactics For Big Data Espionage

Branden R. Williams

Jason Rader

RSA

Session ID: DAS-301

Session Classification: General Interest

What Is This Session About?

- Big Data Analytics
  - They are the future!
  - How do we deal with them in the security world?
- Adopting techniques from Social Engineering
  - How can we learn from other security issues?
  - What snakes may be lurking in the grass?
- Application
  - What to do?

Predictive Analytics - The Future of BI

- Big Data
  - Not LOTS of data
  - But many disparate sources with mixed structure
- Industry trends driving Big Data & analytics
  - Storage is cheap
  - “Anywhere” computing
  - Hardware over-engineering
  - (or distributed computing)
- Math is pretty awesome

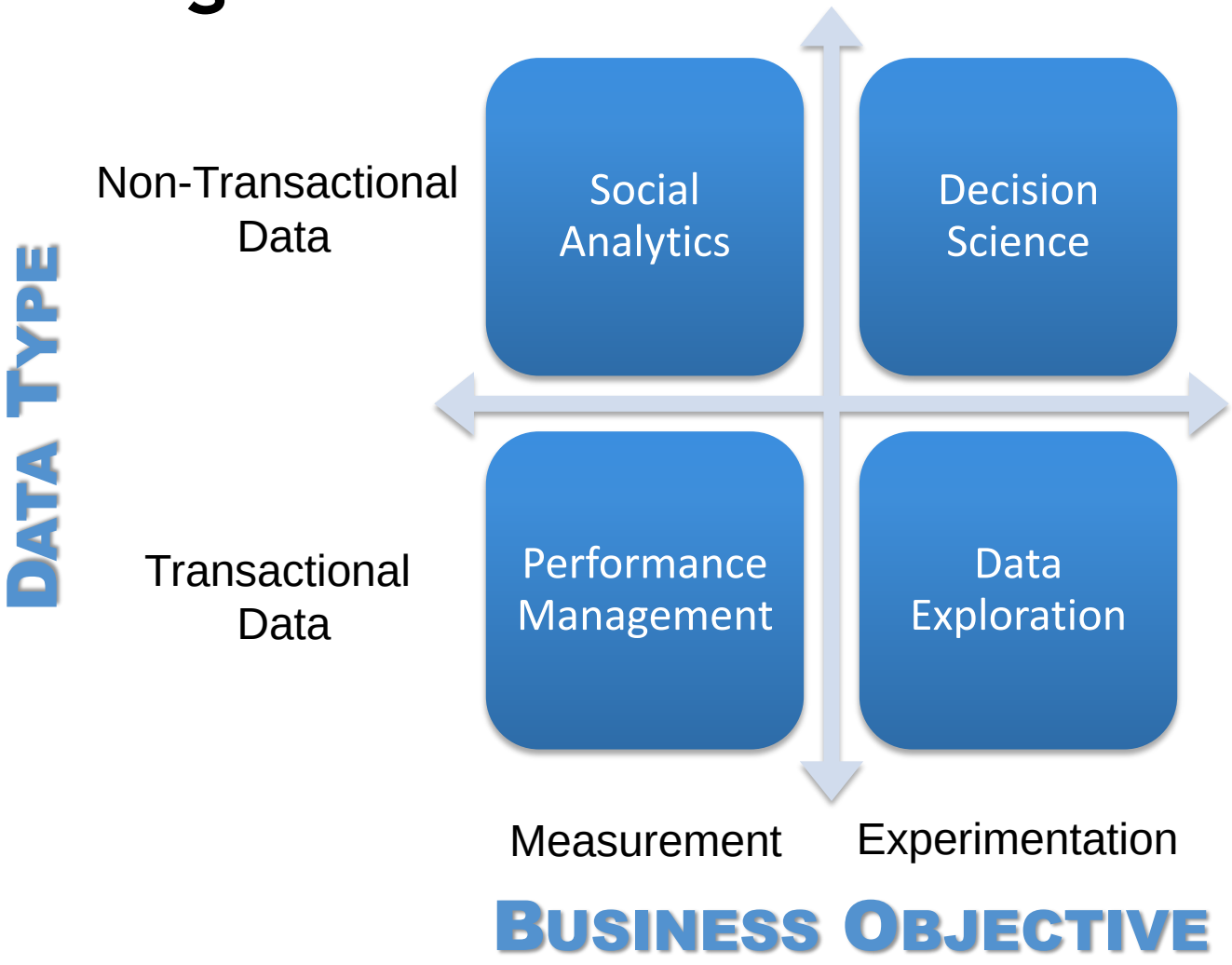
What are Predictive Analytics?

