# Detecting Stuffing of a User's Credentials at Her Own Accounts

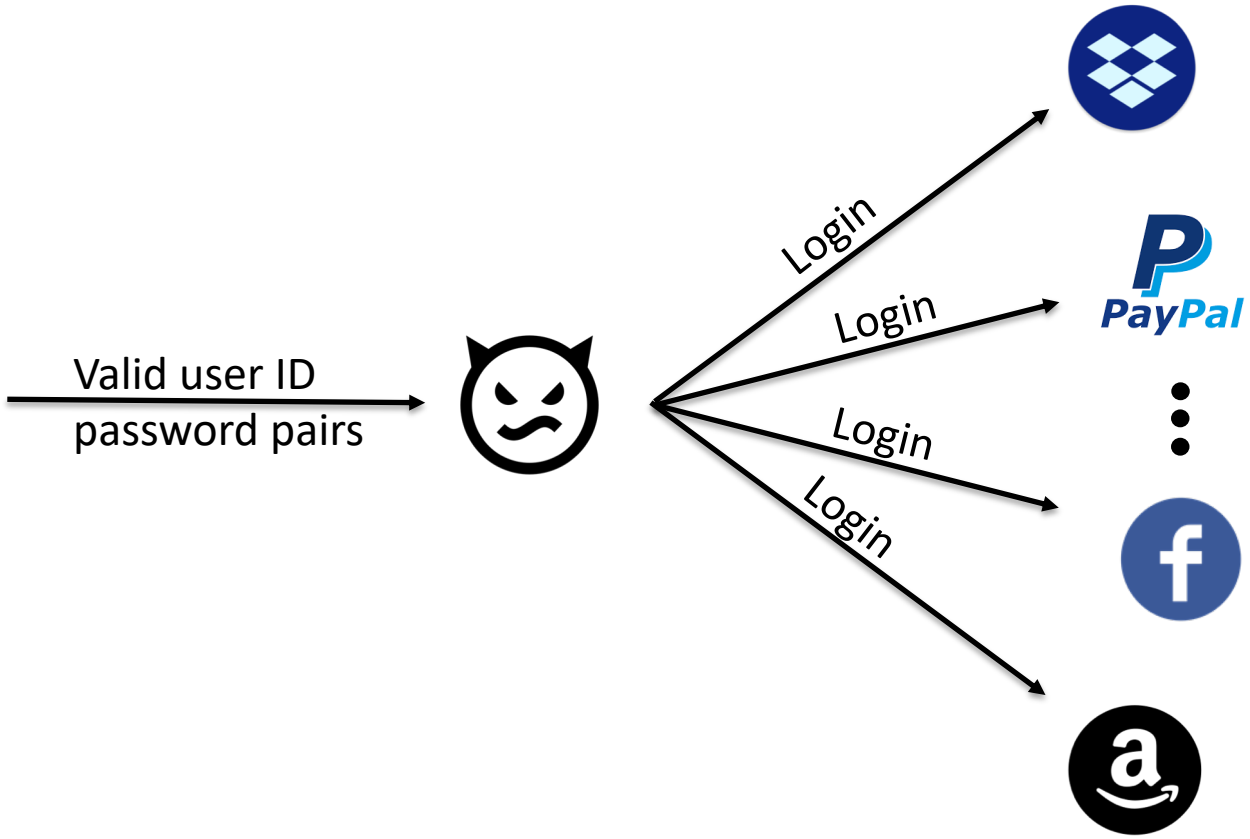
***Ke Coby Wang*** ***Michael