“Lies, Damned Lies, & Statistics”

Improving the Effectiveness of Indicators of Compromise

Industrial Control Security Conference 2014

Seth Bromberger
Principal, NCI Security LLC
seth@ncisecurity.com

NCI Security LLC
Protecting the Nation’s Critical Infrastructure

About NCI Security LLC

“

STATISTICS:

The science of producing

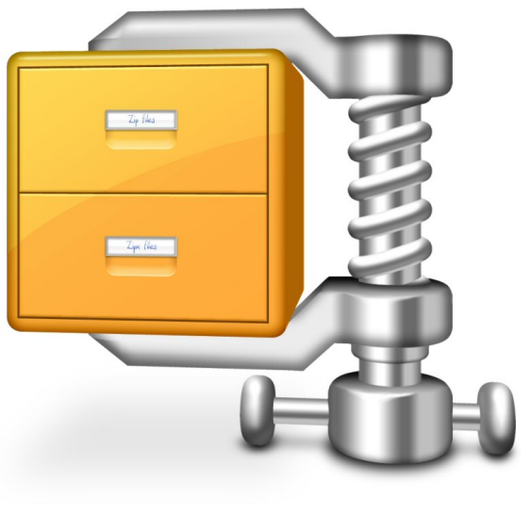
unreliable facts

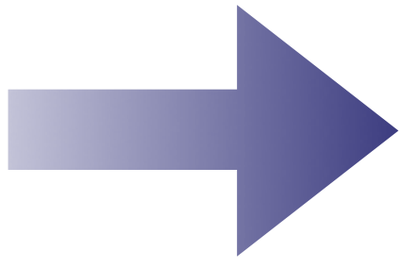
from **reliable figures.**

”

– Evan Esar

Traditional Threat Detection





`35fb761548845431bc1807fbb868caf7`

"Does this hash match any known malware?"

Traditional Threat Detection

- ✓ Low False Positives
- ✓ Fast determination
- ✓ Easily distributed
- ✓ Simple implementation
- × False Negatives wildly variable
- × Relies on up-to-date signatures
- × All-or-nothing

Characteristics are significant

Advanced Threat Detection

✓ Lower False Negatives

✓ Fewer updates required

✓ **Finds the elusive 0-day**

× Higher False Positives

× Slower determination

× Not easily distributed

× Complex implementation

× All-or-nothing

Behavior is significant

Heuristic-based Threat Detection

✓ Balanced Positives

✓ Customizable for organization's risk appetite

✓ All the advantages of Advanced Threat Detection

× Requires Advanced Threat Detection

× Requires feedback loop

History is significant

Data

Data

Information

Data

Information

Knowledge

Data

Information

Knowledge

Understanding

STATISTICS

Statistics Primer, Part I

Single Test

outcome

observation

	sick	well
“sick”	TP	FP
“well”	FN	TN

Statistics Primer, Part I

Multiple Tests (n=55)

	T	F	Σ
P	17	3	20
N	21	14	35
Σ	38	17	

Positive Predictive Value (PPV): $TP / \Sigma_P = 17 / 20 = 0.85$

- AKA “Precision”
- Measures the probability that a “positive” result is ACTUALLY positive.
- Use when the cost of FP is very high relative to missing TP.
- Disadvantageously influenced by the total number of positives in the population (Σ_P influences significance, which is not ideal when unbalanced).
- The rarer the condition is, the more influence FP has.

Statistics Primer, Part I

Multiple Tests (n=55)

	T	F	Σ
P	17	3	20
N	21	14	35
Σ	38	17	

Negative Predictive Value (NPV): $TN / \Sigma_N = 21 / 35 = 0.60$

- Measures the probability that a “negative” result is ACTUALLY negative.
- Use when the cost of a FN is very high.
- Disadvantageously influenced by the total number of negatives in the population (Σ_N influences significance, which is not ideal when unbalanced).