- Simply, deriving value from big data!
- Complexly (is that a word?)
  - Organizing the unorganized
  - Heavy math use
  - The search for leading indicators
- How do we use it?
  - Business strategy
  - Public safety (CDC,NOAA)
  - Predict the future

Who does this well today? Investors!

- Massive historical data
- Lots of math
  - Sometimes bad math
  - The search for trends
  - Execute trades!
- The next step?
  - Implications of social media
  - See through the corporate veil
  - Now we can focus on the individual!

A Big Data Framework



Predictive Analytics - The Why

- Businesses have data coming out of their ears
- SO DO THE BAD GUYS!
- Huge shift from the past
  - Forensics example
- Unused data
- Can we find value?
- Business leaders must innovate[1]
  - But focus on business model innovation
  - PA can drive this!

[1] Amit, R., & Zott, C. (2012). Creating value through business model innovation. *MIT Sloan Management Review, 53*(3), 41-49.

Derived Data & Analytics (more sensitive than original data!)

- Analytics automate “data crunching”
- Could become the new IP!
- Businesses act on output
- What about deriving sensitive info?
  - PII
  - Privacy issues!
- If I accurately guess your info, do I need to protect it?

Problem of derived sensitive data

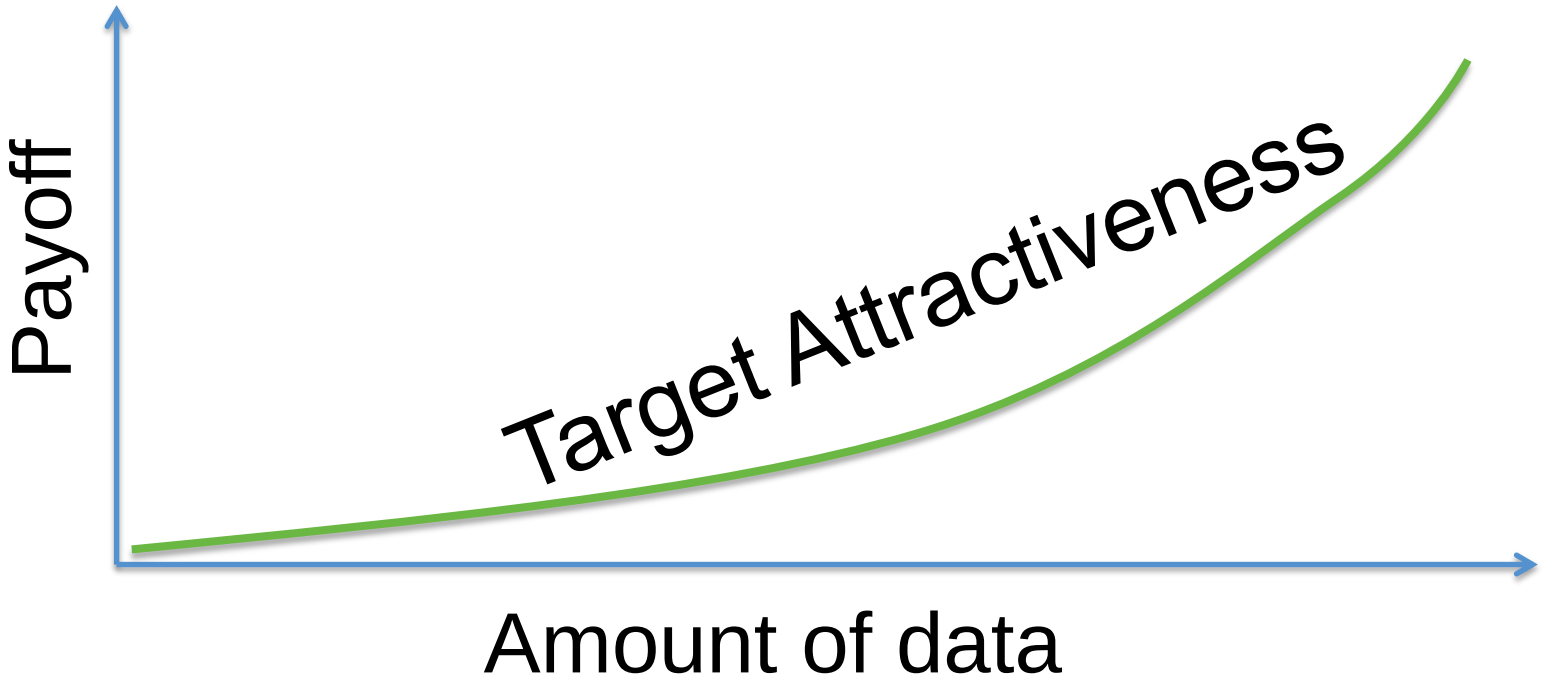
- If I derive your PII, must I protect it?
- Law is WAY behind
- Compliance might be farther
- HUGE IMPACT TO BUSINESS
  - In EU, no question
  - How to apply EU Privacy Updates?