K. Reiter***

*University of North Carolina at Chapel Hill*

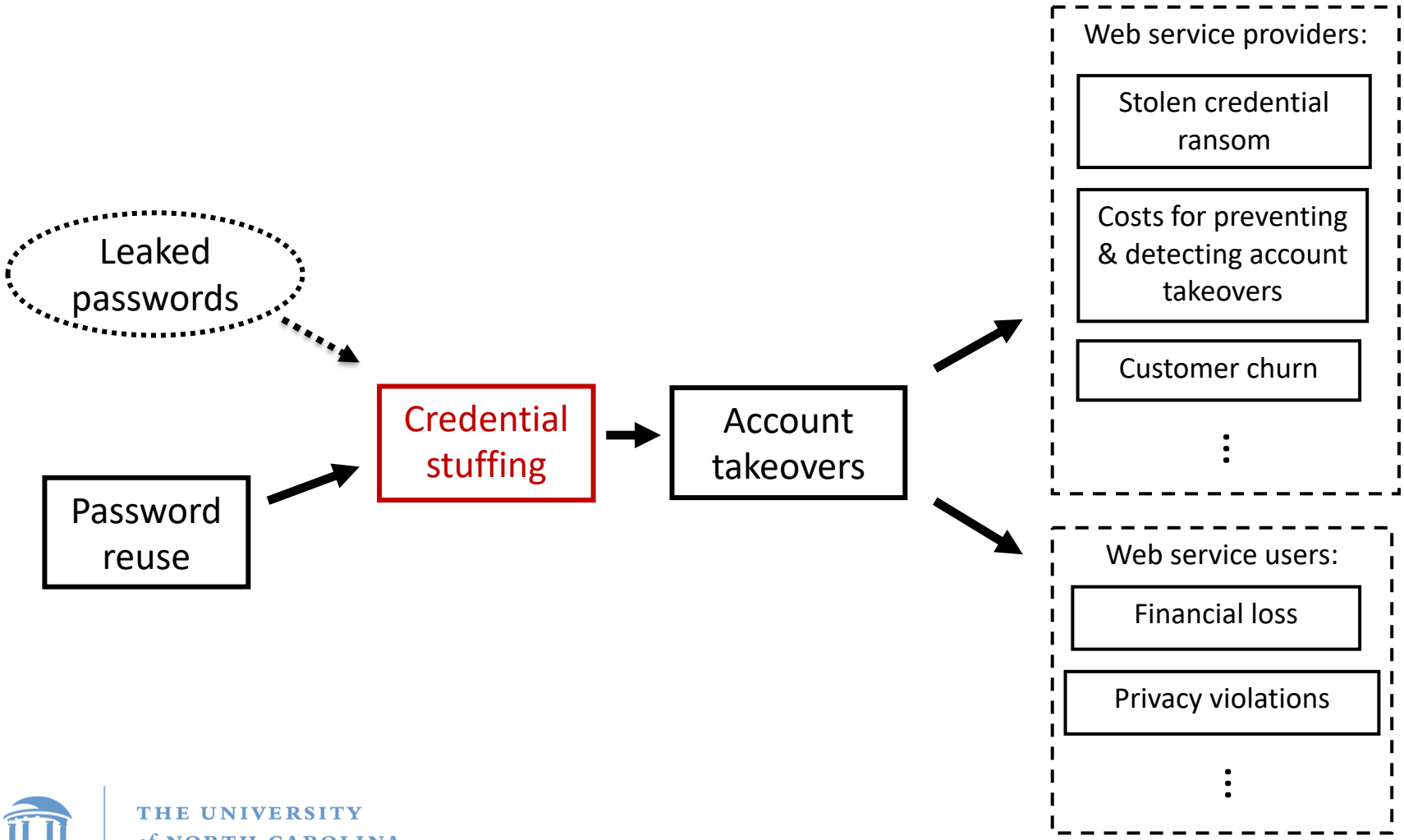
THE UNIVERSITY
*of* NORTH CAROLINA
