CREDC

Forecasting Cyber Security Incidents in Energy Delivery Systems

Michael Bailey and Alfonso Valdes

Reputation Matters

Security posture is an important part of a business relationship.

Security posture is the sum total of all factors, including people, processes, and technology.

The Problem

Measuring security posture is difficult. Need metrics to assess a network or organization from the outside to determine risk.

The Audience

Homeland Security for **Critical Infrastructure**.

E.g., 2014 Breaches

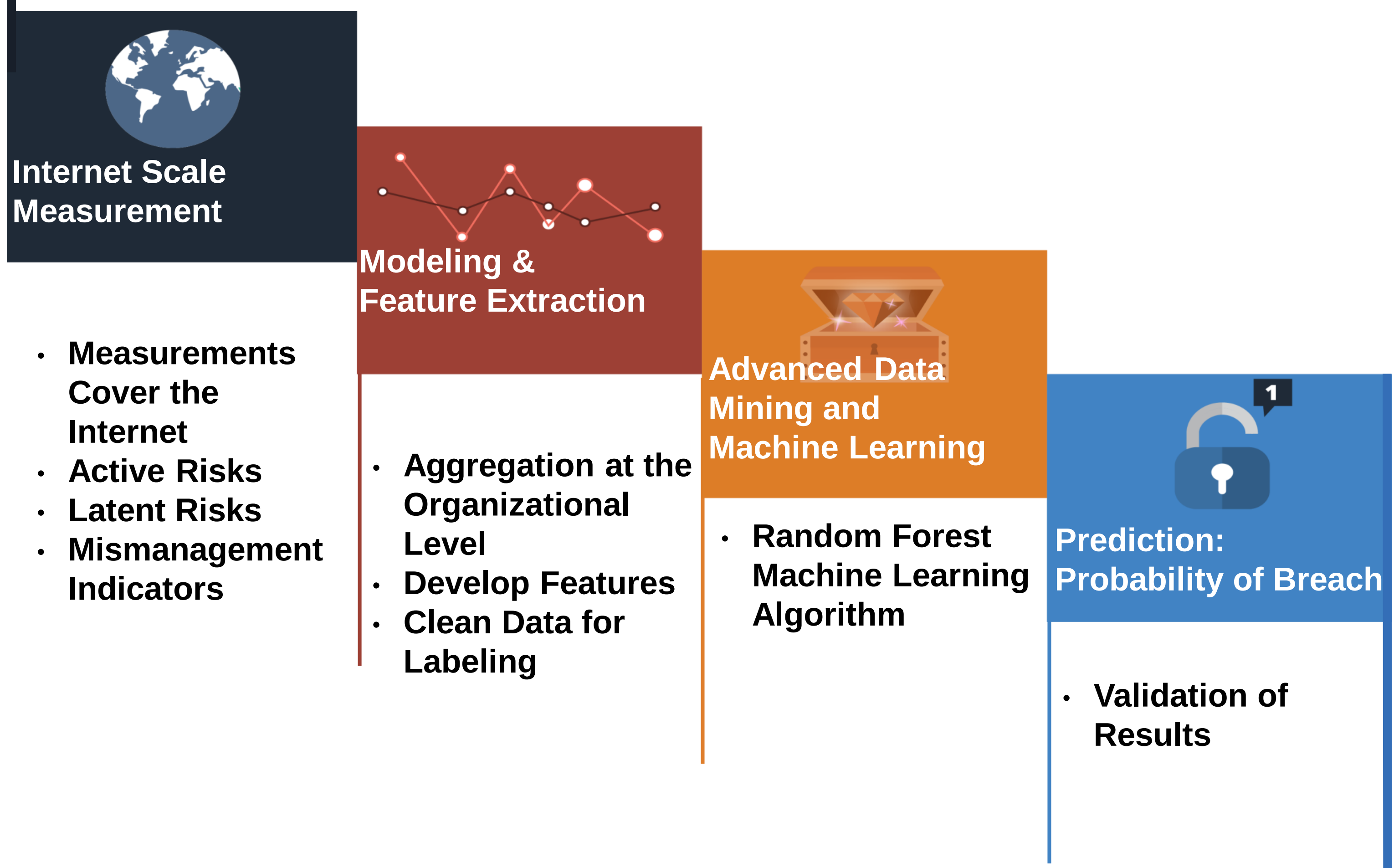
Target, Home Depot, JP Morgan, etc.

Companies for **3rd-party risk**.

Companies for **understanding** their own attack surfaces.

Underwriters for assessing probability of breach to determine premiums.

