Detecting Credential Spearphishing Attacks in Enterprise Settings

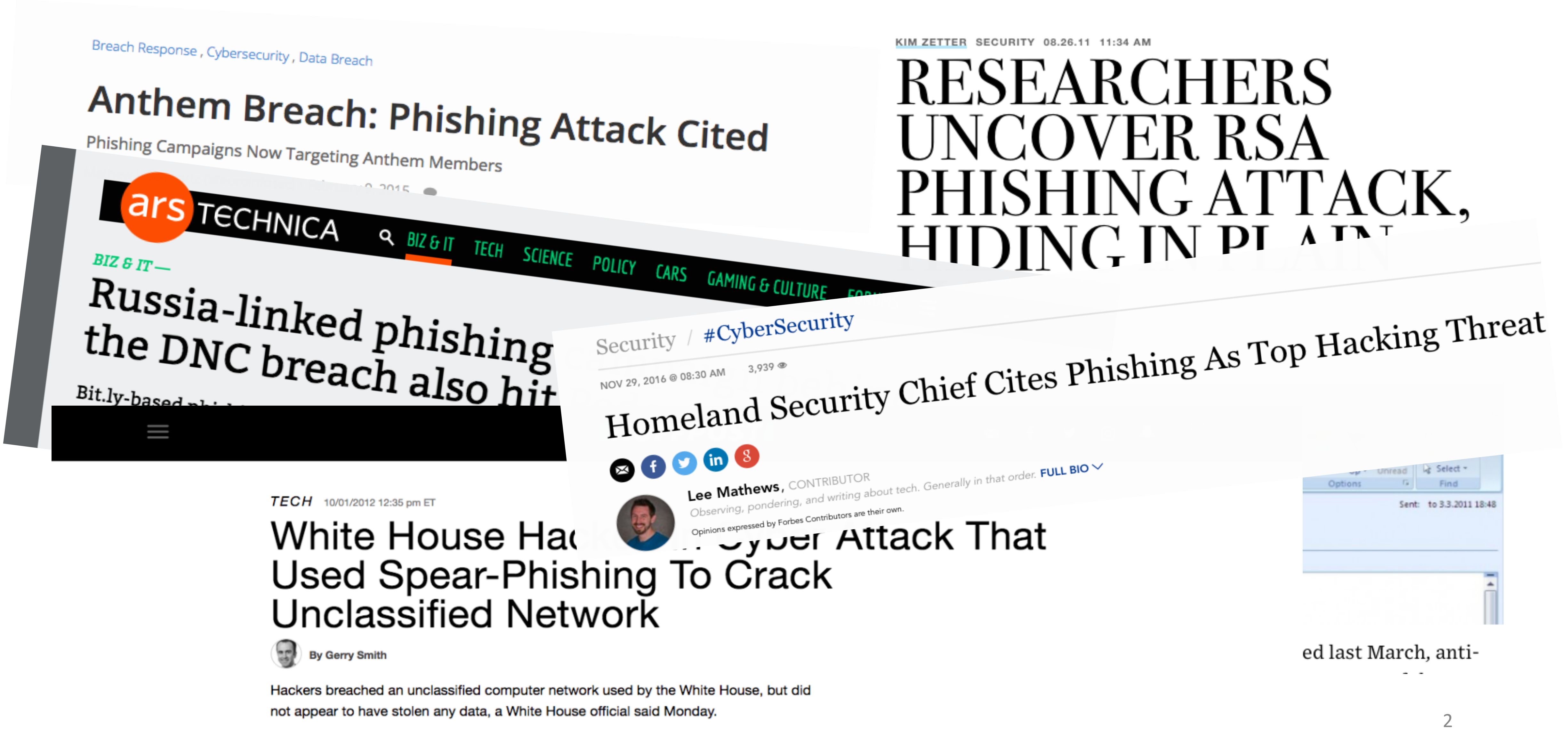
Grant Ho

UC Berkeley

Aashish Sharma, Mobin Javed, Vern Paxson, David Wagner

Spear Phishing

Targeted email that ***tricks*** victim into giving attacker ***privileged capabilities***



Breach Response , Cybersecurity , Data Breach

Anthem Breach: Phishing Attack Cited

Phishing Campaigns Now Targeting Anthem Members

KIM ZETTER SECURITY 08.26.11 11:34 AM

RESEARCHERS UNCOVER RSA PHISHING ATTACK, HIDING IN PLAIN

BIZ & IT —

Russia-linked phishing the DNC breach also hit

Bit.ly-based

Security / #CyberSecurity

NOV 29, 2016 @ 08:30 AM 3,939

Homeland Security Chief Cites Phishing As Top Hacking Threat

Lee Mathews, CONTRIBUTOR

Observing, pondering, and writing about tech. Generally in that order. FULL BIO

Opinions expressed by Forbes Contributors are their own.

TECH 10/01/2012 12:35 pm ET

White House Hac Cyber Attack That Used Spear-Phishing To Crack Unclassified Network

By Gerry Smith

Hackers breached an unclassified computer network used by the White House, but did not appear to have stolen any data, a White House official said Monday.

