RSA Conference 2016

San Francisco | February 29 – March 4 | Moscone Center

Connect to Protect

SESSION ID: AIR-T09

Demystifying Security Analytics: Data, Methods, Use Cases

Dr. Anton Chuvakin
@anton_chuvakin
Research VP
Gartner for Technical Professionals

#RSAC

Easy, Huh?

Security analytics is very easy:

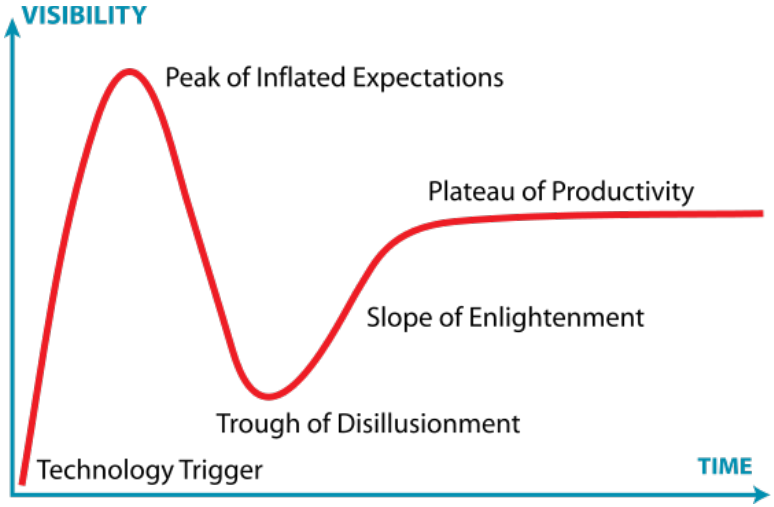
1. Get data.
2. Process with algorithms.
3. Enjoy the insights!

Outline

1. Defining “SECURITY ANALYTICS”
2. Choosing your road to security analytics success
3. Tooling up for security analytics: myth and reality
4. Best practices for security analytics?

A Little Motivation: Why Security Analytics?

- Detect threat better — and detect "better" threats!
- Decide what matters, prioritize alerts and signals.
- Triage alerts faster.
- Solve (some) security problems with less expert labor.



Beware of overhyped tools!

Gartner

Example: Password Guessing

- Manual way
  - How many attempts constitute malicious password guessing?
    - 3? 5? 10? 100?
    - Within 1 second? 1 minute? 10 minutes?
  - How to write a rule here?
- Data-driven / analytics way
  - Run ML on your data, learn from past malicious cases, update models
  - WIN!

What Is Security Analytics?

Definition Drill-Down

Security analytics = *Some* advanced analysis of *some* data to achieve some useful security outcome!

- Q: What is advanced?
- A: Anything better than "known good/known bad" rule matching and basic statistics (like "if 10% above average, alert"). Not only machine learning!

- Q: What data, specifically?
- A: Really, any data: From logs to flows/sessions, from transactions to various context on users, such as HR data, etc.