Statistics Primer, Part I

Multiple Tests (n=55)

	T	F	Σ
P	17	3	20
N	21	14	35
Σ	38	17	

Accuracy: $\Sigma_T / n = 38 / 55 = 0.69$

- Measures the degree to which the test reflects the actual condition.
- High ACC: a given result is likely to be correct.
- Use when it's more important to ensure you've got a balance between FP and FN.
- Does not distinguish between FP and FN: an error is an error.
- This implies that unbalanced data are inappropriate for this sort of statistic.

Real-World Application

- 334 malware investigations over 16+ months
- 65 indicators of compromise used
  - Investigation if at least one indicator triggered
- 40 instances of malware found
  - We assume accuracy of outcome (forensics = “gold standard”)

Raw Data Extract

MALWARE?	Deleted Itself	Spawned New Process	Modified Registries	Started/ Stopped System Service	Injected code into process	Attempted to sleep
F	F	T	T	F	F	F
F	F	F	T	F	F	F
T	F	T	T	T	F	T
F	F	F	T	F	T	F
T	T	F	T	T	T	T
F	F	F	T	F	F	F
T	F	F	T	F	T	T
F	F	F	T	F	F	F

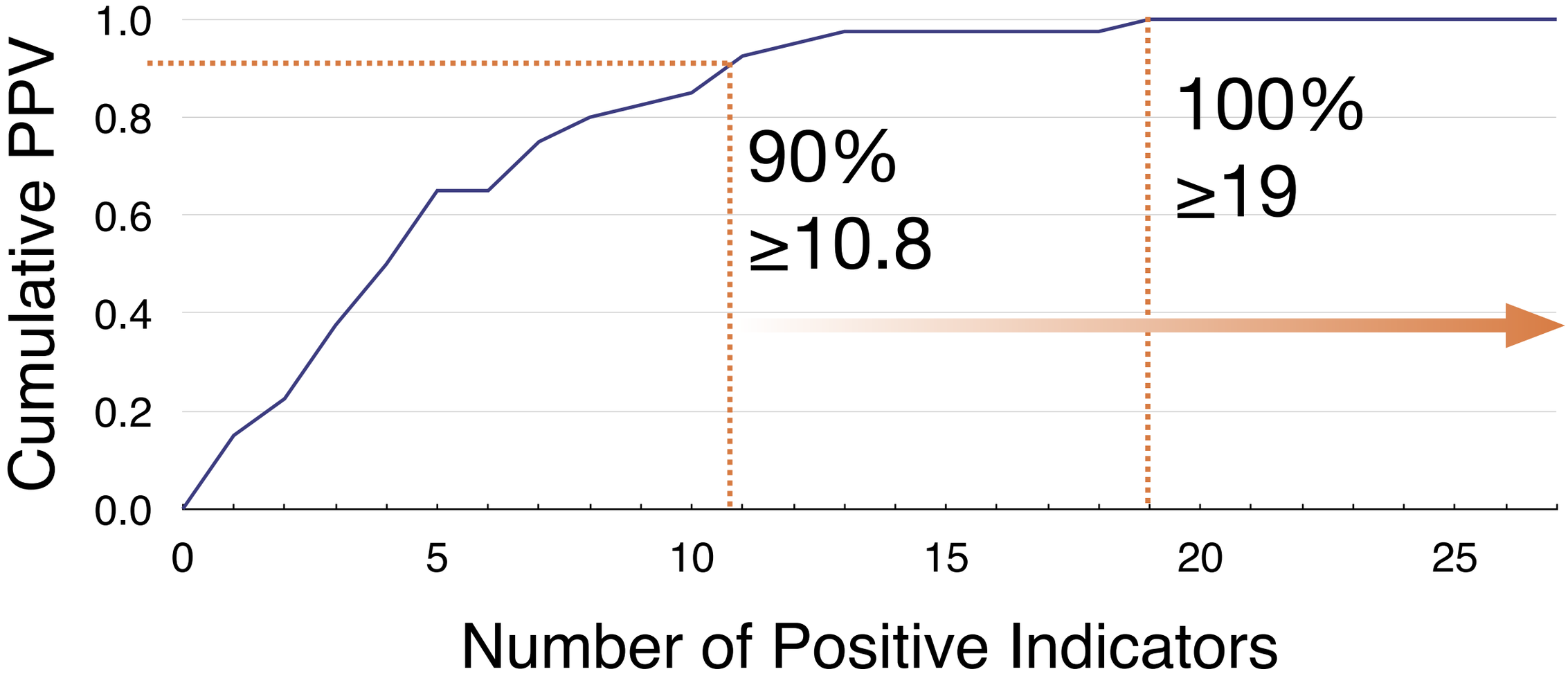
Real-World Data

Multiple Tests (n=334)

