What are the Chances?

Quantitative Methods for Managing Cyber Risk

Doug Clare
VP, Security Solutions
FICO

A History in Addressing Risk Quantification Challenges

FICO Decisions

1980 Consumer Credit Scoring

Opportunity

- Apply behavioral analytics and predictive scoring to drive cost efficiency and scale in consumer credit underwriting and portfolio management

Solution

- FICO® Consumer Credit Score
- Rank-order consumers based on likelihood of paying their credit obligations

Result

- Greatly expanded access to consumer credit
- 10+ billion FICO Scores purchased annually
- Most widely used credit score in the world

2018 Cyber Risk Scoring

Opportunity

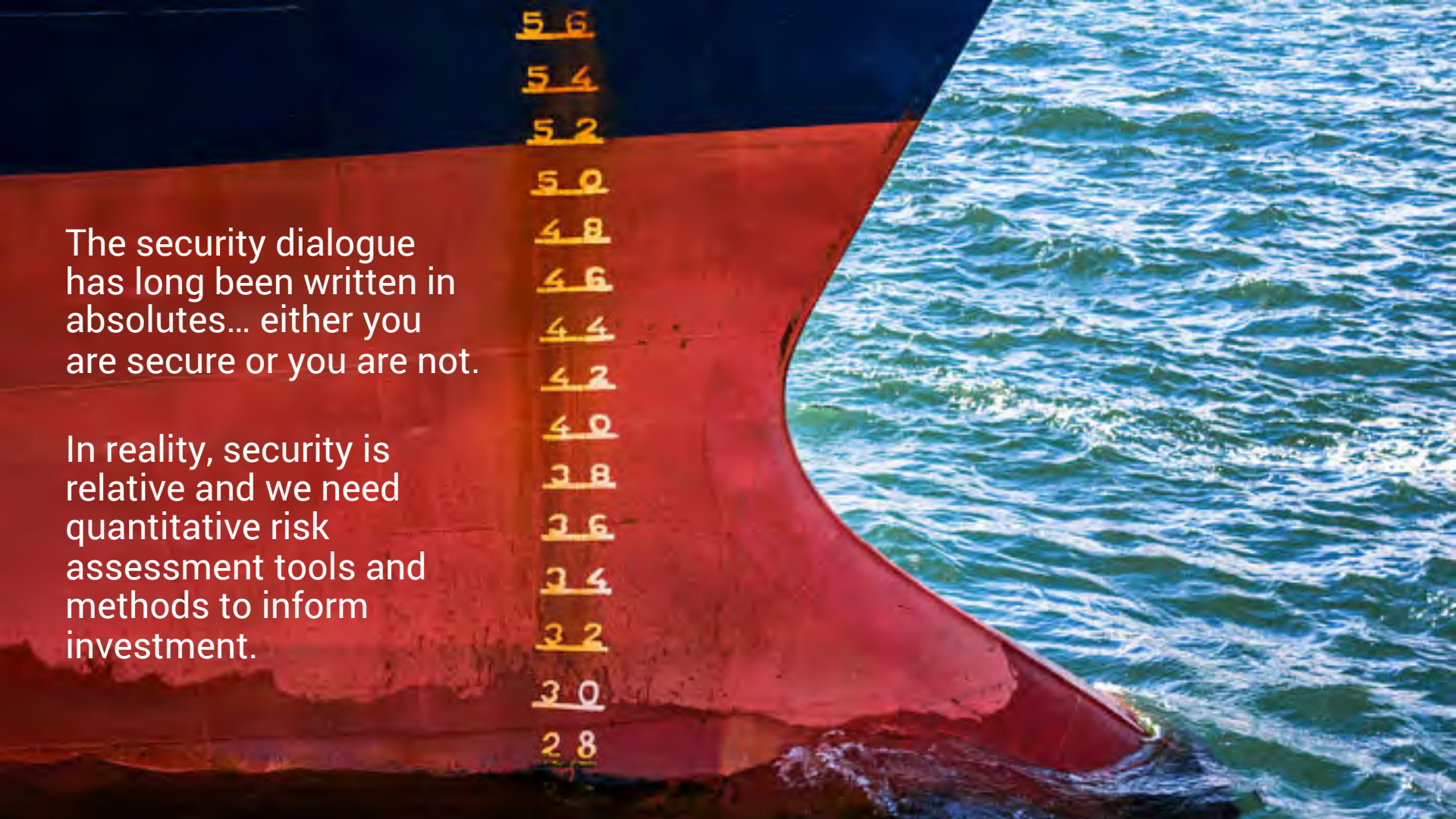
- Apply behavioral analytics and predictive scoring to drive cost efficiency and scale in third party risk management and cyber insurance underwriting

