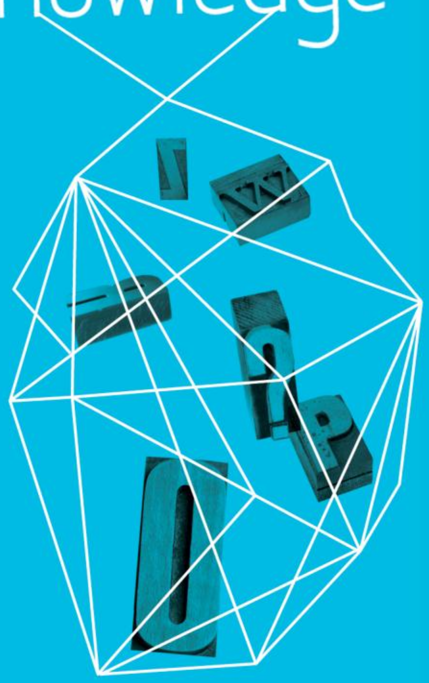
RSACONFERENCE2013

Security in knowledge

THIN SLICING A BLACK SWAN: A SEARCH FOR THE UNKNOWNS



Michele Chubirka

Transaction Network Services/Packetpushers.net

Session ID: MASH-F41A

Session Classification: Intermediate

Something's Broken

In Verizon's 2012 Data Breach Investigations Report, it was found that across organizations, an external party discovers 92% of breaches.

