Better Security Metrics

Hate on metrics all you want – they pay the bills.

$whoami

- Exec. Director of Threat Intelligence at SCYTHE
- IANS Faculty, former SANS Instructor
- Former NSA Hacker, endorsed by Shadow Brokers
  - aka Russian Intelligence
- Digital terrorist, breaker of software, responder of incidents, reverser of malware, injector of code, spaces > tabs
- **Dislikes:** those who call themselves “thought leaders,” “crypto bros,” and anyone who **needlessly adds blockchain** to a software solution

Agenda

- Why Metrics?
- Foundations of Metrics (That Don’t Suck)
- Example Blue Team Metrics
  - SOC Metrics
  - Incident Response Metrics
  - CTI Metrics
  - Threat Hunting Metrics
- Closing Thoughts

Ground Rules

- Photos are fine
- Posting online is fine
  - In case you were previously confused, **this is what consent means**
- I'll post slides later and this will be repeated in the coming months as a webcast
  - Follow my social media (@MalwareJake) for scheduling details
  - I'm sure it will be recorded then too, so if you want to see another talk

Why Metrics?

- Because stakeholders said so.
  - But why do they value (er, demand) metrics so much?
- What we do in security is inherently very technical
- We need to be able to communicate clearly to stakeholders:
  - What we do
  - How to measure our success
  - How to measure process growth

Foundation of Metrics

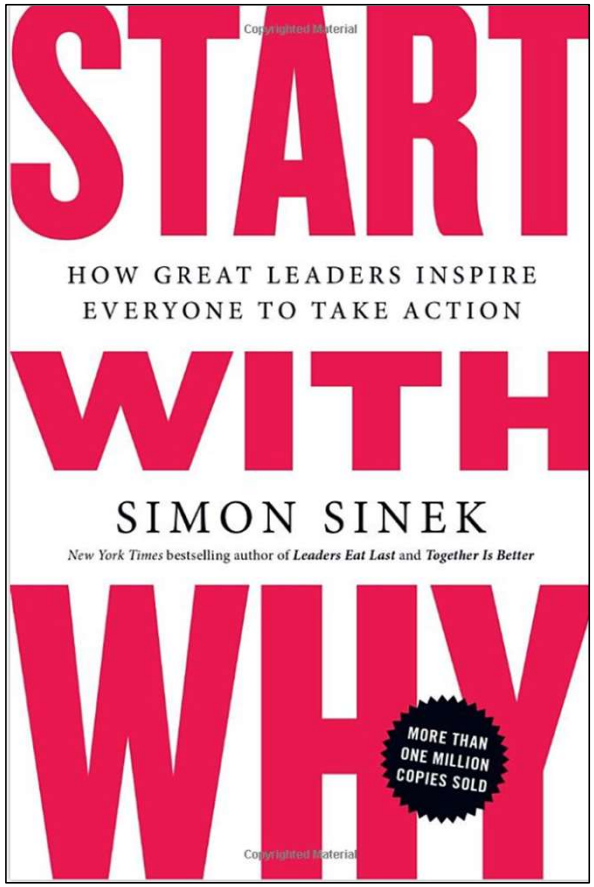
At least foundations of metrics that don't suck…

Principles of Metrics

- **Metrics are a decision support tool for stakeholders**
- Good metrics are first and foremost:
  - Quantifiable or objectively measurable
  - Targeted to a specific audience
  - Denotes the success or failure of a process
  - Start with why
  - What story are you trying to tell?
  - What conclusion should the audience draw from my data?
  - Can I reasonably expect them to infer my intent from the metrics?