Solution

- FICO® Cyber Risk Score
- Quantifying organizational likelihood of suffering a material cyber breach

Result

- Only empirically-derived quantification of risk
- Continuous supply chain monitoring
- Cyber risk underwriting and portfolio mgmt



The security dialogue has long been written in absolutes… either you are secure or you are not.

In reality, security is relative and we need quantitative risk assessment tools and methods to inform investment.

The FICO Cyber Risk Score

- Empirical, passive assessment of forward-looking breach risk based on condition and behavior
  - Supervised machine learning
  - Thousands of breach exemplars
  - Billions of data points
- 3-digit score encapsulates the future likelihood of a significant breach event
- Reason codes detail primary risk vectors
- Range from 300 – 850 (higher = less risk)
- Serving distinct use cases:
  - Objective self-assessment
  - Supply chain risk management
  - Insurance underwriting and pricing
- Strong separation of goods and bads
  - 24X dynamic range (relative odds)

FICO Cyber Risk Score IP Lineage

DHS-funded research on global internet security threat identification / quantification

Internet-scale data collection

Entity-level security risk correlation

Internet-wide historical signal database

FICO™

Deep expertise in model characteristic engineering

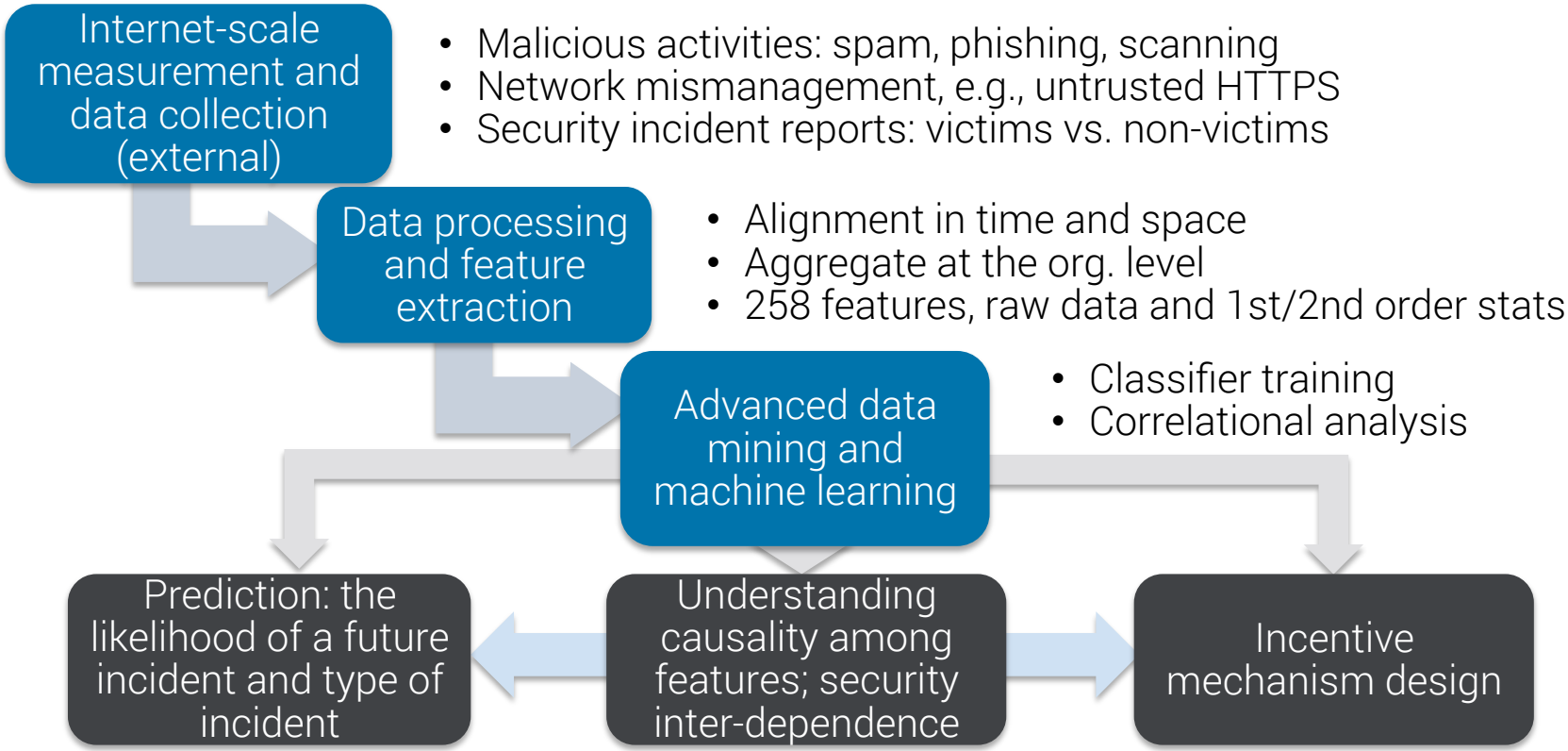
Multiple patented machine-learning techniques

Predictive analytics IP

Operational know-how / software assets for deploying analytics into production workflow

Research funded by the DHS: Predictive Data Analytics

- Data collection followed by supervised learning



Research Study 1: Are Network misconfiguration correlated with maliciousness?

- Measured, at Internet-scale, correlation between networks following best practices and malicious activity from the networks (e.g., botnet infections, spam).
- Example best practices studied: not allowing public access to DNS resolvers, managing SSL certificates carefully; disallowing untrusted email, etc.