*at* CHAPEL HILL

# Credential Stuffing

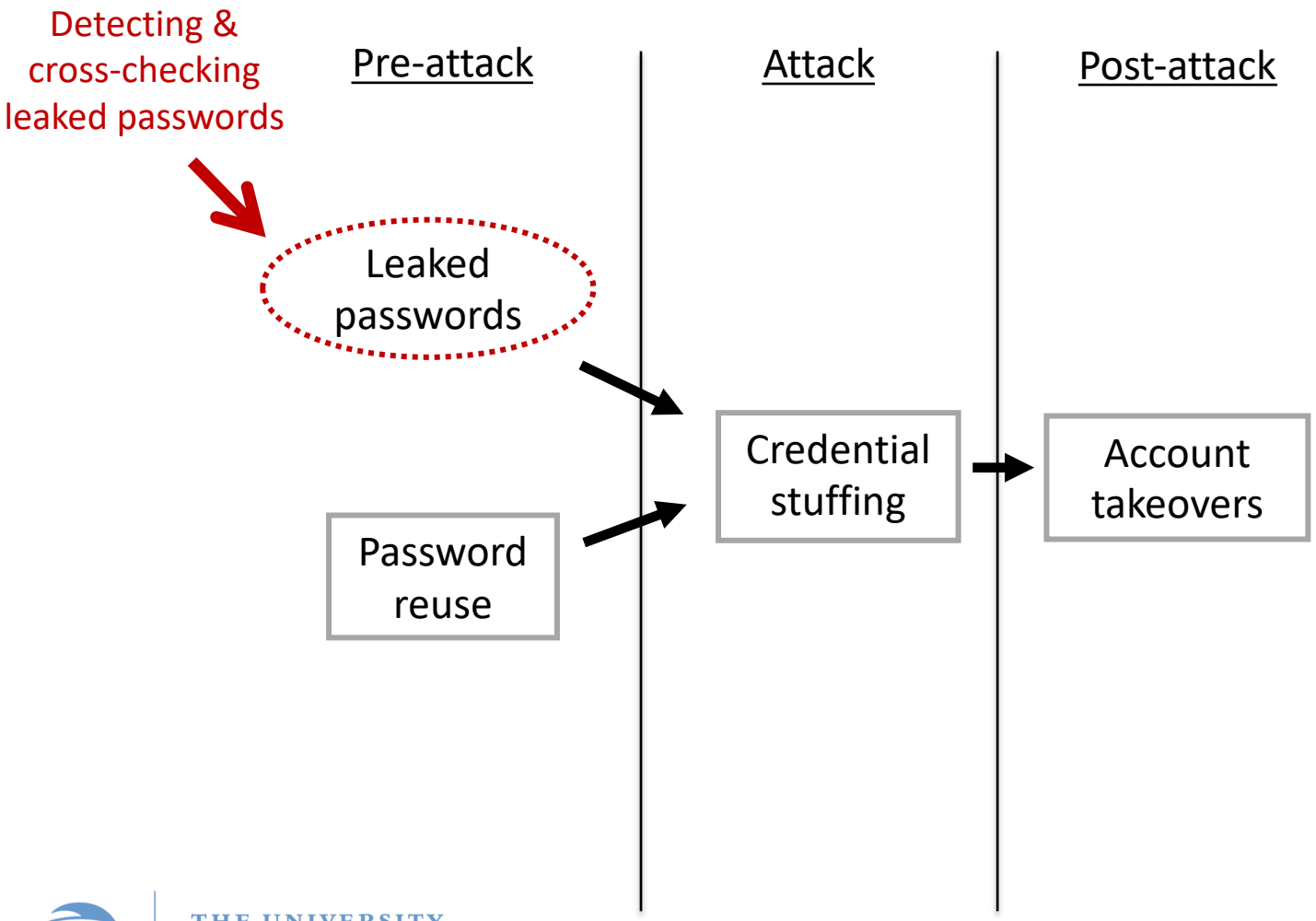
Database breaches, phishing, malware, social engineering, etc.