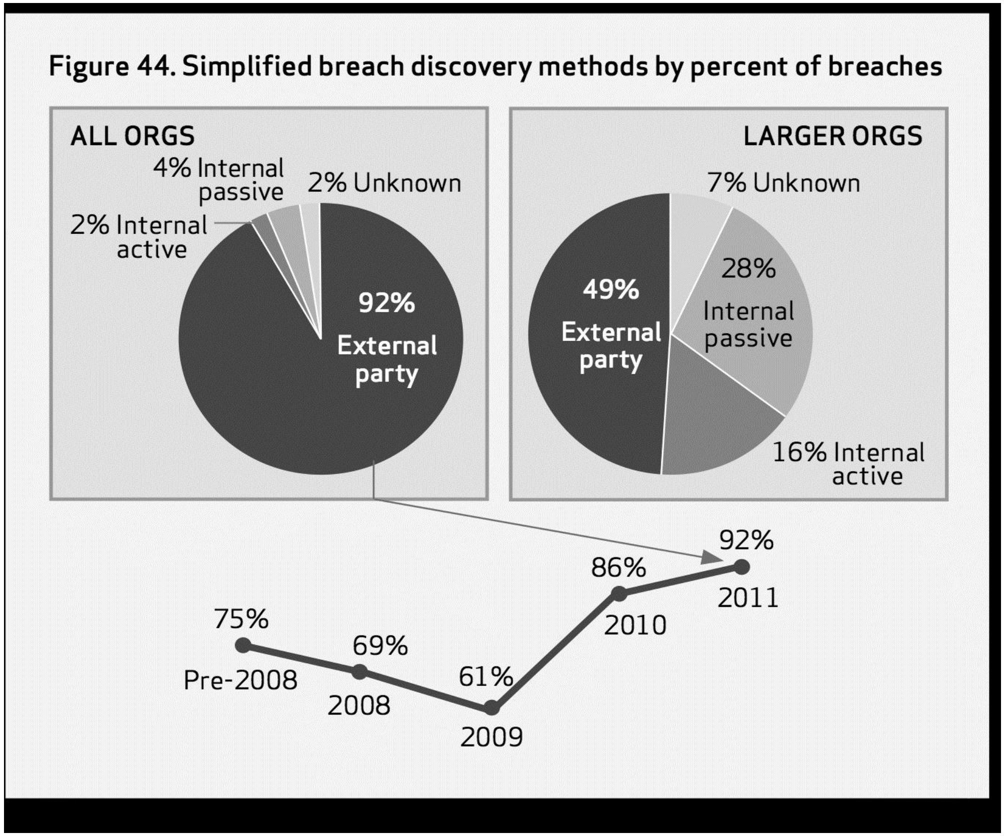


Figure 44. Simplified breach discovery methods by percent of breaches

From Compromise To Discovery

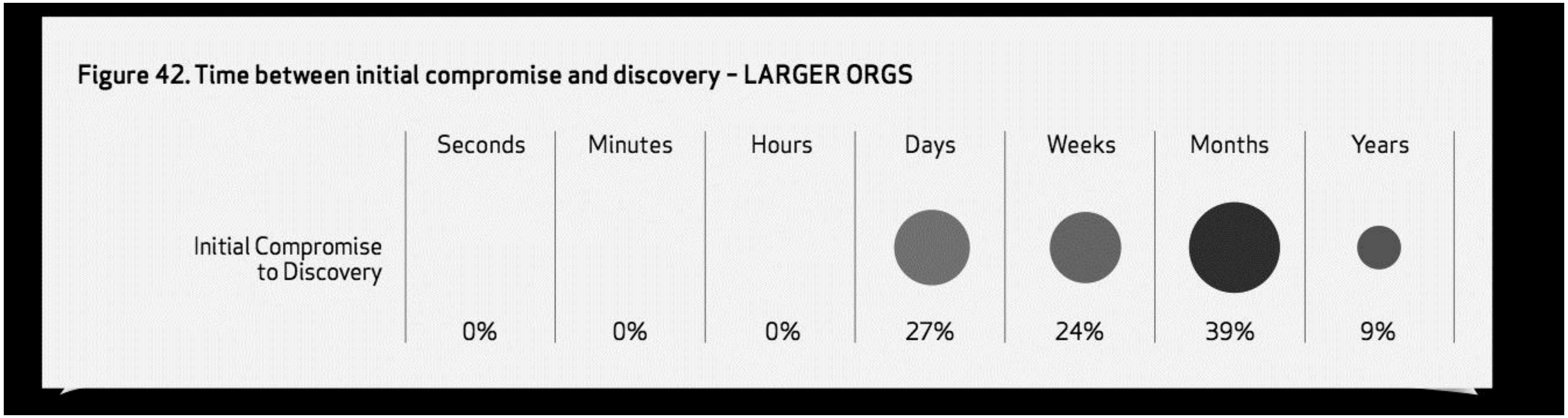


Figure 42. Time between initial compromise and discovery – LARGER ORGS

- We believe we can solve the issue of the *unknowns*, intrusions, with more data.
- The more information we have, the less we know.
- This makes us no better than security archeologists.

The Black Swan Event

- An *unknown unknown.*
- Can’t be predicted by probability theories.
- Rationalized after the fact.
- How often do we try to predict the Black Swan Event in security and fail?

Information Gluttony?

“Military drone operators amass untold amounts of data that never is fully analyzed because it is simply too much.”

Michael W. Isherwood, defense analyst and former Air Force fighter pilot.

Digital Kudzu

- From beginning of recorded time to 2003 - five exabytes of information.
- 2011 - that much created every two days.
- 2012 - prediction is every 10 minutes.

Current Solutions

- SIEMs: never gets fully implemented.
- Predictions using Logistic Regression/Bayesian Probability.
- Huge amounts of data, not enough time.
- “Open world” problem using “closed world” assumptions.
- More staff, more money.

Alternative Model: Thin Slicing

“…the ability of our unconscious to find patterns in situations and behavior based on very narrow slices of experience.”

Malcolm Gladwell, ***Blink***

Case Study: A Hospital in Trouble

- Cook County Hospital struggled with identifying patients in danger of an imminent heart attack.
- Coronary care unit was overwhelmed.
- Public hospital, limited resources.