Building Good Metrics – Start With Why

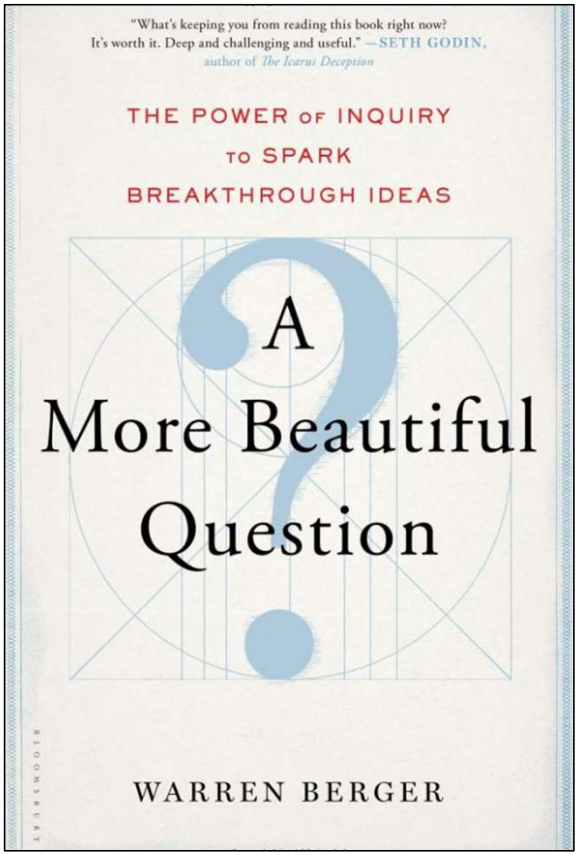
- The famous management book “Start With Why” answers the question of what really motivates us by looking at the Golden Circle (Think, Act, Communicate)
  - Inner Circle: Why
  - Middle Circle: How
  - Outer Circle: What
- We need to be able to answer these questions **for our stakeholders** *before* we start building security metrics



Building Good Metrics – Ask The Right Question

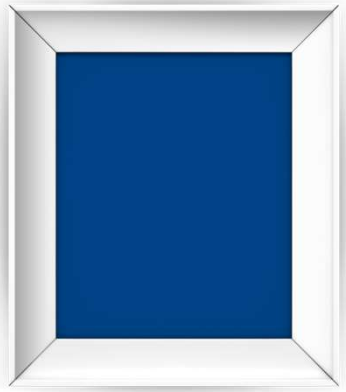
- Make sure you're asking the right question in the first place
  - If you don't ask the right question, getting the right answer is just luck
  - Aka: garbage in, garbage out

- As anyone with significant consulting experience can confirm, many orgs struggle with solving problems because they're asking the **wrong questions**



Building Good Metrics – Use Frameworks If You Can

- Use frameworks if they exist for what you're trying to measure
  - Frameworks show academic rigor
  - Even when not intended to create metrics, anything that has measurable success criteria (and most frameworks do) can be turned into a metric
- If the framework is proscriptive, but success criteria aren't present, ask:
  - What's the intent of this?
  - Are success measurements binary or scalar?
  - If scalar, how do we measure/rate it?

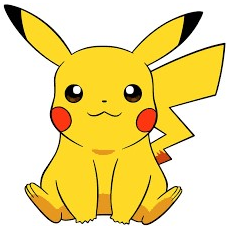


Metrics – Frameworks Example

- Google developed the HEART framework to address UX
  - Happiness
  - Engagement
  - Adoption
  - Retention
  - Task success

- Do any of these areas support good metrics?
  - Which one is best?
  - Why?

Building Good Metrics – Don't Measure Everything



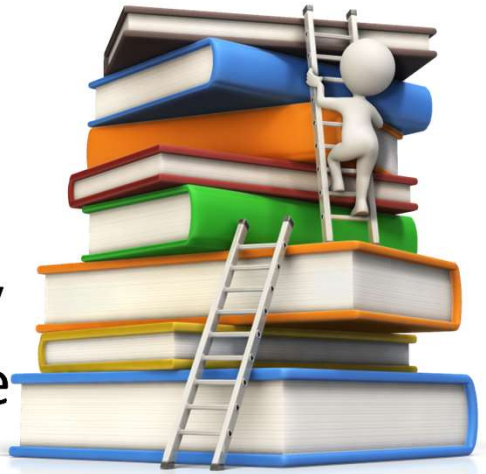
- Trying to measure everything is a fool's errand
- Many organizations treat metrics like Pokemon
  - Not only do these orgs drown in low quality data, they often miss better quality metrics
  - Something, something, quality over quantity…
- Remember (or realize) that every metric has a compliance cost
  - The data you don't store can't be compromised
- Metrics also impose cost on operations teams
  - ***And*** there's a cost for stakeholders to consume them

Building Good Metrics – Avoid Vanity Metrics

- Vanity metrics are metrics that make you feel good/look awesome, but don't really tell a coherent story
  - In many cases, they actually ***mislead*** stakeholders
- Unfortunately, vanity metrics are often the easiest to collect
- Security examples:
  - Number of port scans blocked by the boundary firewall
  - Number of log events collected in the SIEM
  - Number of IPs blocked via a threat intelligence feed
- These only look impressive if you don't understand them