Sent: to 3.3.2011 18:48

ed last March, anti-

Our Focus: Enterprise *Credential* Spearphishing

“Credentials are king”

- Rob Joyce, Director of NSA’s Tailored Access Operations

- Wealth of access & lower barrier than 0-day malicious attachments
- What about 2FA?
  - Cost, usability , incomplete deployment, often still phish-able
- Detection today: user reporting, phish-able 2FA, post-mortem forensics

Our Work

Practical detection system for an enterprise’s security team

1. Extremely low FP burden (Goal: < ***minutes per day***)

2. Raises bar & detects many attacks, but ***not*** silver bullet

Our Work

Worked with the Lawrence Berkeley National Laboratory (LBL)

- US DoE National Lab w/ 5,000 employees

Anonymized datasets:

- SMTP header information (From and RCPT-TO headers)
- URLs in emails
- Network traffic logs
- LDAP logs

Key Challenges

1. Small set of labeled attack data
   - < 10 known successful credential spearphishing attacks

2. Base rate
   - **372 million** emails over **4 years** (Mar 2013 – Jan 2017)
   - Even detector w/ 99.9% accuracy = 372,000 alerts

Structure-Driven Features

Spearphishing Attack Taxonomy

- Successful spearphishing attacks have two necessary stages:

1. **The Lure**
    - Successful attacks ***lure***/convince victim to perform an action

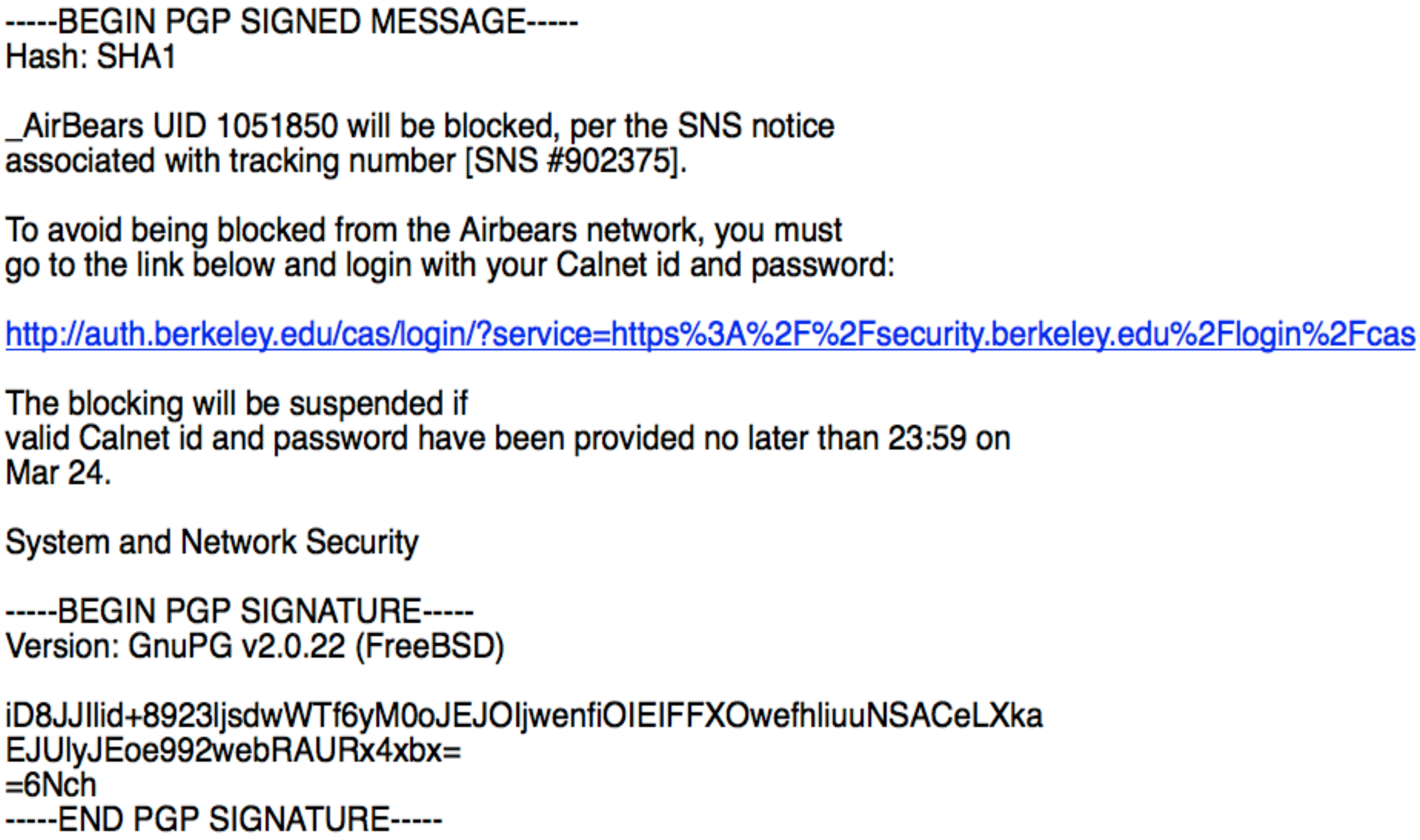
2. **The Exploit**
    - Successful attacks execute some ***exploit*** on behalf of the attacker
    - Malware, revealing credentials, wiring money to “corporate partner”