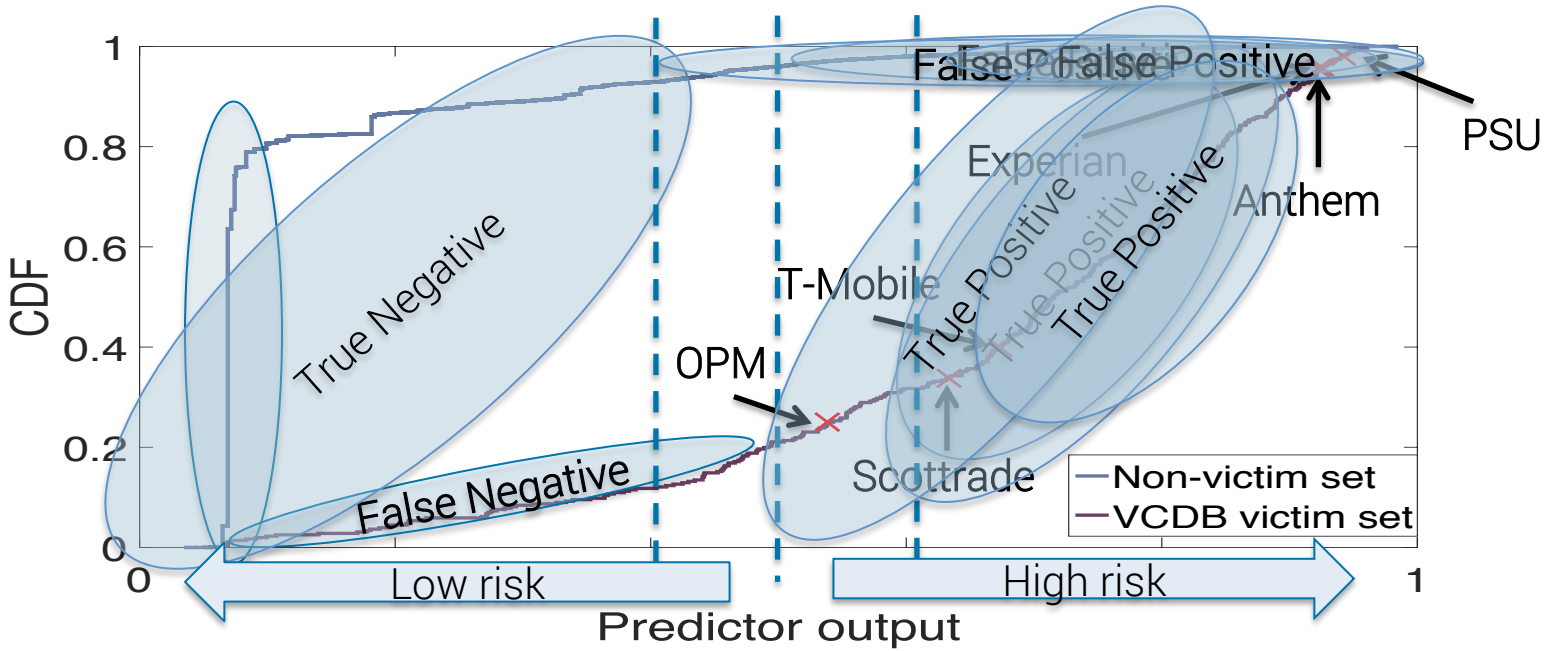
Misconfiguration Metric (examples from publication)	Correlation with Maliciousness (-1.0 to +1.0)
Open DNS Resolvers	+0.59
DNS Source Port non-Random	+0.45
Untrusted HTTPS Certs	+0.44
OVERALL	**+0.64**

Networks that do not follow best practices for configuration are more likely to harbor compromised hosts or malicious insiders.