Building Good Metrics – Don’t Educate in the Metric

- There's an old sales adage that you rarely close a sale in the same meeting where you introduce the product
- Applying this to metrics, you shouldn't educating the audience about a problem ***with*** a metric
- Educate the audience on the situation, ensure they understand it, then use metrics to demonstrate the degree or scope
- Put another way: without appropriate context, the data you are showing is data, **NOT** information



SOC Metrics

Not to be confused with "sock metrics"

SOC Metrics

- A few example SOC metrics (depending on intended audience):
  - Person hours committed to working alarms
  - Person hours committed to engineering new and better detections
  - Number of new detection rules created (and source for each)
  - Number of tuned detection rules
  - Number (and severity) for alarms by business unit
  - Detection source for alarms

SOC Metrics – BU Alarm Breakdown

- The "Number (and severity) for alarms by business unit" is a **VERY** easy metric to get **VERY** wrong
  - Even assuming that the data is correct, it can still be VERY misleading to the audience
- Differences in BU work habits will impact the data
  - Manufacturing line workers are less likely to be phishing victims than knowledge workers
    - Are they really better at avoiding phishing or are they just in their email less?
  - DevOps teams were responsible for most watering hole attack alarms
    - Most users can't install their own software, so watering hole attacks would probably impact them less

SOC Metrics – Detection Sources for Alarms

- Reporting on detection sources for alarms helps to drive understanding of where to dedicate tool training dollars
  - Do not confuse this metric with the **types** of alarms
  - While the type of alarm and detection source are often tightly correlated, these do not represent the same information
- Note: ensure to communicate to your audience defense in depth may result in some tools never seeing data needed to generate an alarm

SOC Metrics –Detection Engineering Hours

- Good detection engineering is one of the most important measures of SOC maturity
- If hours aren't dedicated to detection engineering:
  - Senior analysts are overworked with alarms?
  - Analysts lack the skills necessary skills to perform the task?
  - The organization isn't prioritizing detection engineering?
- None of these are good and illuminate opportunities for improvement
  - To argue otherwise is to claim your detections are just fine as-is

Cyber Threat Intelligence (CTI) Metrics

The number of IOCs in your automated feed does NOT count…

CTI Metrics

- A few example CTI metrics (depending on intended audience):
  - Number of RFIs answered
    - Subdivided by analyst
  - Quantity of person hours per RFI (by business units)
  - Number of CTI-enabled detections
  - Percentage of indicators CTI enabled advance warning for
    - This is before the indicators were generally available (e.g., "FBI scoop")
  - Net promoter score for RFIs (and potentially other services)
    - This definitely warrants separating by analyst or team

CTI Metrics – Net Promoter Score

- Because CTI reporting is extremely subjective, it is important to measure the quality of reporting
  - Note that feedback (e.g., "how do I make this reporting more valuable to you?") is **not** a metric (fails the measurement test)
- Some organizations use a Likert scale for measuring the quality of CTI reporting
- The Net Promoter Score (NPS) is often a better fit
  - NPS is a well-understood measure of how likely a consumer is to recommend a product or service to others
  - This is understood to generally align with quality of the overall process

CTI Metrics – "Scooping the FBI"

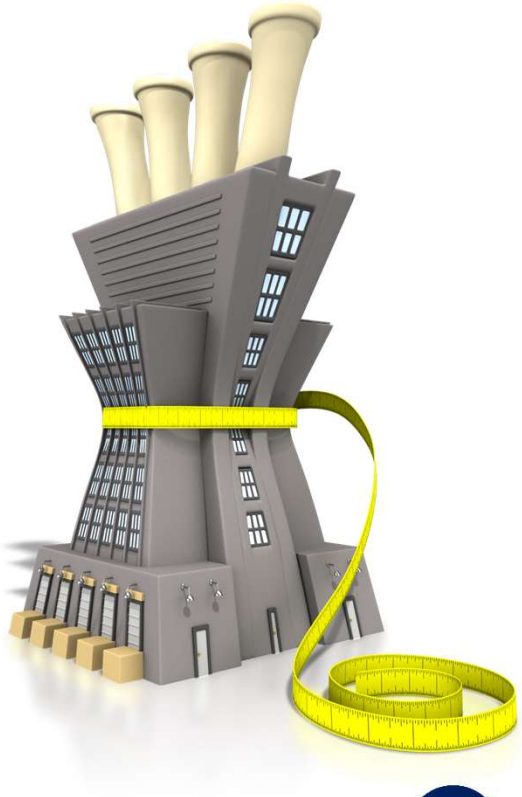
- When a new "FLASH" report is issued by CISA or the FBI, parse it and extract indicators
- For each indicator, search your Threat Intel Platform (TIP) and determine whether you already knew of the indicator, whether it has been operationalized, and when for both elements
  - Report the percentage of indicators already covered
- Bonus points for reporting:
  - Average age each indicator has been on coverage
  - Number of detections enabled with the indicator