Spearphishing Attack Taxonomy

- Successful spearphishing attacks have two necessary stages:

1. **The Lure**
    - Successful attacks ***lure***/convince victim to perform an action

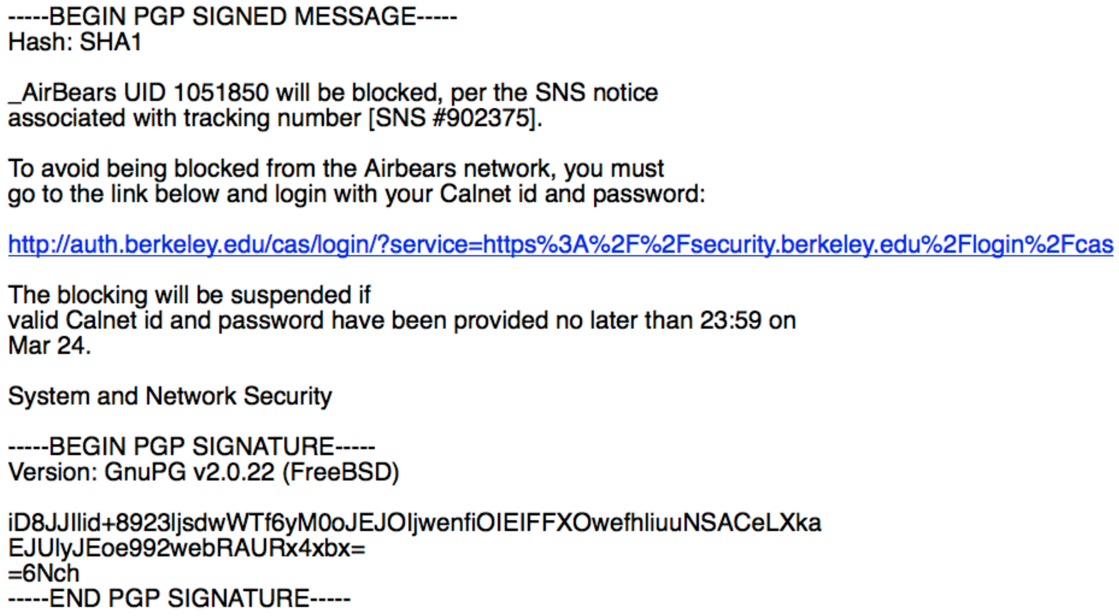
2. **The Exploit**
    - Successful attacks execute some ***exploit*** on behalf of the attacker
    - Malware, **revealing credentials**, wiring money to “corporate partner”



```
-----BEGIN PGP SIGNED MESSAGE-----
Hash: SHA1

_AirBears UID 1051850 will be blocked, per the SNS notice
associated with tracking number [SNS #902375].

To avoid being blocked from the Airbears network, you must
go to the link below and login with your Calnet id and password:

http://auth.berkeley.edu/cas/login/?service=https%3A%2F%2Fsecurity.berkeley.edu%2Flogin%2Fcas

The blocking will be suspended if
valid Calnet id and password have been provided no later than 23:59 on
Mar 24.

System and Network Security

-----BEGIN PGP SIGNATURE-----
Version: GnuPG v2.0.22 (FreeBSD)

iD8JJIlid+8923ljsdwWTf6yM0oJEJOIjwenfiOIEIFFXOwefhliuuNSACeLXka
EJUIyJEoe992webRAURx4xbx=
=6Nch
-----END PGP SIGNATURE-----
```