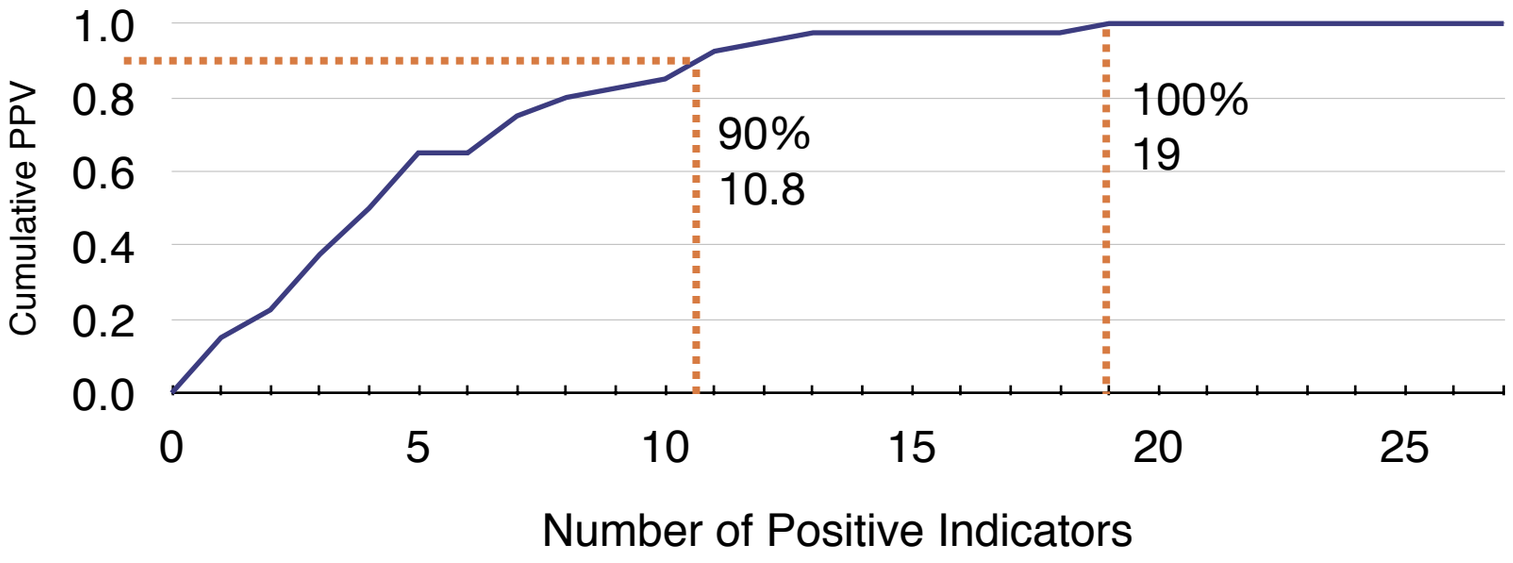
	Avg # of Indicators	Standard Deviation	Min	Max
P	11.18	4.88	2	27
N	5.15	3.14	1	17

Malware samples had, on average, 6 more positive indicators than non-malware

Real-World Data



Real-World Data



- 90% probability that event with 11+ indicators is malware
- Malware is (almost) assured when you hit 19 indicators

Indicator Analysis

- We haven’t looked at the actual indicators yet.
- Are there indicators that are more likely to indicate malware? (Are there correlations?)
- What about **combinations** of indicators?