Applied Thin-Slicing

- Lee Goldman, a cardiologist, created a protocol based upon an algorithm developed in partnership with mathematicians.
- After two years of using a decision tree, hospital staff were 70% more effective at recognizing patients at risk.
- **Less** information led to greater success.
- Technique used by first-responders every day.

Fast and Frugal Trees

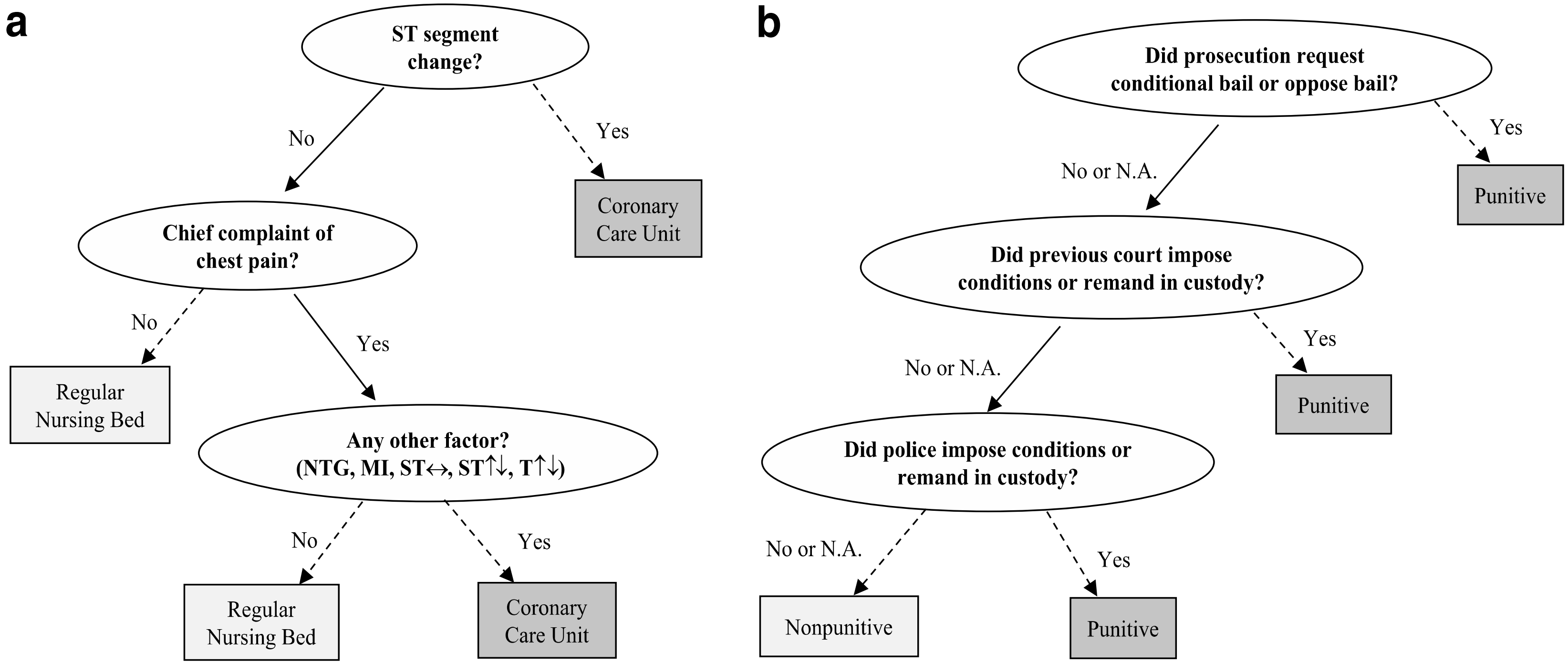
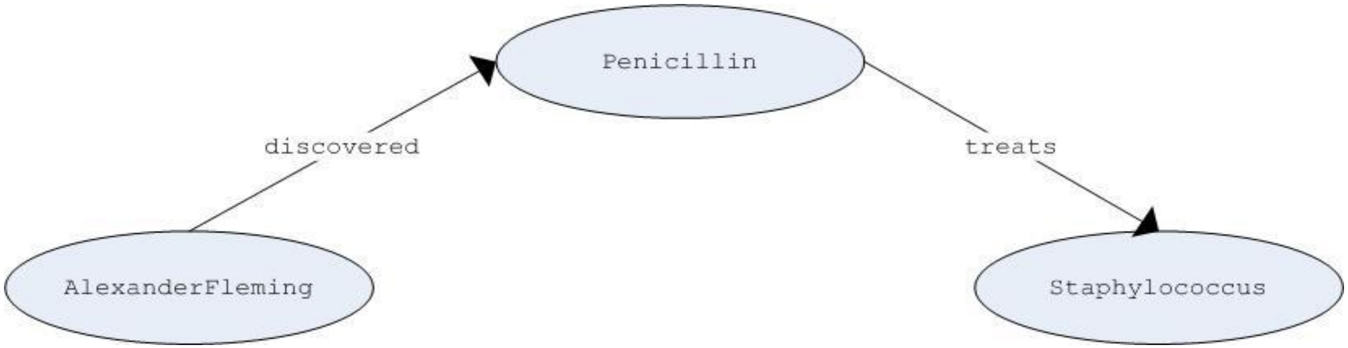