Modern Credential Spearphishing: The Lure

From: “Berkeley IT Staff” <security@berkeley.net>

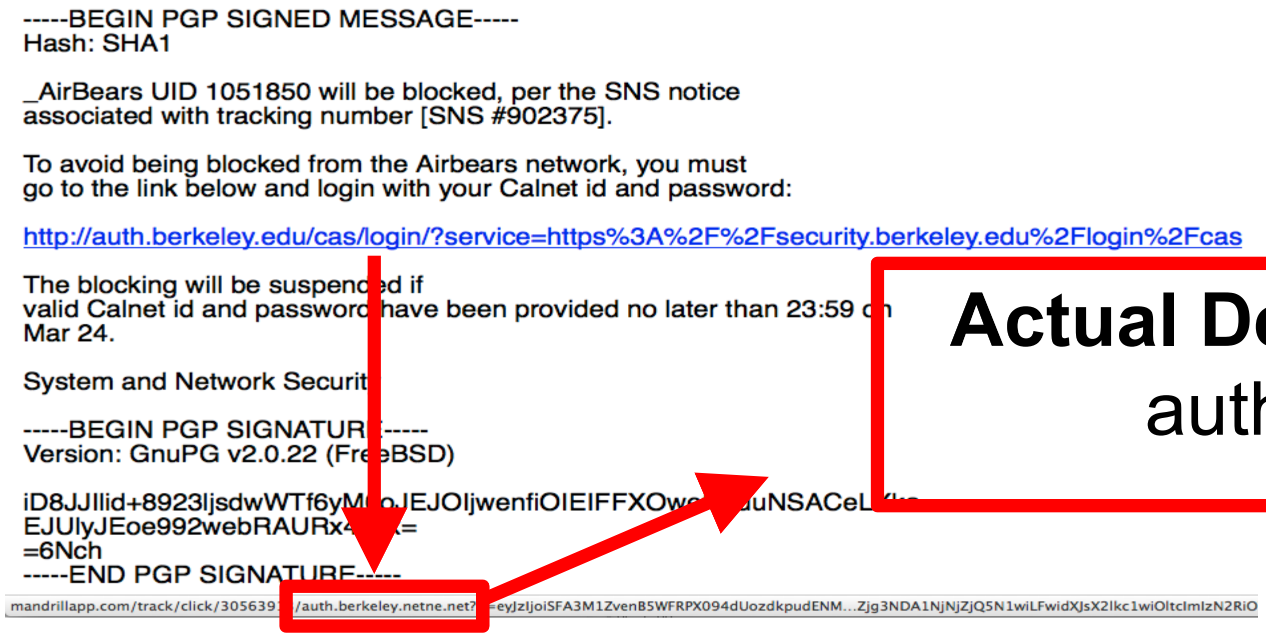


```
-----BEGIN PGP SIGNED MESSAGE-----
Hash: SHA1

_AirBears UID 1051850 will be blocked, per the SNS notice
associated with tracking number [SNS #902375].

To avoid being blocked from the Airbears network, you must
go to the link below and login with your Calnet id and password:

http://auth.berkeley.edu/cas/login/?service=https%3A%2F%2Fsecurity.berkeley.edu%2Flogin%2Fcas

The blocking will be suspended if
valid Calnet id and password have been provided no later than 23:59 on
Mar 24.

System and Network Security

-----BEGIN PGP SIGNATURE-----
Version: GnuPG v2.0.22 (FreeBSD)

iD8JJIlid+8923ljsdwWTf6yM0oJEJOIjwenfiOIEIFFXOwefhliuuNSACeLXka
EJUlyJEoe992webRAURx4xbx=
=6Nch
-----END PGP SIGNATURE-----
```

Lure

1. Attacker sends catchy email under ***trusted/authoritative identity***

Modern Credential Spearphishing: The Exploit



```
-----BEGIN PGP SIGNED MESSAGE-----
Hash: SHA1

_AirBears UID 1051850 will be blocked, per the SNS notice
associated with tracking number [SNS #902375].

To avoid being blocked from the Airbears network, you must
go to the link below and login with your Calnet id and password:

http://auth.berkeley.edu/cas/login/?service=https%3A%2F%2Fsecurity.berkeley.edu%2Flogin%2Fcas

The blocking will be suspended if
valid Calnet id and password have been provided no later than 23:59 on
Mar 24.

System and Network Security

-----BEGIN PGP SIGNATURE-----
Version: GnuPG v2.0.22 (FreeBSD)

iD8JJIlid+8923ljsdwWTf6yM oJEJOljwenfiOIEIFFXOw uNSACeL
EJUlyJEoe992webRAURx4 =
=6Nch
-----END PGP SIGNATURE-----
mandrillapp.com/track/click/305639 /auth.berkeley.netne.net? =eyJzIjoiSFA3M1ZvenB5WFRPX094dUozdkpudENM...Zjg3NDA1NjNjZjQ5N1wiLFwidXJsX2lkc1wiOltcImIzN2RiO
```

Actual Destination for linked text:
auth.berkeley.**netne.net**

Exploit

1. Victim ***clicks on embedded link***
2. Victim arrives at phishing website & submits credentials

Lure Features: Suspicious Sender Present

- Common lure: impersonate a trusted or authoritative entity

- Four “impersonation” classes - each has own set of *lure* features
  1. Name spoofing attacker
  2. Address spoofing attacker
  3. Previously unseen attacker
  4. Lateral attacker

- This talk: ***lateral attackers***

Lure Features (Cont.): Suspicious Sender Present

- Lateral spearphishing lure: attacker compromises trusted entity’s account

- Feature intuition: email = suspicious if employee sent it during a suspicious login session

- **Lure** features for **lateral spearphishing**:
  - was email sent in a session where sender logged in w/ new IP address?
  - # prior logins by the sender from the geolocated city of login IP addr
  - # of other employees who’ve also logged in from city of login IP addr

Exploit Features: Suspicious Action Occurred

- Winnow pool of candidate alerts to:

  Emails where recipient clicked on embedded URL (a ***click-in-email*** action)

- **Exploit** features: URL's **Fully-qualified domain** (hostname) is suspicious
  - # of prior visits to FQDN across all enterprise's network traffic
  - # of days between 1st employee's visit to FQDN & current email's arrival

Using Features for Detection

How do we leverage our features?

- Combine lure + exploit features to get FVs for emails

- How do we use these features for detecting attacks?