Threat

The best "

Threat Hunting Metrics

- A few example thrunting metrics (depending on audience):
  - Number of hypotheses tested
  - Source for hypotheses tested
  - Number of intrusions detected (**DANGER!!!**)
  - Number of security hygiene items detected
  - Number of unique MITRE ATT&CK techniques tested in hypotheses
  - Number of hypotheses converted to SOC detections



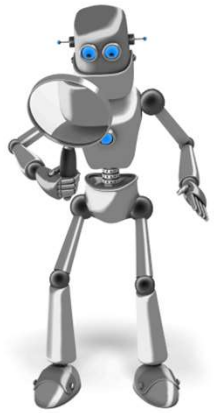
Thrunting Metrics – Hypotheses Sources

- Thrunting is inherently driven by hypotheses
- Ensure that analysts track:
  - The source of hypotheses
  - Which sources produce the highest number of detections
  - The sources that highlight telemetry gaps
- This allows analysts to prioritize and obtain optimal outcomes
  - Over time, it will become clear where to dedicate limited threat hunting resources

Thrunting Metrics – Intrusions Detected

- **DO NOT USE THIS METRIC WITHOUT FIRST EDUCATING YOUR AUDIENCE OR I WILL CURSE YOU UNTIL THE END OF DAYS**

- If a hypotheses is tested and returns no detections, that is NOT a threat hunting failure
  - You still have knowledge you didn't ***before*** the test
- Contextualize reporting of any intrusions detected
  - This number should almost always be low



Thrunting Metrics – Detection Engineering

- A primary output of threat hunting is detection engineering
- If an intrusion is detected, the analyst should ask:
  - Why did our existing systems miss this intrusion?
  - What telemetry am I seeing now that was previously missed?
  - How can this telemetry search be turned into a detection?

- Note: sometimes acceptable rates of false positive reduction cannot be achieved to create ongoing detection rules
  - Over time, the "why not" (telemetry gaps, unacceptably high background noise, etc.) creates another metric of its own

Incident Response (IR) Metrics

Because incident response sucks enough without bad metrics…

Incident Response (IR) Metrics

- A few example IR metrics (depending on audience):
  - Number and type of incident escalations
  - Percentage of escalations that could have been handled by SOC
  - Incidents handled without outside assistance
    - Percentage of work performed by in-house analysts
  - Detection methods for escalations (grouped by incident type)
  - Percentage of person hours spent overcoming telemetry gaps
  - Number of person hours spent on investigation per incident
    - Subdivided by incident type or severity
  - Number of lessons learned documented during the incident
    - Lessons learned owned by the IR team and actioned within *n* days

IR Metrics – Lessons Learned

- So much of incident response feels like déjà vu
  - That's because most orgs treat lessons learned as a check box action
- Lessons learned is my favorite metric for maturing an IR team
- Tracking lessons learned ***during*** the incident is paramount
  - High numbers may indicate complex incidents or process failures
- Lessons learned actioned measures continuous improvement
  - It's important to subdivide this metric by the owning business unit
  - The IR team should be measured for the lessons learned they can action in-house

IR Metrics – Evidence Acquisition Wait Time

- Far too much time in IR is consumed waiting for the best evidence (which changes during the investigation) to analyze
- Tracking the time between evidence request (e.g. "all firewall logs for the last 7 days) and evidence delivery highlights which teams may be roadblocks in the process
  - The evidence tracking spreadsheet denotes whether a particular evidence request is blocking
- **Note:** It is critical to communicate that IR teams don't simply wait on this evidence to arrive (because sometimes it never does) and instead analyze evidence that is already available

Closing Thoughts

Because I can't close talking about thrunting…

Closing Thoughts

- Start by defining what story you want to tell
  - How will this provide decision support to my audience?
- Make your metrics meaningful to stakeholders
  - Ideally, metrics should also drive practitioner behaviors
  - Ensure measurement is consistent and repeatable
- Vanity metrics are the devil
  - They confuse stakeholders (and will eventually torpedo your credibility)

Jake Williams

@MalwareJake

SCYTHE

@scythe_io