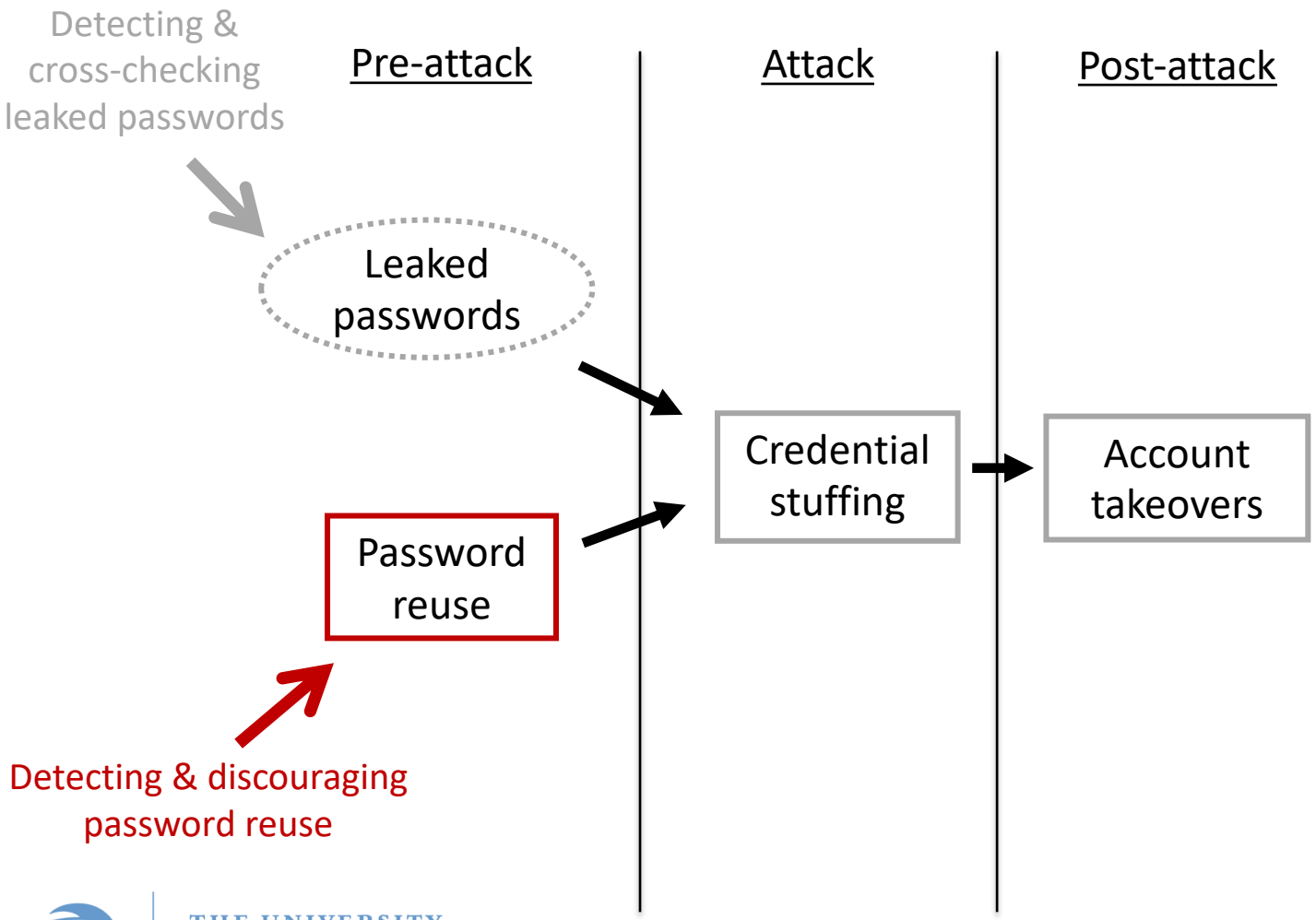
# Harm of Credential