The Approach



Internet Scale Measurement

- **Measurements Cover the Internet**
- **Active Risks**
- **Latent Risks**
- **Mismanagement Indicators**

Modeling & Feature Extraction

- **Aggregation at the Organizational Level**
- **Develop Features**
- **Clean Data for Labeling**

Advanced Data Mining and Machine Learning

- **Random Forest Machine Learning Algorithm**

Prediction: Probability of Breach

- **Validation of Results**

The Data

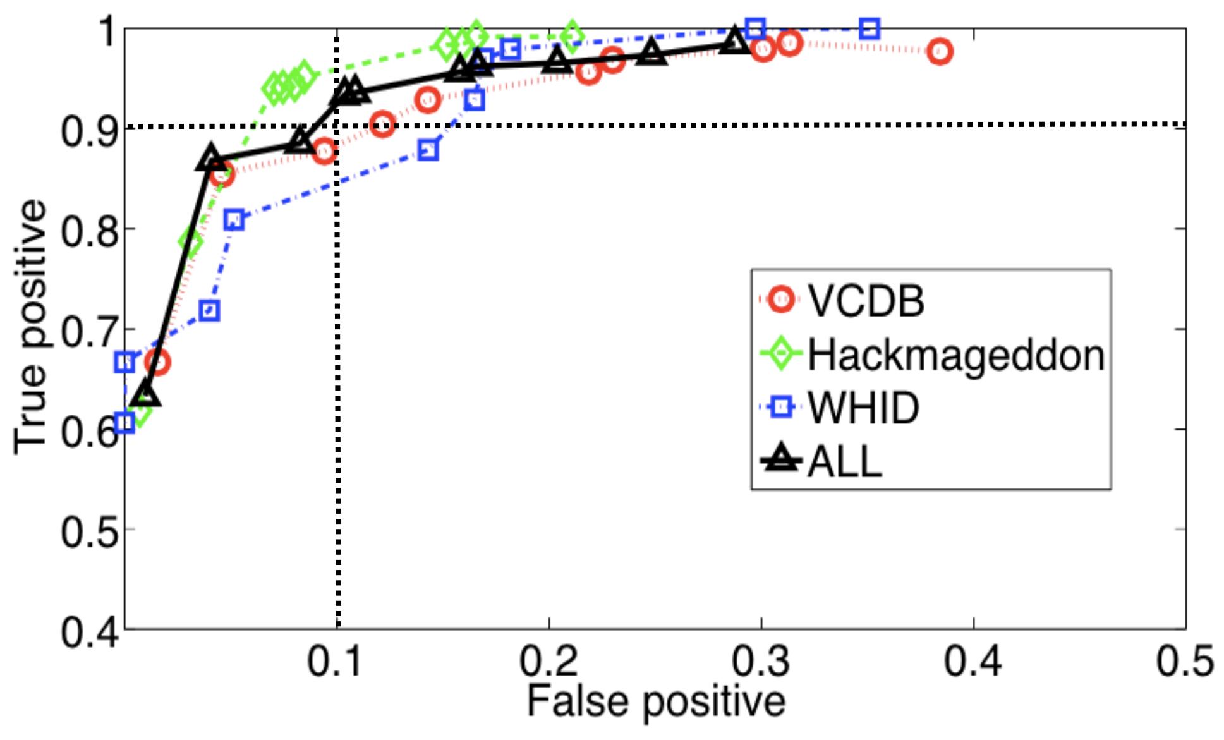
Inferred malicious activities from RBL lists:

- SPAMHAUS-XBL, SpamCop.
- UCE-PROTECT, SURBL,WPBL, PhishTank, HpHosts.
- Darknet Scanners, DSHIELD, OpenBL.

Ground-truth data used for identifying data breaches:

- VCDB: Veris Community Database (basis for Verizon Data Breach Investigations Report).
- Hackmageddon.
- Web Hacking Incidents Database.