(Zhang, et al. 2014)

Research Study 2: Supervised learning approach to predicting breaches

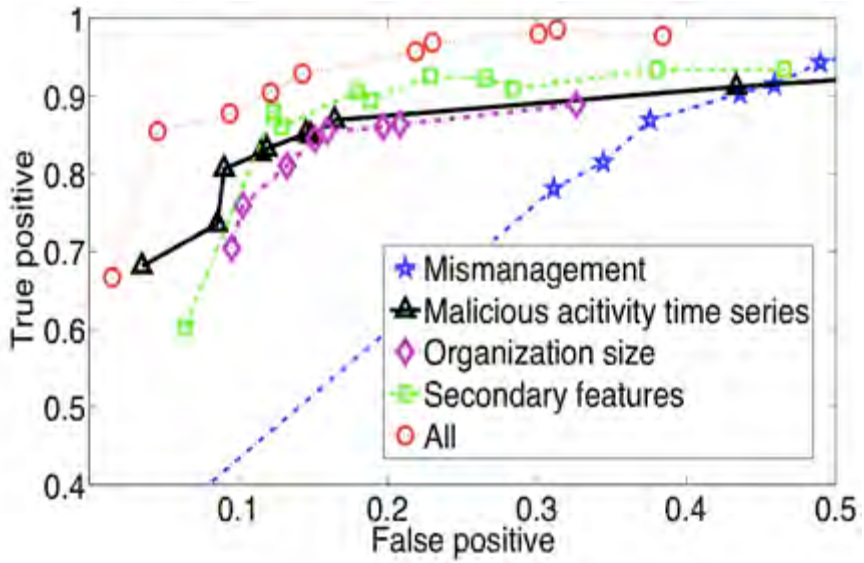
- Supervised learning using reported data breaches; built classifier/predictor
- Output is a number between 0 and 1



(Liu, et al., 2015)

Research Study 2: Can Misconfiguration and maliciousness predict breaches?