“

STATISTICS

are no substitute for

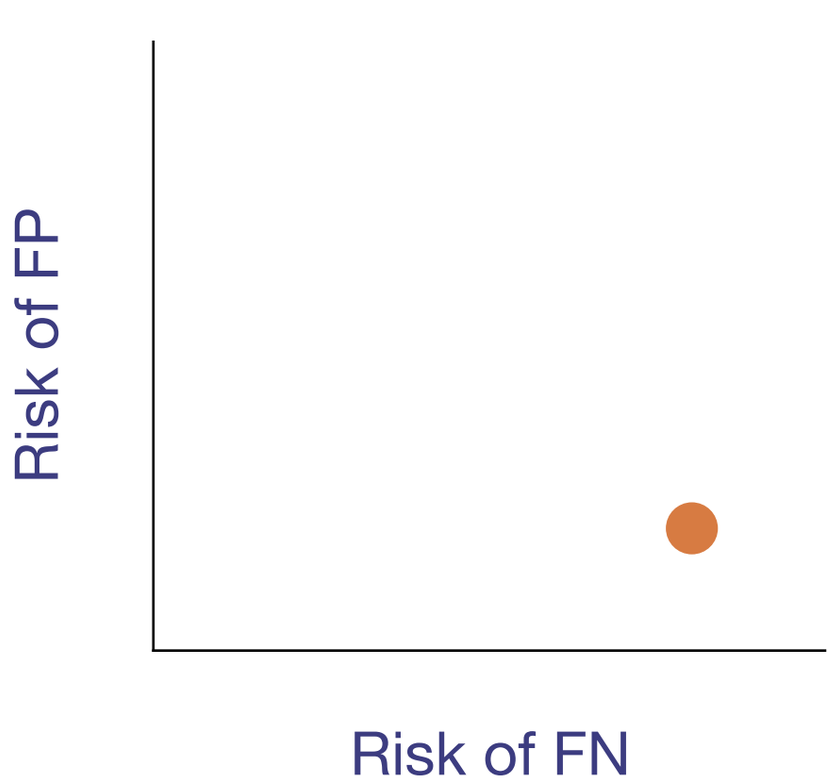
JUDGMENT

”

– Henry Clay

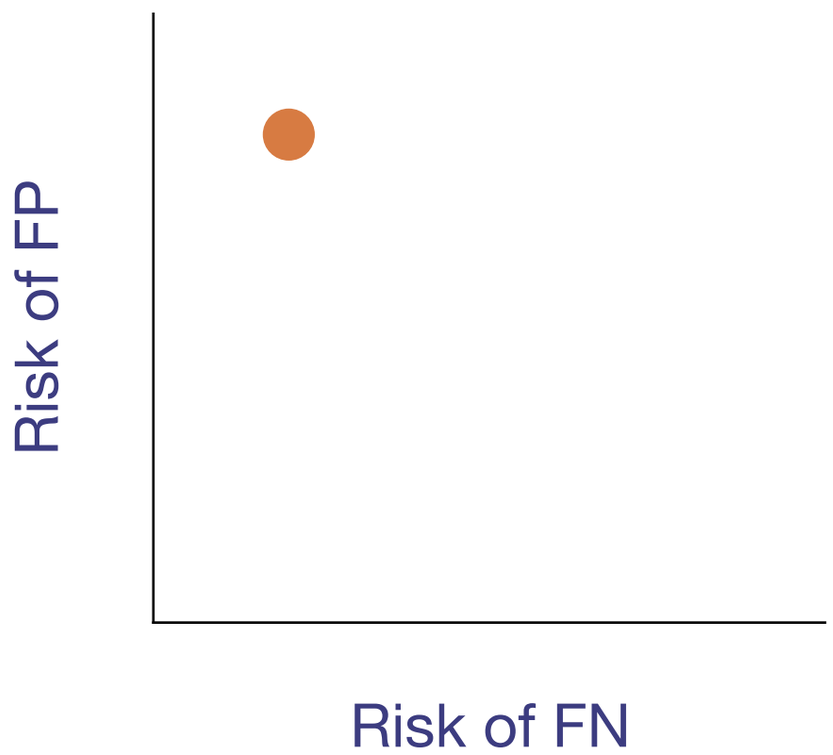
What Statistics Are Important?