Stuffing



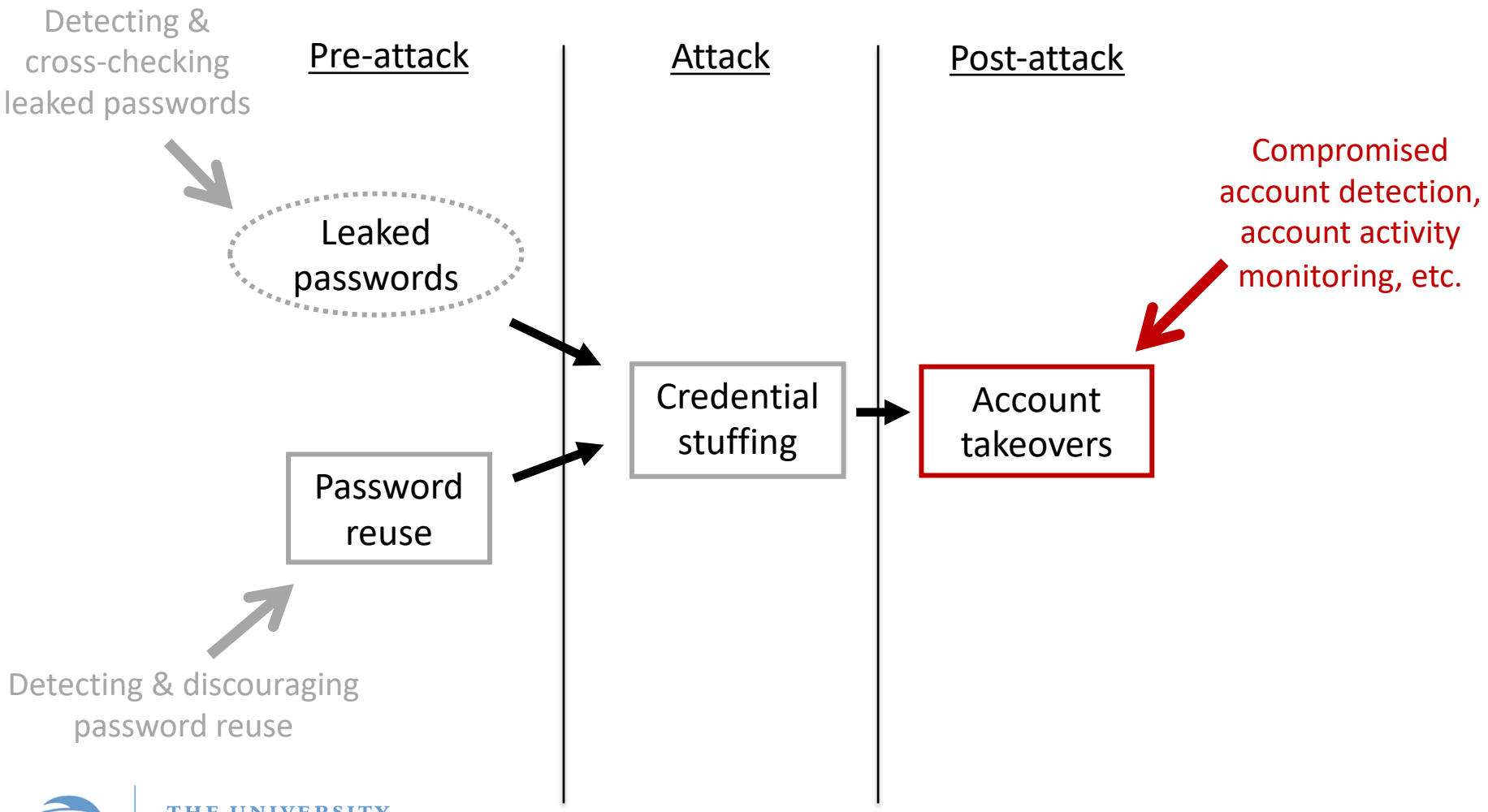
# Existing Approaches