a
ST segment change?
No
Yes
Coronary Care Unit
Chief complaint of chest pain?
No
Yes
Regular Nursing Bed
Any other factor? (NTG, MI, ST↔, ST↑↓, T↑↓)
No
Yes
Regular Nursing Bed
Coronary Care Unit
b
Did prosecution request conditional bail or oppose bail?
Yes
Punitive
No or N.A.
Did previous court impose conditions or remand in custody?
Yes
Punitive
No or N.A.
Did police impose conditions or remand in custody?
No or N.A.
Yes
Nonpunitive
Punitive

Figure 4. Two examples of fast-and-frugal trees (FFTs) applied to large world problems. The left tree (a) is designed to help emergency room doctors decide whether to send a patient with severe chest pain to the Coronary Care Unit (CCU) or a regular nursing bed (Green & Mehr, 1997). The right tree (b) is a model of how British judges decide whether to make a punitive bail decision (Dhami, 2003).

Method: Resource Description Framework (RDF)



- Semantic Web technology.
- Queries based on relationships or mental associations.
- Graphs treat each packet from capture file as a discrete event with properties.
- TCP header info in a metadata model.
- Model replicates human cognitive economy.

Thin-Slicing with SPARQL

- SPARQL query language uses a concise approach for quickly traversing large data sets while capturing similarities between packets as generalizations.
- RDF statement contains a subject, predicate and an object.
  - Subject defines the event.
  - Predicate defines a characteristic or property.
  - Object contains the value for the predicate.