Practical Applications

- How Target knew a teenager was pregnant before her own father did:
  - Pause for story time!

- Implications:
  - Target can predict due dates/sex
  - Aggregations of data can accurately predict the future
  - Better placement (upside?)
- Not Singular pieces of data!
  - It's the CORRELATION that matters

The Relevance of Data Mass



The Compartmentalization problem

- Big data requires big flexibility
  - But that impacts security
  - Think about SQL features:
    - Views
    - Complex Queries (with aggregates)
    - Column-level RBAC
- Classic approaches (where have we solved this problem before)
  - DBAs still a risk (so is the Data Scientist)
  - Operators may not know what they are looking at

Attack Scenarios for BI

- Hacking the Coca Cola formula
  - Can you derive the formula?
  - YES YOU CAN!
  - Shipping/Receiving/Production data
    - Think Crypto plaintext attack
  - KFC would be harder, but possible!
- IP/Strategy Derivation
  - Attackers can figure out your IP by looking at data sets you don't necessarily protect
  - Macro trends repeat themselves over multiple sets of data, can I fill the gaps from missing data sets?

How Does This Fit Into Social Engineering?

- Social engineers rarely blatantly ask for all of the information
  - Patchwork over time
  - Desired outcomes are known
  - Surprises happen to change outcome
- Given 20 sources of input data:
  - If a criminal has access to say 15
  - Can he replicate analytics & results?
  - Can he fill in the gaps?
- Eeek.

Side By Side Comparison

Social Engineering

- Gather Intelligence
- Go after willing targets (sheeple)
- Gather innocuous information in bits
- Piece together bits
- Exploit

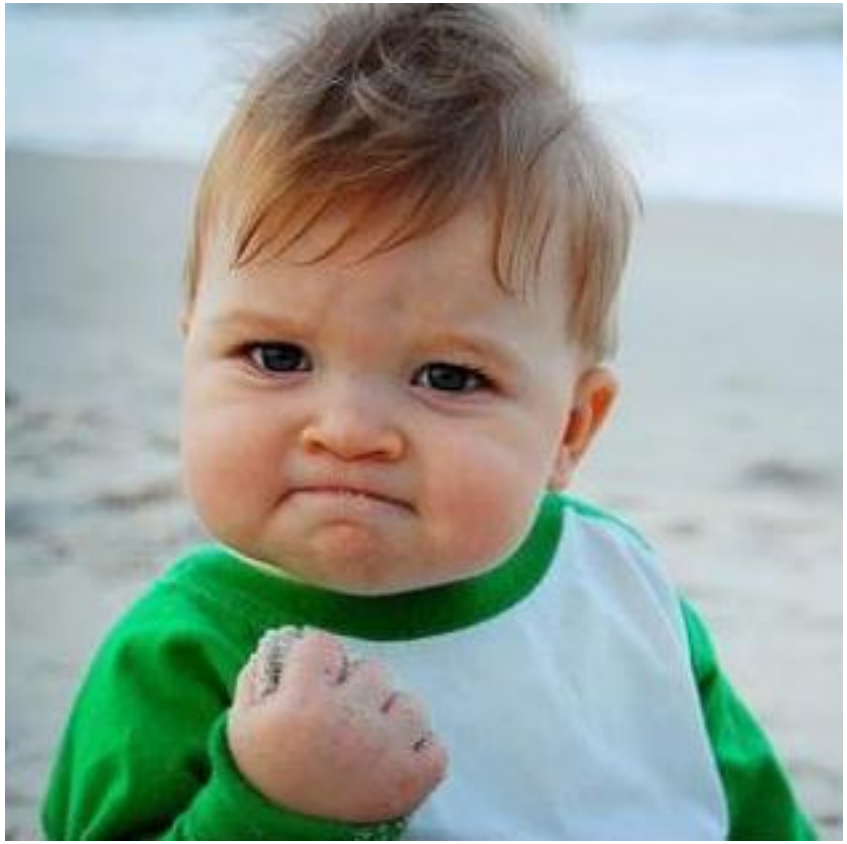
Big Data Attacks

- Source Data
- Go after IP rich companies
- Go after under-protected information
- Find and fill gaps
- Exploit

How To Apply Content From This Session

- In the first three months:
  - Understand the sausage factory operation
    - How does your business use analytics/data to operate or make strategic decisions?
    - What happens in real-time?
  - Catalogue data sources and uses
    - Where are those inputs and how are they protected?
- Within six months you should:
  - Identify data where inference leads to espionage
    - Unused business data or seemingly mundane data
  - Build sourcing plans (in or out!)

That’s all folks!



Q/A

@BrandenWilliams
@JRaderRSA

Thank you!