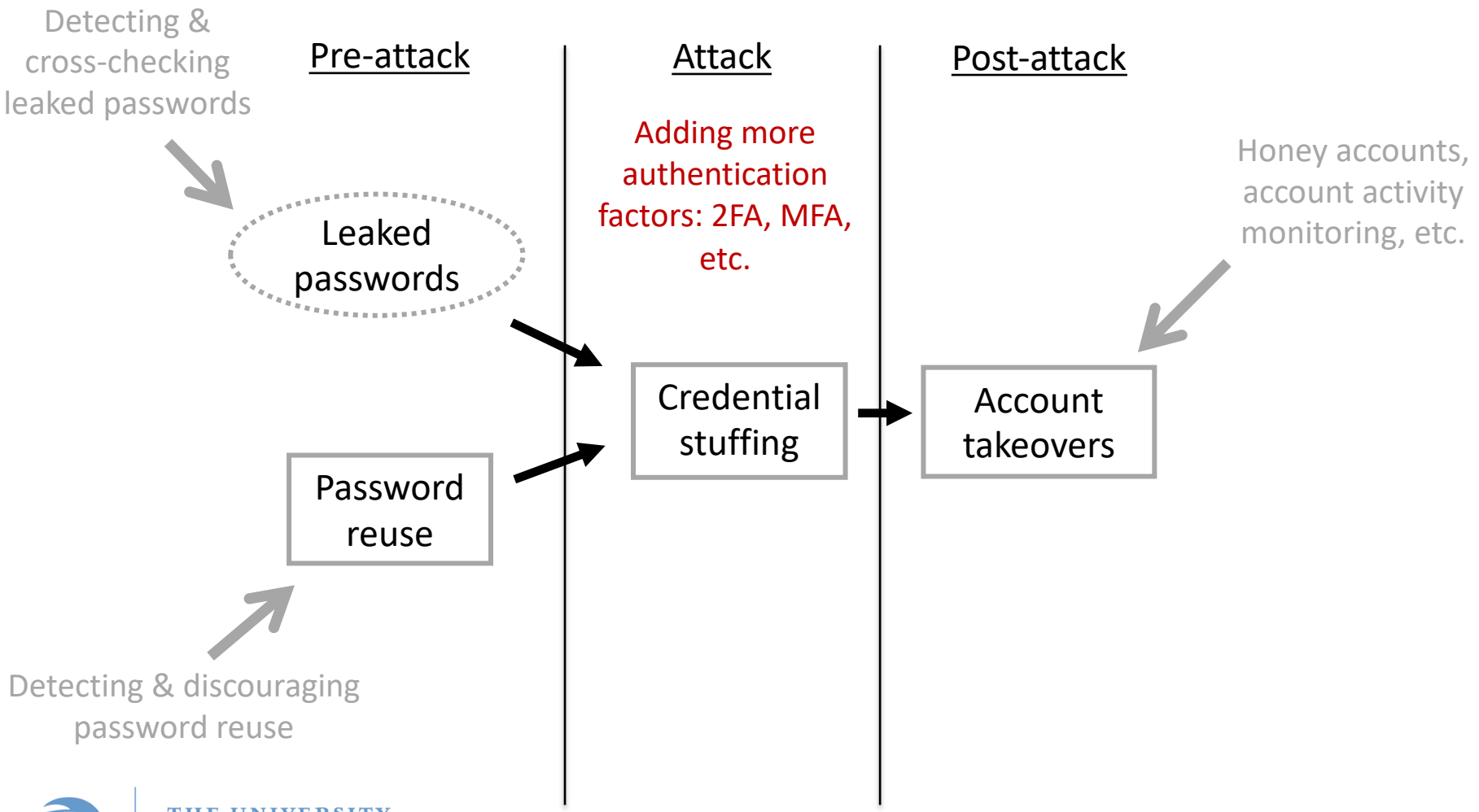
# Existing Approaches