Example: Building A Query

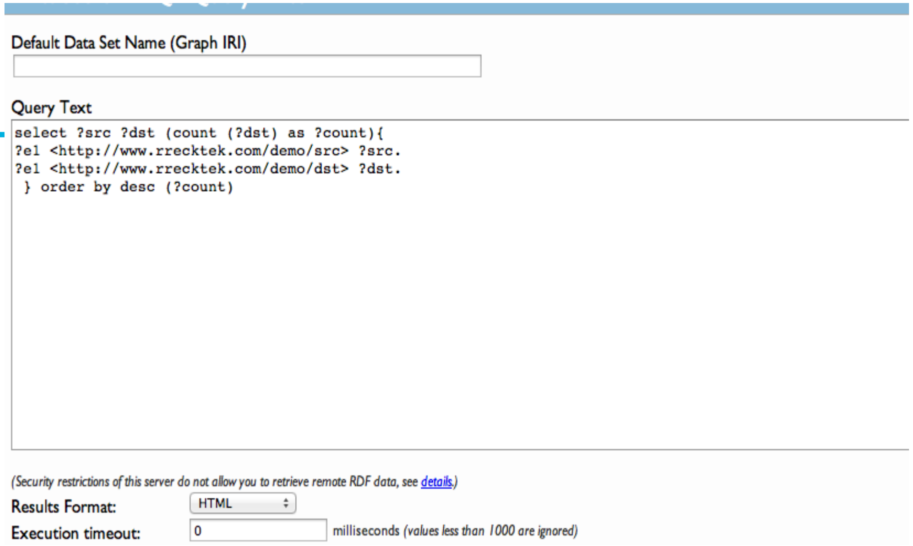
```
sparql select * {
?s
?p
?o.};

sparql select *{
?e1
<http://www.rrecktek.com/demo/src>
?ip1.};
```

Example

- All source IPs and their destination IPs.
- For each source, count how many times it went to a destination.
- Report source destination and count.

```
sparql SELECT ?src ?dst (count (?dst) as ?count) {
?e1 <http://www.rrecktek.com/demo/src> ?src.
?e1 <http://www.rrecktek.com/demo/dst> ?dst.
 } ORDER BY DESC (?count);
```

SPARQL web interface



Default Data Set Name (Graph IRI)