- Classifier's prediction performance per independent feature set shown in graph
- Dynamics of malicious activity ("secondary features") captures org security response nimbleness and efficacy
- Diversity of signals best picks up underlying phenomena leading to breach



- Combining mismanagement and malicious activity dynamics features was most effective for prediction, as capture org policies and behaviors well

(Liu, et al., 2015)

What does Quantitative Security Research Teach Us?

1. A lot of information is contained in externally measurable data

2. A diversity of data measurements can be used to accurately predict breaches

Example Misconfiguration Features

Exposed internal services, globally reachable

- Development web servers (e.g., TCP/8080)
  - (example screenshot from FICO ESS)

IP	Hostname	Indicator	First Observed
[redacted]109.78	v5.eecs.umi[redacted]	ALT-HTTP(TCP/8080)	2017-11-27
[redacted]14.21	cfinelli-lapto[redacted]	ALT-HTTP(TCP/8080)	2017-12-18
[redacted]14.128	blindspot.ee[redacted]	ALT-HTTP(TCP/8080)	2017-11-27

  - Common attack vector as often not policy-compliant, unfinished code, buggy, unmonitored, etc.

- MySQL Database (TCP/3306)
  - Should not be globally accessible, as provides ripe target for attackers
  - Often, exploitable configuration or unpatched bugs can lead to total date exposure

Example Misconfigurations, Cont'd

Exposed misconfigured Infrastructure examples

- SNMP – Simple Network Management Protocol (UDP/161)
  - default community/password: "public"
  - Allows gleaning inside operational information about networks
  - Allows use as DDoS reflector; may allow remote reconfiguration of device

- NTP – Network Time Protocol (UDP/123)
  - NTP monlist and version command responses to external probes
  - Same two reasons as above, and has been a very common DDoS vector (Czyz, et al., 2014)
  - Can leak internal network addresses, as well as hardware versions

Example Misconfigurations, Cont'd

Mismanaged SSL Certificates

- SSL Certificates are used whenever website presents https:// URL; they are a bedrock of the modern Internet
- When mismanaged, can lead to man-in-the-middle attacks or encourage users to "just click ok" on warnings
- Example management problems: include **expired, self-signed, or untrusted** certs
- Most importantly for us, **bad certificate hygiene speaks to poor network management**