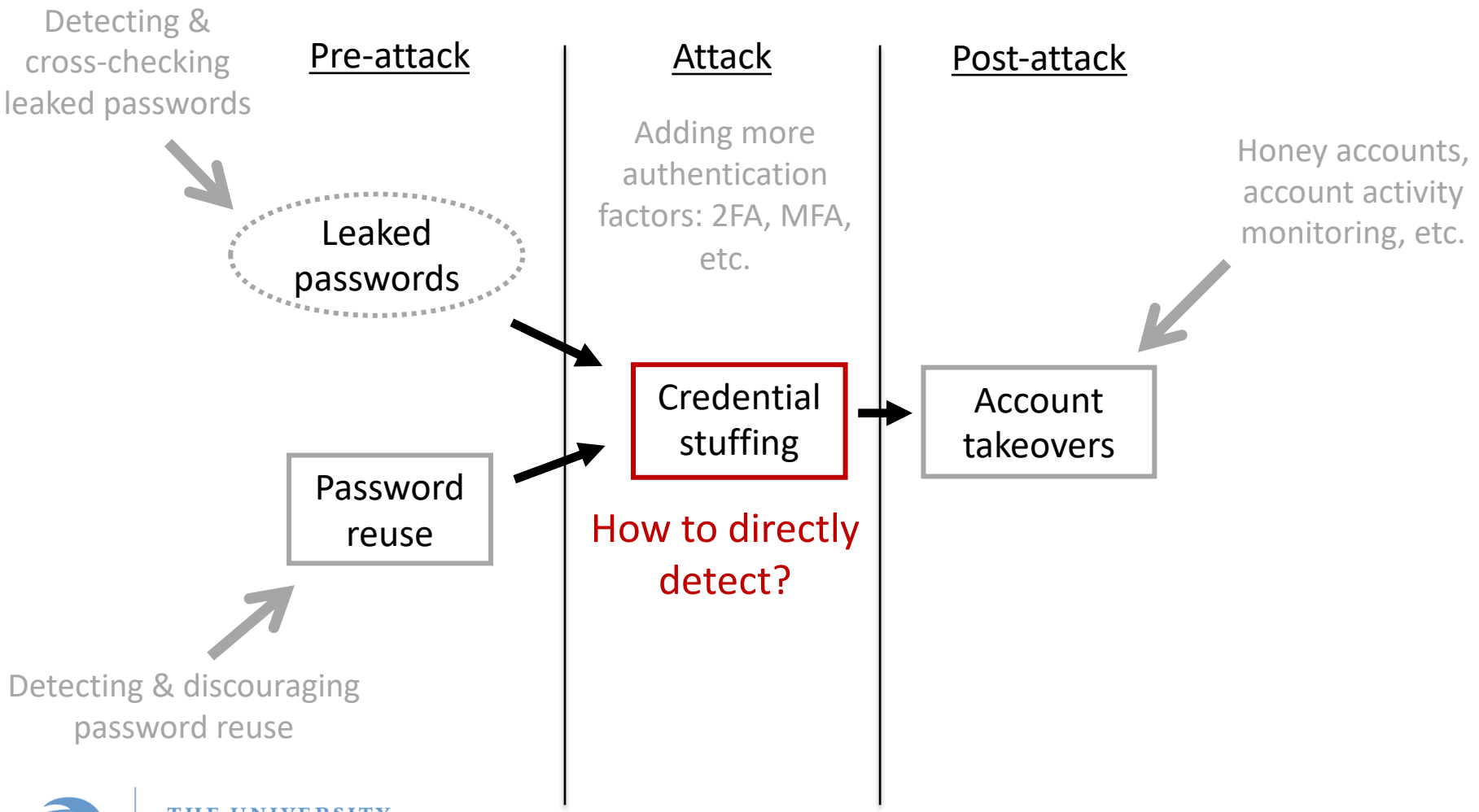
# Existing Approaches