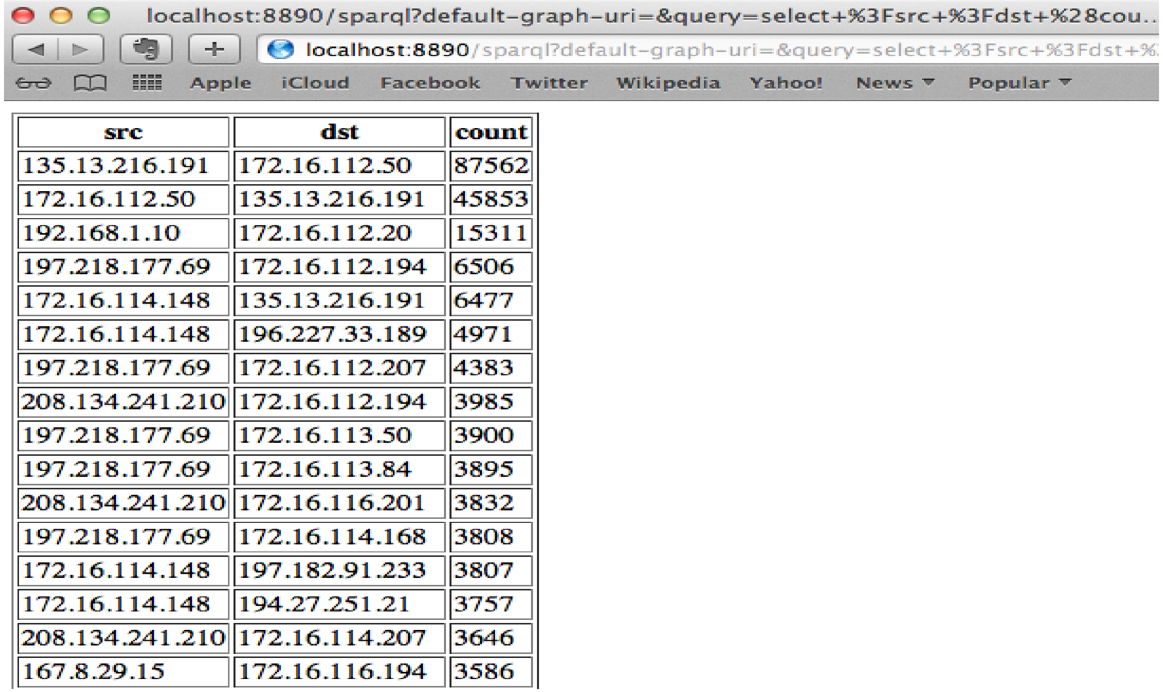
Query Text

```
select ?src ?dst (count (?dst) as ?count){
?el <http://www.rrecktek.com/demo/src> ?src.
?el <http://www.rrecktek.com/demo/dst> ?dst.
 } order by desc (?count)
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see details.)

Results Format: HTML

Execution timeout: 0 milliseconds *(values less than 1000 are ignored)*



localhost:8890/sparql?default-graph-uri=&query=select+%3Fsrc+%3Fdst+%28cou..

localhost:8890/sparql?default-graph-uri=&query=select+%3Fsrc+%3Fdst+%

src	dst	count
135.13.216.191	172.16.112.50	87562
172.16.112.50	135.13.216.191	45853
192.168.1.10	172.16.112.20	15311
197.218.177.69	172.16.112.194	6506
172.16.114.148	135.13.216.191	6477
172.16.114.148	196.227.33.189	4971
197.218.177.69	172.16.112.207	4383
208.134.241.210	172.16.112.194	3985
197.218.177.69	172.16.113.50	3900
197.218.177.69	172.16.113.84	3895
208.134.241.210	172.16.116.201	3832
197.218.177.69	172.16.114.168	3808
172.16.114.148	197.182.91.233	3807
172.16.114.148	194.27.251.21	3757
208.134.241.210	172.16.114.207	3646
167.8.29.15	172.16.116.194	3586

We Can’t Fight All Unknowns

- What we *can* do
  - Build strong infrastructures minimizing technical debt.
  - Add the equivalent of air bags to the architecture for when intrusions occur.
  - Recognize signature limitations.
  - Investigate the creation of real-time fast and frugal trees.

Our patient is dying on the table. It’s up to us to change the outcome.

Thanks!

- Michele Chubirka

  Twitter @MrsYisWhy
  networksecurityprincess@gmail.com
- RDF/SPARQL contribution courtesy of Ronald P. Reck

  rreck@rrecktek.com

References

"Eclectic Tech." *Semantic Web Introduction*. N.p., n.d. Web. 20 Dec. 2012.

Erwin, Sandra I. "Too Much Information, Not Enough Intelligence." *National Defense Magazine*. N.p., May 2012. Web. <http://www.nationaldefense.org>.

Gigerenzer, Gerd. *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking, 2007. Print.

Gladwell, Malcolm. *Blink: The Power of Thinking without Thinking*. New York: Little, Brown and, 2005. Print.

Luan, Shenghua, Lael J. Schooler, and Gerd Gigerenzer. "A Signal-detection Analysis of Fast-and-frugal Trees." *Psychological Review* 118.2 (2011): 316-38. Print.

Marewski, Julian N., PhD, and Gerd Gigerenzer, PhD. "Heuristic Decision Making in Medicine." *Dialogues in Clinical Neuroscience* 14.1 (2012): 77-89. Print.

Messmer, Ellen. "SANS Warns IT Groups Fail to Focus on Logs for Security Clues." *TechWorld*. IDG, May 2012. Web.

"RDF." *-Semantic Web Standards*. W3C, n.d. Web. 02 Jan. 2013.

"Resource Description Framework (RDF)Model and Syntax." *RDF Model and Syntax*. W3C, n.d. Web. 02 Jan. 2013.

Rieland, Randy. "Big Data or Too Much Information?" *Innovations*. Smithsonian, 7 May 2012. Web.

"Semantic Web Standards." *W3C*. W3C, n.d. Web. 02 Jan. 2013.

Taleb, Nassim. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House, 2007. Print.

Turek, Dave. "The Case Against Digital Sprawl." *The Management Blog*. Bloomberg Businessweek, 2 May 2012. Web.

Verizon 2012 Data Breach Investigation Report. Rep. N.p.: Verizon, n.d. Print.