Example Malicious Activity

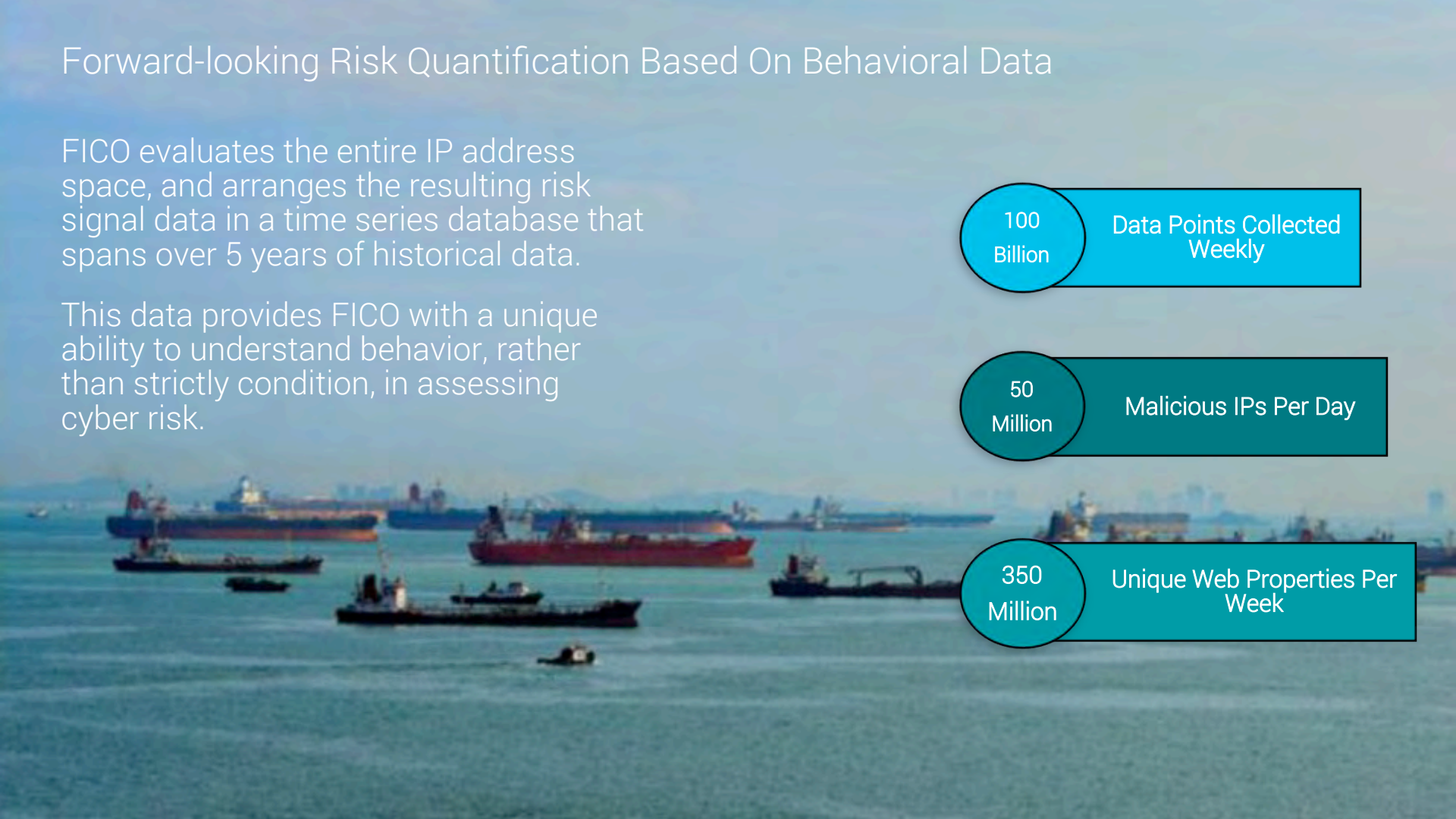
Endpoints on a network that are hosting a phishing site

- A direct indicator of a host being compromised, usually due to a botnet
- Compromise itself is not great (poor endpoint protection or user education),
- But, when we see it persist over time, it is a strong signal that the operators are not minding the shop (e.g., poor monitoring)

Forward-looking Risk Quantification Based On Behavioral Data

FICO evaluates the entire IP address space, and arranges the resulting risk signal data in a time series database that spans over 5 years of historical data.

This data provides FICO with a unique ability to understand behavior, rather than strictly condition, in assessing cyber risk.



How is it Built?

- Thousands of breach exemplars
- Billions of time-series data points
- Dozens of engineered features designed to expose and amplify risk signal
- Supervised machine-learning algorithms correlate condition, behavior with target outcomes
- Validated against hold-out and out-of-band samples
- Strong good-bad separation – 24X dynamic range
- Proven in the field

FICO Enterprise Security Score
Distribution and outcome odds (v2.3)