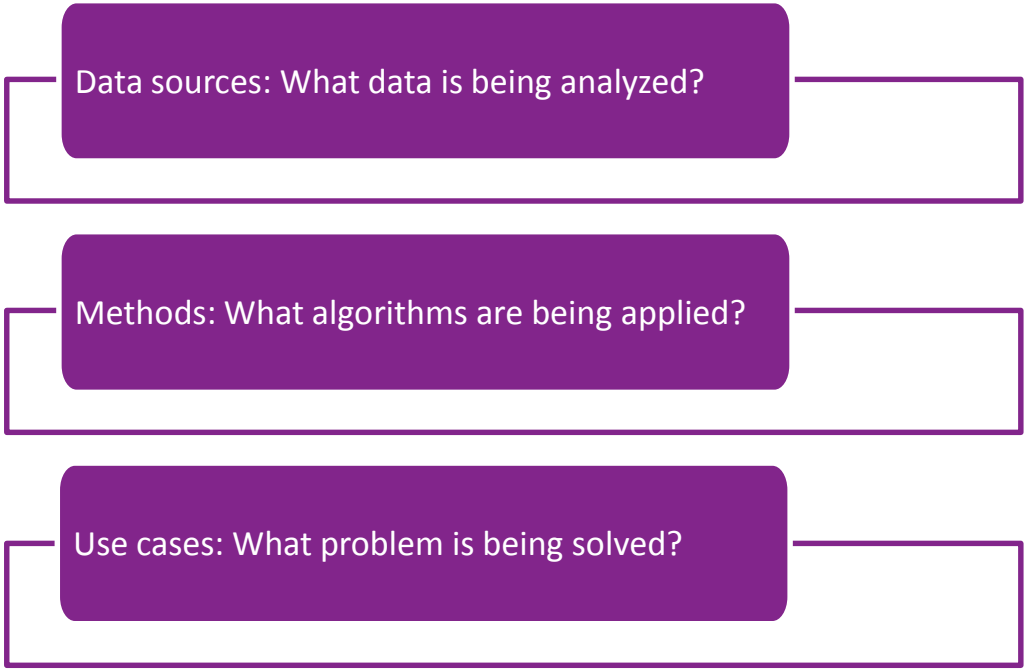
Gartner

Still Mystified?

Analytics Demystifying Framework!

Data/Methods/Use cases:

- What **data** is being analyzed?
- What **methods** and algorithms are being used?
- What specific **problems** are being solved?



Gartner

Example: DLP Alert Prioritization

Data/Methods/Use cases:

- *What **data** is being analyzed?* - DLP alerts, user identity context, data access logs
- *What **methods** and algorithms are being used?* – PCA to find out which dimensions of the alert matter and how much
- *What specific **problems** are being solved?* – Identifying which alerts indicate significant and dangerous data theft

Starting Your Security Analytics Journey

Gartner

RSAConference2016

Security Analytics Key Success Factors

1. Analytic, data-driven mindset (!)
2. Willingness to explore data
3. Ability to collect and retain data (SIEM helps — isn't a must)
4. Some understanding of data science approaches
5. Availability of people to use the tools, prepare the data and refine the questions

Really, A Mindset??!

- Are you a data explorer or an appliance buyer ("OOBster")?
- Would you look at your data or at your vendor for answers?
- Must you have out-of-the-box content (rules, signatures, etc.) for everything?
- Do you accept that data may have the answers, not the gut feel?

Analytic mindset seems to determine the success of analytics and big data initiatives for security more than anything else!

Products or People: Buy, Build, Partner

Gartner

RSAConference2016

Before You Buy!

Does it say
BUY THE WRONG PRODUCT
anywhere?

Figure 2. A Framework for Preventing Big Data Failures

Source: Gartner (August 2015)

Gartner

Buy Choice

Pros:

- Solve at least some problems immediately upon installation
- Solve select problems automatically, without tuning

Cons:

- Risk of limited applicability of the tool beyond specific use cases
- Rely on vendor to pick and solve problems

Build Choice

Pros:

- Leads to development of the capability that can be used to solve future problems
- Focus resources on organization-specific problem

Cons:

- Extensive effort to build and then mature an analytic capability
- Skills requirements uncommon for IT organizations (statistics, etc.)

How to Avoid "Science Project" Syndrome?

- Start with DATA
- Explore data to find and solve PROBLEMS
- Grow SKILLS to identify other useful problems and solve them
- RETAIN KNOWLEDGE and grow analytics capability

- Start with PROBLEM AT HAND
- Find the right DATA to solve it
- Try different ALGORITHMS, tuning the data too
- LEARN and solve next problem better
- LOOK at other data and problems

Gartner

Partner Choice

Pros:

- Leads to development of the customized capability
- Leverages vendor/provider expertise gained from previous projects

Cons:

- Longer time to value compared to off-the-shelf products
- High cost of specialty consulting labor

Warning: Buy Is Often "Buy Then Build"?