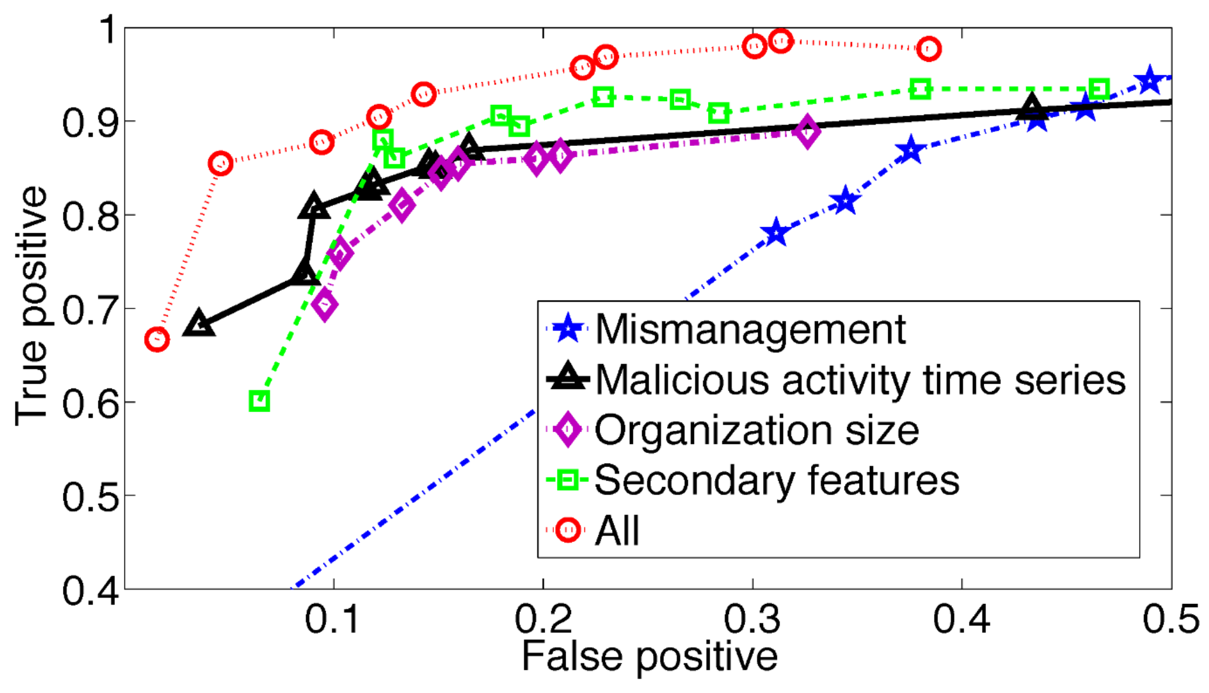
RESEARCH RESULTS



- Overall a combined 90% true-positive, 10% false-positive.
- 90% of the organizations in the test victim set were correctly rated as "high-risk" according to the models.

Feature Category	Normalized Importance
21	0.3229
Time Series Data	0.2994
Recent-60 Secondary Features	0.2602
Organization Size	0.0976
Recent-14 Secondary Features	0.02

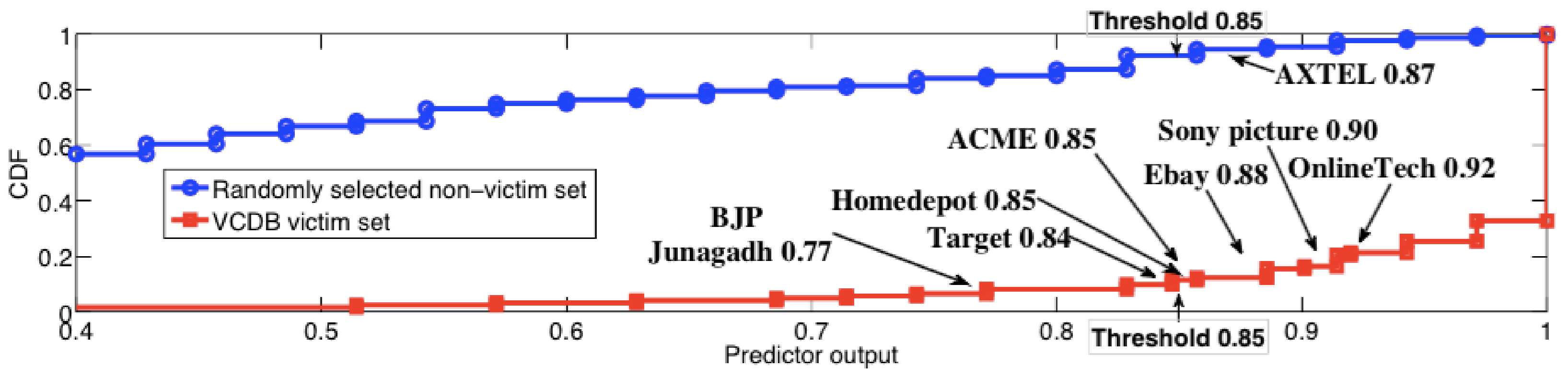
- Overall mismanagement features that are the most directly related with human factors have the largest normalized weight.
- *This confirms the intuitive understanding that the human element is the most important factor in cyber security.*



- Mismanagement features by themselves are not sufficiently good predictors;

BUT

- *In combination with other features, such as malicious activities, they add the MOST value.*



- **65% of incidents in blind-test dataset were predicted as 100% chance of breach.**
- Using a threshold of 0.85, we predict 92% of breaches.

SUMMARY

- **It is possible to *statistically predict* cyber security incidents on the basis of historical incidents and pre-incident security posture data. For example, we have shown we can predict 92% of the 2014 Verizon Data Breach Investigation Report breaches.**
- Difference between *detection* and *prediction* is key: one relies on signatures, while the other looks at patterns and trends in data that might appear to be unrelated.
- Security posture is many-dimensional and requires data from many parts of an organization, including Web applications, network configurations, and DNS.
- Protecting against data breaches requires fighting a battle on many fronts, and the key almost always is people.

FUTURE EFFORTS

- While these features and models may retain predictive power in the Energy Delivery System domain, we anticipate that the unique features of this domain will require new measurement methodologies; additional, at-scale, and in-practice measurements; the identification of new predictive features; and new models and classifiers for prediction.

CYBER RESILIENT ENERGY DELIVERY CONSORTIUM | CRED-C.ORG

FUNDING SUPPORT PROVIDED BY THE DEPARTMENT OF ENERGY, OFFICE OF ELECTRICITY DELIVERY & ENERGY RELIABILITY