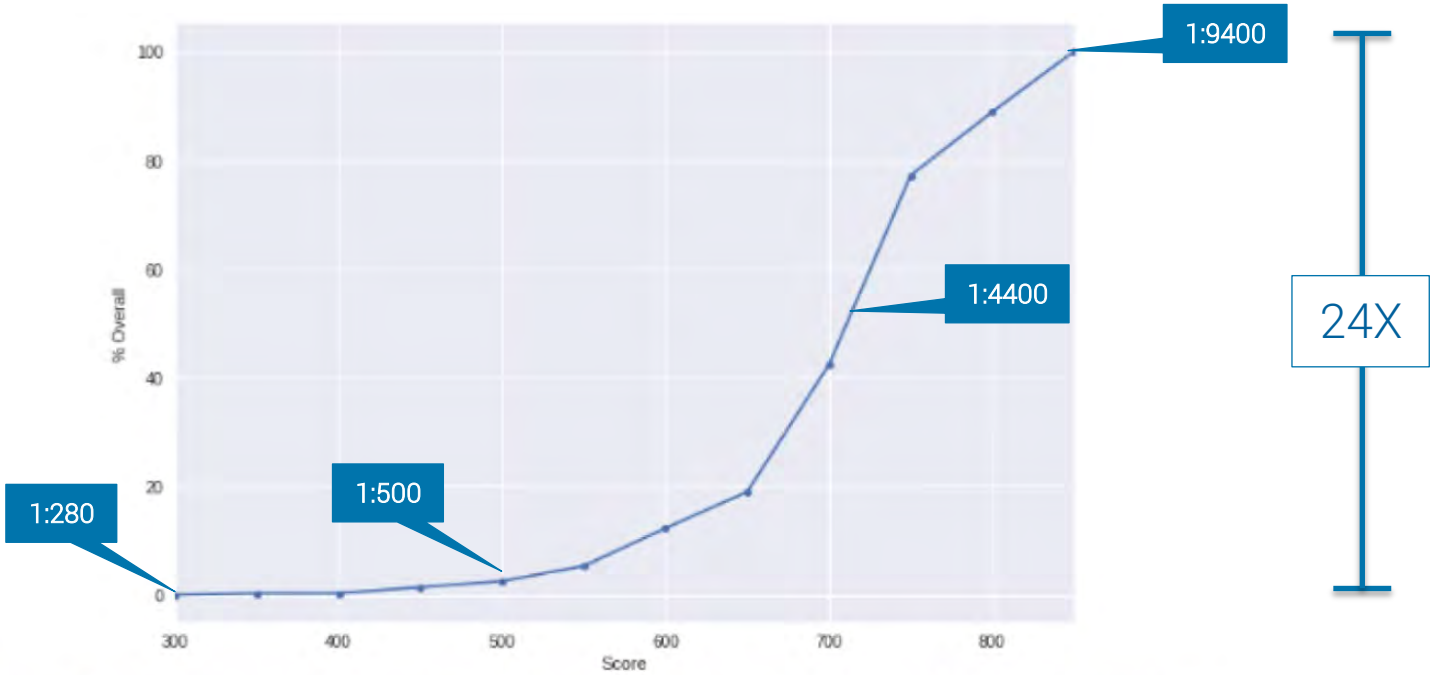


Figure 1: Organization cumulative score distribution. Each point indicates the percentage of organization scores at or below that score threshold, as observed in the "holdout" evaluation dataset used in model development. Because 850 is the maximum score, 100% of organizations score at or below that threshold.

Odds of a significant breach event double with each 84-point drop in the score

What is the ABC?

- A quantitative benchmark for measuring and tracking progress in cybersecurity for US business
  - Published quarterly
  - Based on the FICO Cyber Risk Score – an empirical assessment of organizational cyber breach risk
- A catalyst for discussion
- A means of better understanding change over time
- A rallying point for action in improving cyber risk posture

Resources:

Get the US Chamber of Commerce Assessment of Business Cybersecurity at:

https://www.cyber-abc.com

Get your free FICO Cyber Risk Score at:

https://cyberscore.fico.com