Buy advantages:

- Solve at least some problems within 30 days after deployment
- Solve select problems automatically, without much tuning
- Supported commercial product may be more comforting to security leaders
- Rely on vendor expertise with algorithms and statistics; vendor can employ scientists and statisticians

However … even if you decide to buy, you may still need to build and definitely tune (both initially and over time!)

Tooling Up

Tools: The Mythology Edition

- There is no one **"security analytics market."**
- There is no specific **"security analytics technology."**
  - (and no "big data security analytics technology").

- Security analytics is a concept!
- Several types of *very* different commercial tools use it.
- You can also build your own tools — and use OSS heavily:
  - Ever heard of Hadoop? ELK stack? R?

Tools: The Reality Edition

1. User behavior analytics (UBA, sometimes UEBA for “entity”)
2. Network traffic analysis (NTA, not a common name yet)

… and also:

- Broad-scope data analysis tools with **solid** and **proven** security use cases

Gartner®

But Wait … Where Is SIEM?

Requirement	Likely tool type to use
Collect log data for compliance, run reports	Log management
Perform near-real-time security correlation	SIEM
Detect user anomalies and solve other typical data-intensive problems	UBA or other commercial tool
Solve organization-specific data intensive problems; collect and analyze diverse data types at high volume	Custom-built big data security platform

Source: Gartner (May 2015)

UBA or UEBA: What's Inside the Box?

	Details	Example
Data	Logs (from SIEM or directly), DLP alerts, flows, network metadata, IAM data, HR data, etc.	System authentication logs and Active Directory user information.
Methods	Supervised machine learning, unsupervised learning, statistical modeling, etc.	Self-to-self comparison, peer comparison and activity model vs. time.
Use cases	Compromised account detection, predeparture data theft, employee sabotage, shared account abuse, etc.	Detect account takeover by a malicious external attacker.

Example UBA Tool Wins!

- Compromised Accounts Found
- Departing Users Stealing IP
- Geolocation Anomaly
- Anomalous Behavior in VPN Activity
- Customer Service Rep. Privacy Breaches
- Source Code Compromised
- Compromised System Behavior
- Retired Devices Still in Service
- Unauthorized Access to Patient Records
- Privileged Accounts Shared

Gartner

Network Traffic Analysis: What's Inside the Box?

	Details	Example
Data	Flows data and network session metadata (to Layer 7), payload data; flows DNS and WHOIS data as context.	Application layer traffic data coupled with WHOIS domain registration data.
Methods	Supervised machine learning, clustering, network modeling, etc.	Network activity model over time, traffic volume model by protocol, etc.
Use cases	Data exfiltration, attacker lateral movement within the network and malware spreading.	Detect data exfiltration by the attacker based on traffic volume per protocol per use model.

One Implementation: How They Did It?

1. SIEM collectors feed SIEM and Hadoop (some direct data collection into Hadoop):
   - One data repository for everything ("data lake")
2. Selective data structuring (Hadoop to MongoDB and PostgreSQL):
   - Tableau fed from Postgres, custom tools fed from Mongo
   - Also, Solr runs off Hadoop
3. Shared data scientist pool (+ 1 "security data scientist")
4. Visualization and query tools built for junior analysts

Not Yet!

Recommendations

- ✓ Given an early stage of these technologies, **tool acquisition needs to be targeted at specific problems** because there are no "general security analytics tools" on sale. (BUY route)
- ✓ **There is not enough data on the comparative effectiveness of various analytic approaches** and algorithms (implemented in vendor tools) versus real-world problems. (BUY route)
- ✓ Think of the problems first, target purchases at problems, do validate that the vendor of choice has a record of solving such problems. (BUY route)
- ✓ Think of the data first, and start exploring, then bring tools and components as needed (BUILD route)

INDUSTRY CALL TO ACTION!

- ✓ Make the commercial tools EASY TO TEST and PROVE VALUE
- ✓ Evolve analytics tools to ALGORITHM PORTABILITY
- ✓ Create CATALOGUE of USE CASES where various analytics tools work well IN REAL WORLD
- ✓ Add more BRAINPOWER to all security tools!