- If you have overworked / small teams, PPV is probably ok



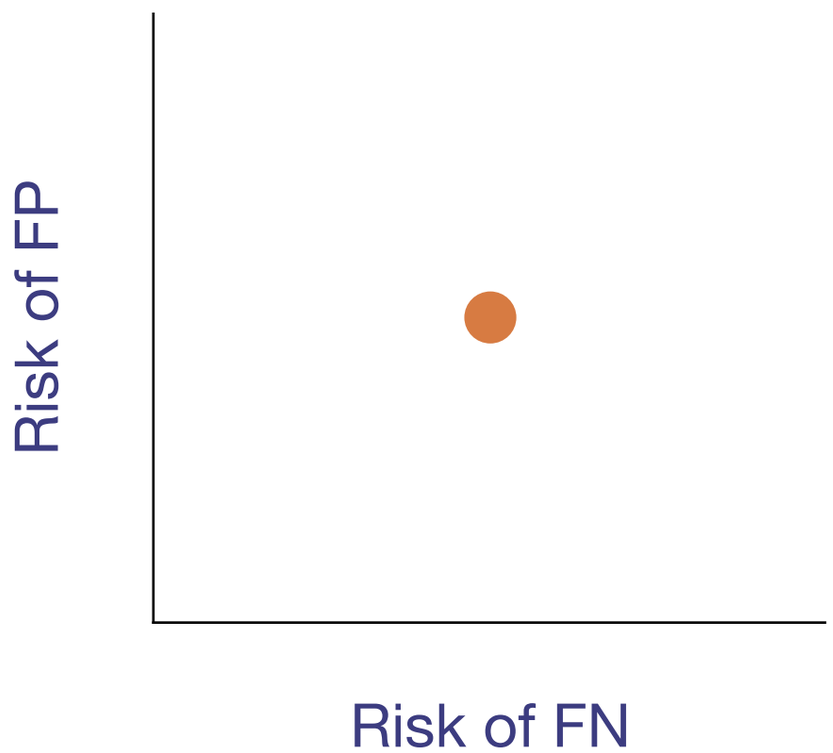
What Statistics Are Important?

- Resources to spare? NPV is probably ok.



What Statistics Are Important?

- Unbalanced Data? Look at MCC.



Statistics Primer, Part II

Multiple Tests (n=55)

	T	F	Σ
P	17	3	20
N	21	14	35
Σ	38	17	

Matthews Correlation Coefficient (MCC):

$$\frac{(TP)(TN) - (FP)(FN)}{\sqrt{(\Sigma_P)(\Sigma_N)(TP + FN)(TN + FP)}} = 0.44$$

- Does not require balanced data.
- Range is [-1, 1] - different from other statistics. 0 is “no better than random correlation”, 1 is “perfect correlation”, and -1 is “perfect disagreement”.

“

STATISTICS

are like

BIKINIS:

What they reveal is ***suggestive***,
but what they conceal is ***vital***.

”

– Aaron Levenstein

Real-World Results

- Plenty of PPV = 1
    - Re-ordered by accuracy to penalize false negatives without impacting efficiency
- Decision to respond automatically resulted in significant time savings
    - …at the cost of decreased data training / modeling opportunities
- Using combinations of observations yielded better (more specific) results ($n < 8$)
    - Compute-intensive: $\binom{65}{8}$ = ~5.04 billion combinations
- Approach can be used as a basis for first-level incident response

Key Messages

- Statistical analysis is an effective method for improving efficiency of incident response
- Choose a statistical approach based on characteristics of actual data
- Understand limitations and benefits of selected approach. Make informed risk decisions.
- Minimize manual analysis. AUTOMATE!

Thank you!

NCI Security LLC

Protecting the Nation's Critical Infrastructure

info@ncisecurity.com

+1 415 890 2233