Approach 1: Manual rules

- Problems: soundly choosing thresholds & generalizability

Approach 2: Supervised ML

- Problems: tiny # of labeled attacks and base rate

Limitations of Standard Techniques

Approach 3: Unsupervised learning/anomaly detection

- Clustering/Distance Based: kNN
- Density-based: KDE, GMM
- Many others...

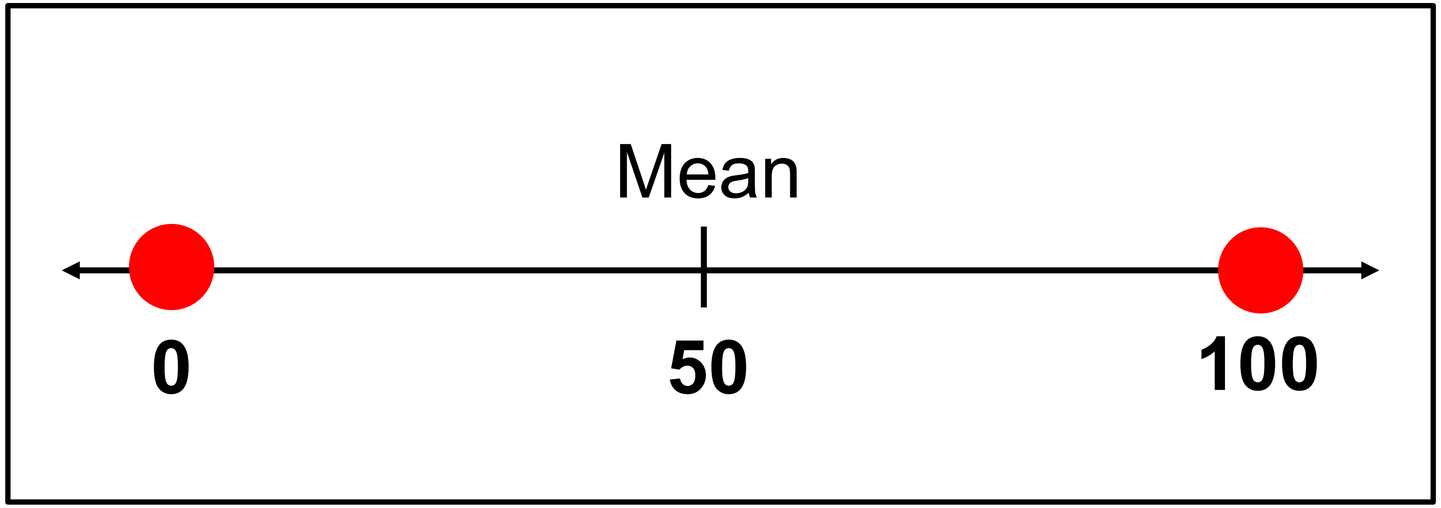
Three common problems:

1. Require hyperparameter tuning

Classical Anomaly Detection: Limitations

Three thematic problems:

1. Parametric and/or hyperparameter tuning
2. **Direction-agnostic** (standard dev of +3 just as anomalous as -3)

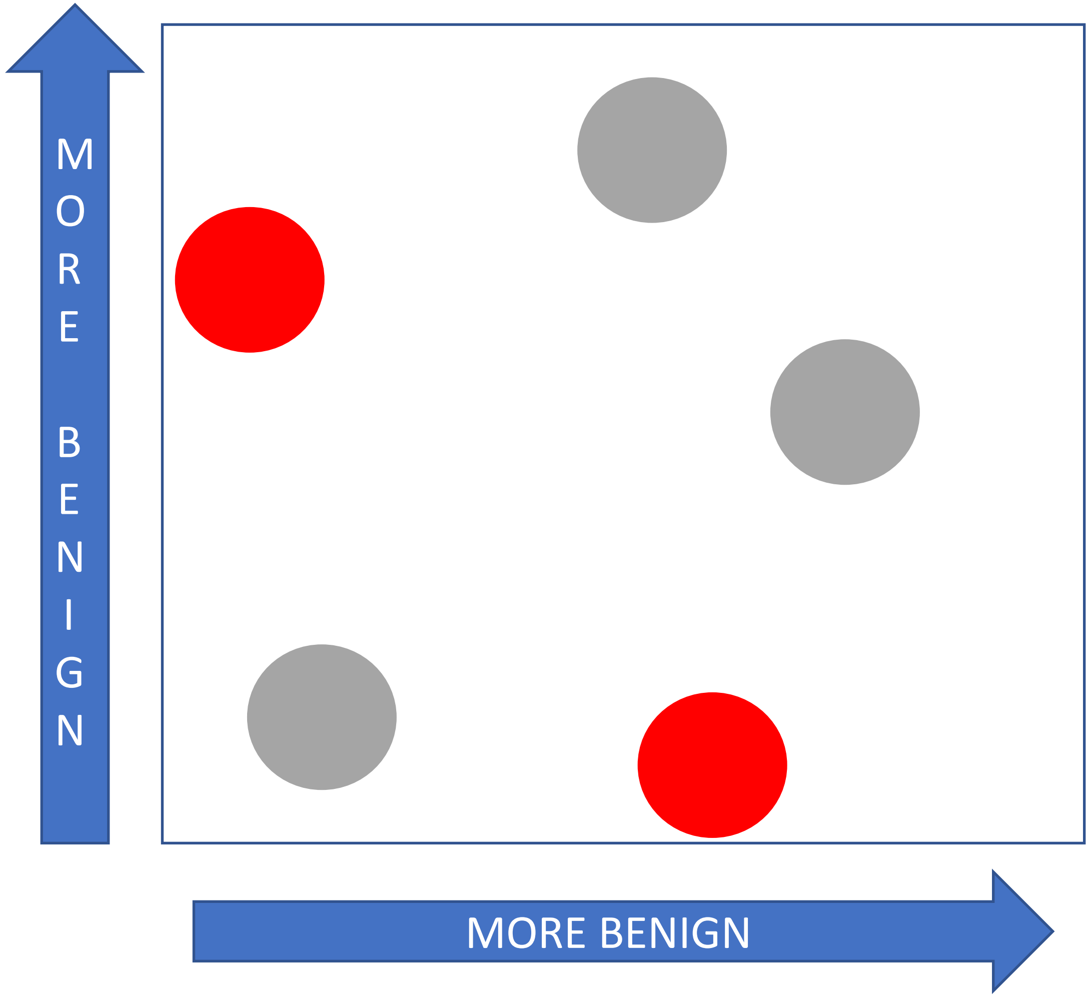


Feature:
# prior logins by current employee from city of new IP addr

Classical Anomaly Detection: Limitations

Three thematic problems:

1. Parametric and/or hyperparameter tuning
2. Direction-agnostic
3. **Alert if anomalous in only one dimension**

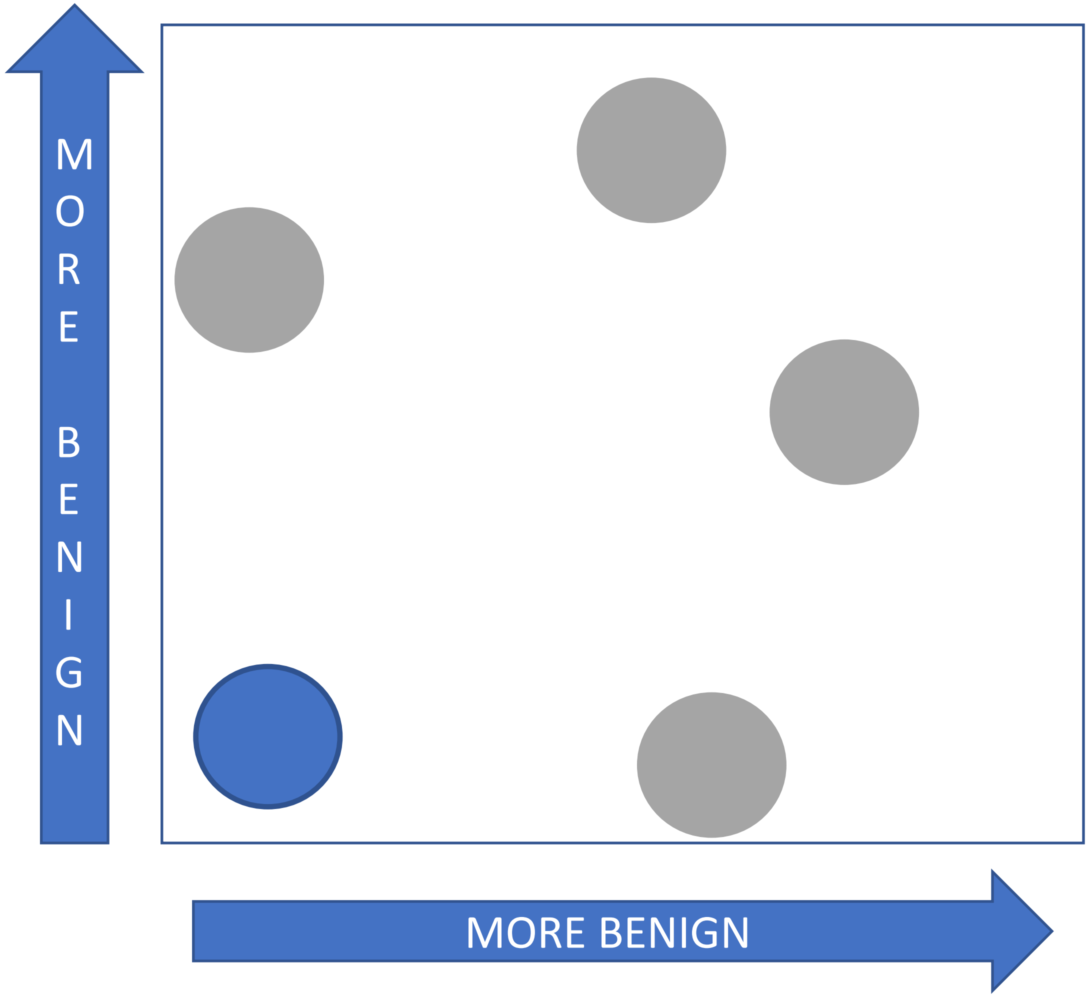


Classical Anomaly Detection: Limitations

Three thematic problems:

1. Parametric and/or hyperparameter tuning
2. Direction-agonistic
3. Alert if anomalous in only one dimension

- **DAS**: *simple*, new method that overcomes these 3 problems



DAS: Directed Anomaly Scoring

1. Security analysts w/ limited time: specify ***B*** = alert budget

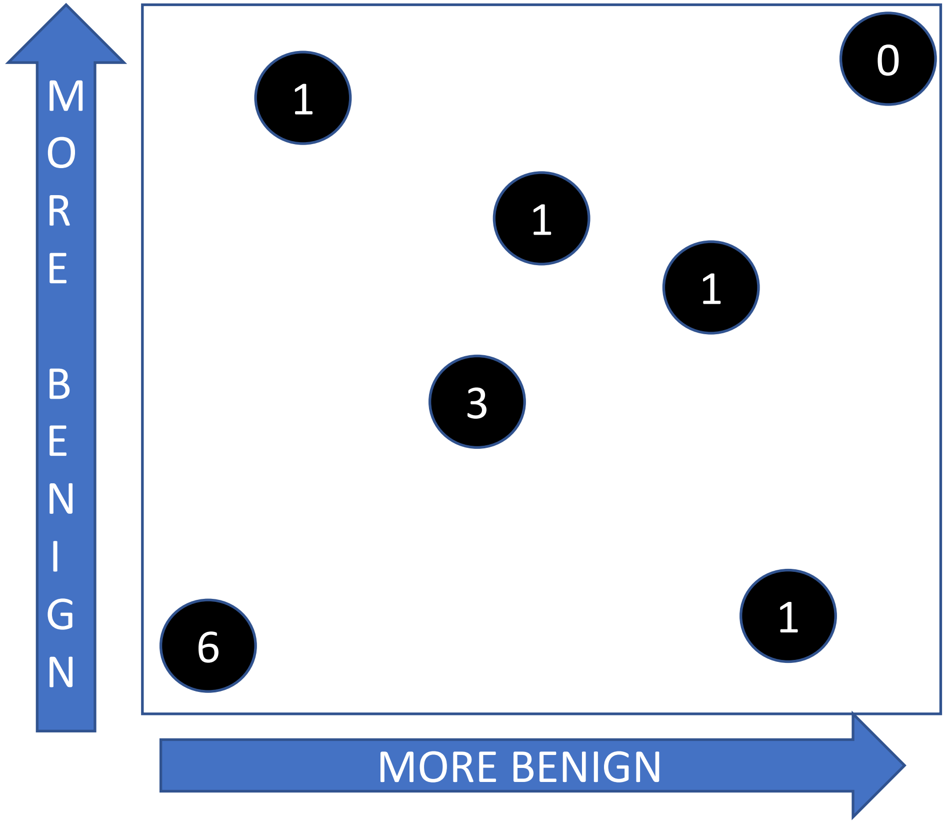
2. For set of events, assign each event a “suspiciousness” score

3. Rank events by their “suspiciousness”

4. Output the ***B*** most suspicious events for security team

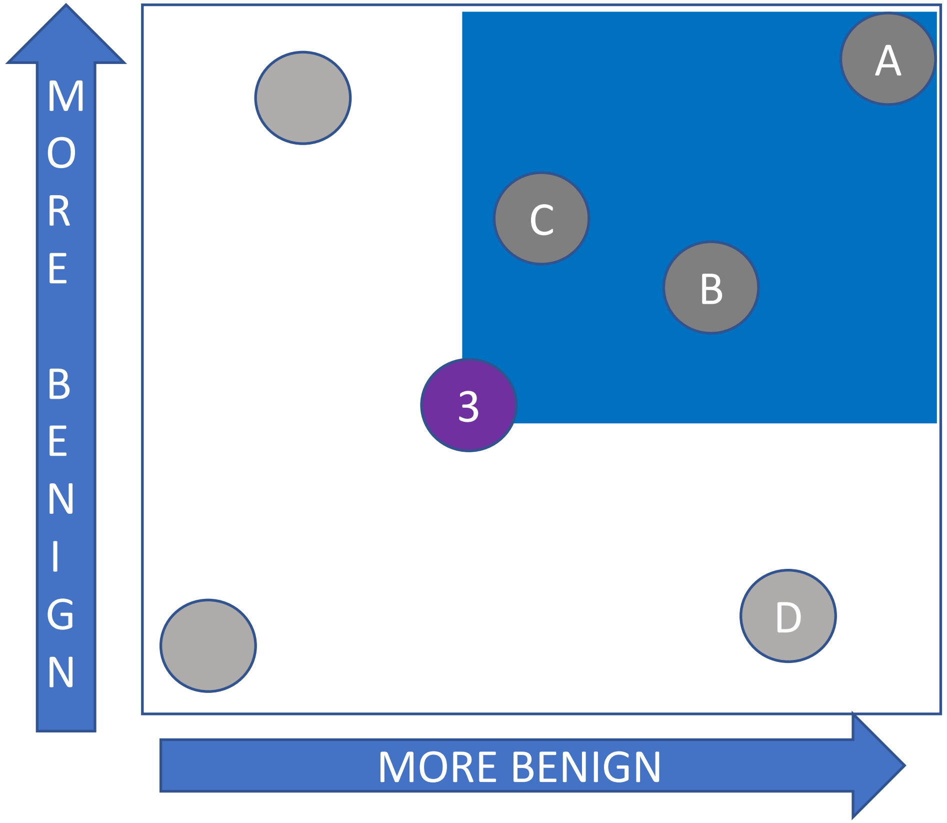
DAS: Directed Anomaly Scoring

- Score(Event X) = # of other events that are as **benign** as X in ***every*** dimension
  - i.e., Large score = many other events are more benign than X



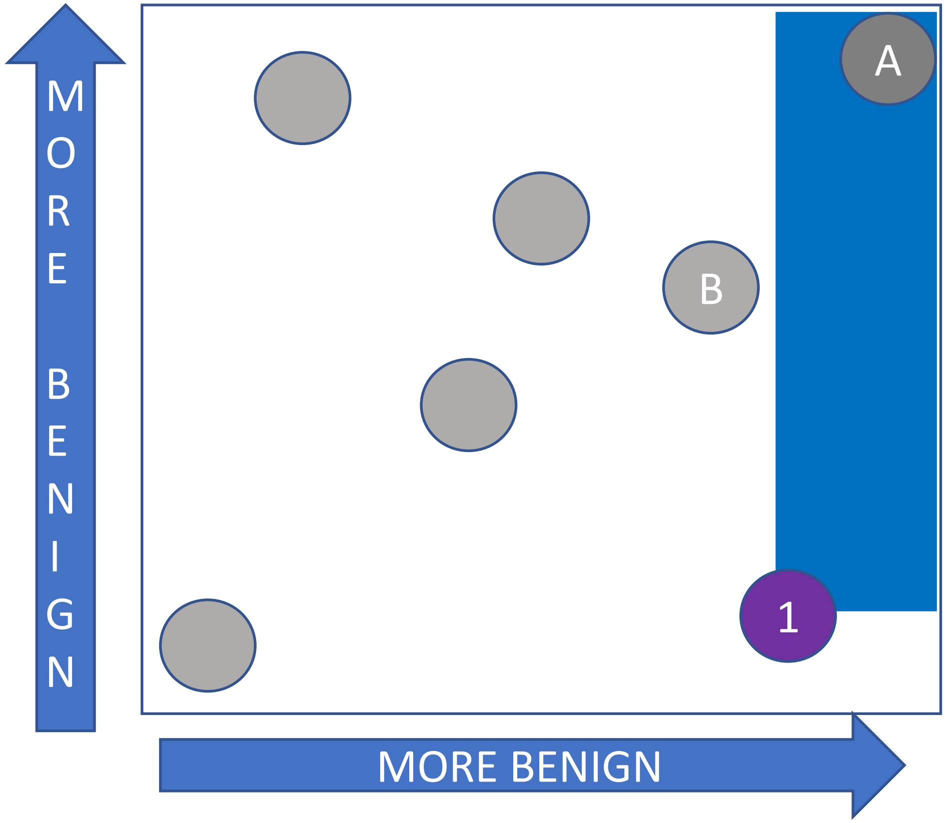
DAS: Directed Anomaly Scoring

- Score(Event X) = # of other events that are as **benign** as X in ***every*** dimension



DAS: Directed Anomaly Scoring

- Score(Event X) = # of other events that are as **benign** as X in ***every*** dimension



Detection Results

- Real-time detector on 370 million emails over ~4 years

- Ran detector w/ total budget of **10 alerts/day**
  - Practical for LBL’s security team (~240 alerts/day typical)

- Detected **17 / 19** spearphishing attacks (89% TP)
  - 2 / 17 detected attacks were ***previously undiscovered***

- Best classical anomaly detection: **4/19** attacks for same budget
  - Need budget >= **91 alerts/day** to detect same # of attacks as DAS

Results: Cost of False Positives

- **10 alarms / day:** How much time does this cost the security team?

- LBL's security staff manually investigated all our alerts
  - 24 alerts / minute (avg rate for one analyst)
  - **< 15 minutes** for 1 analyst to investigate alerts from **an entire month**

- Subject + URL + "From:" = quick semantic filter
  - "Never Lose Your Keys, Wallet, or Purse Again!"
  - "Invitation to Speak at Summit for Energy..."

Conclusion

- Real-time system for detecting credential spearphishing attacks
  - TP = 89%: detects known + previously undiscovered attacks
  - FP = 0.004%: 10 alerts / day (alerts processed in < minutes per day)

<u>Key ideas</u>

1. Leverage lure + exploit structure of spearphishing to design features
2. DAS: unsupervised, non-parametric technique for anomaly detection
   1. Generalizes beyond spearphishing
   2. “Needle-in-haystack” problems w/ curated & directional features

grantho@cs.berkeley.edu