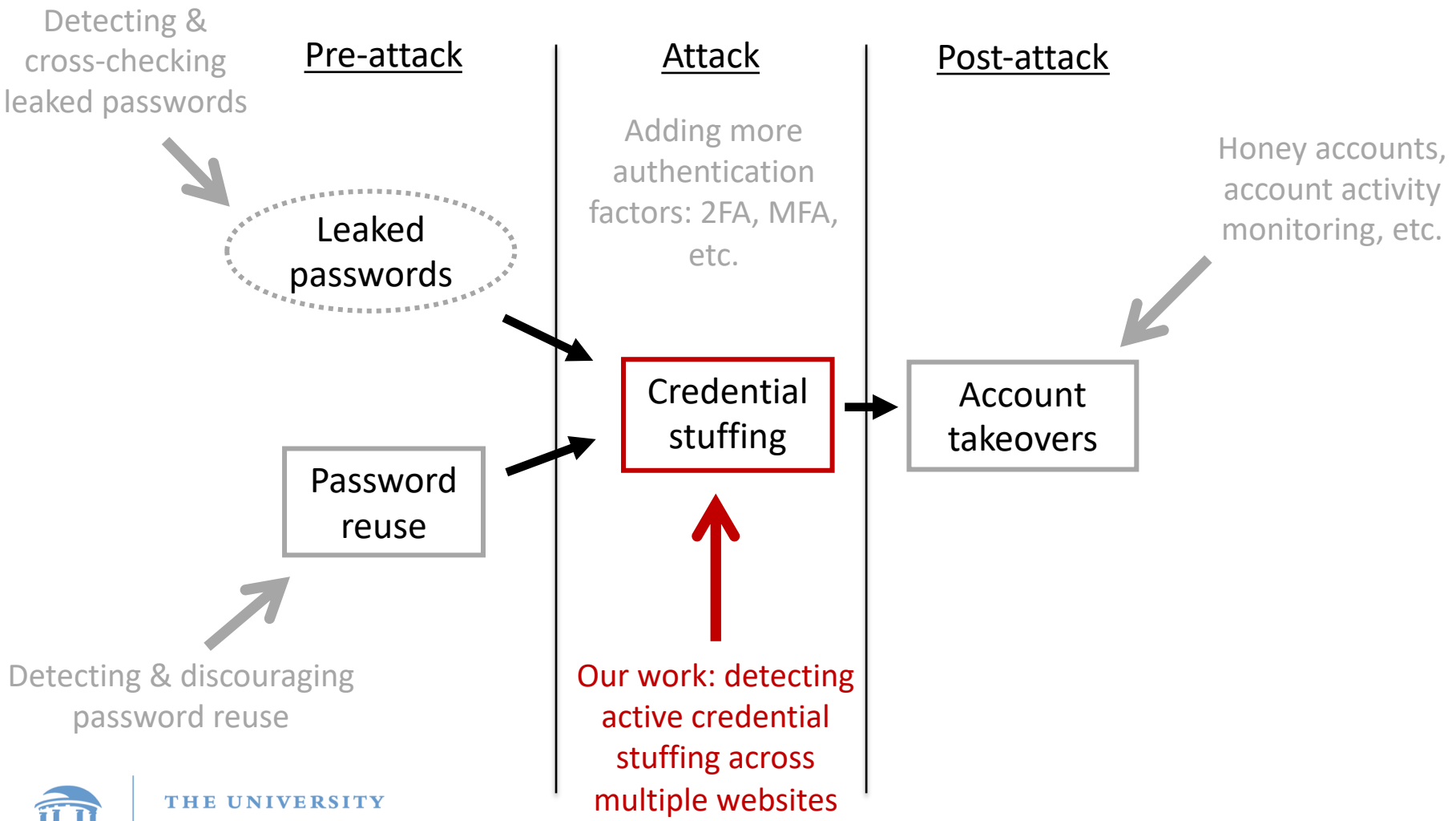
# Existing Approaches



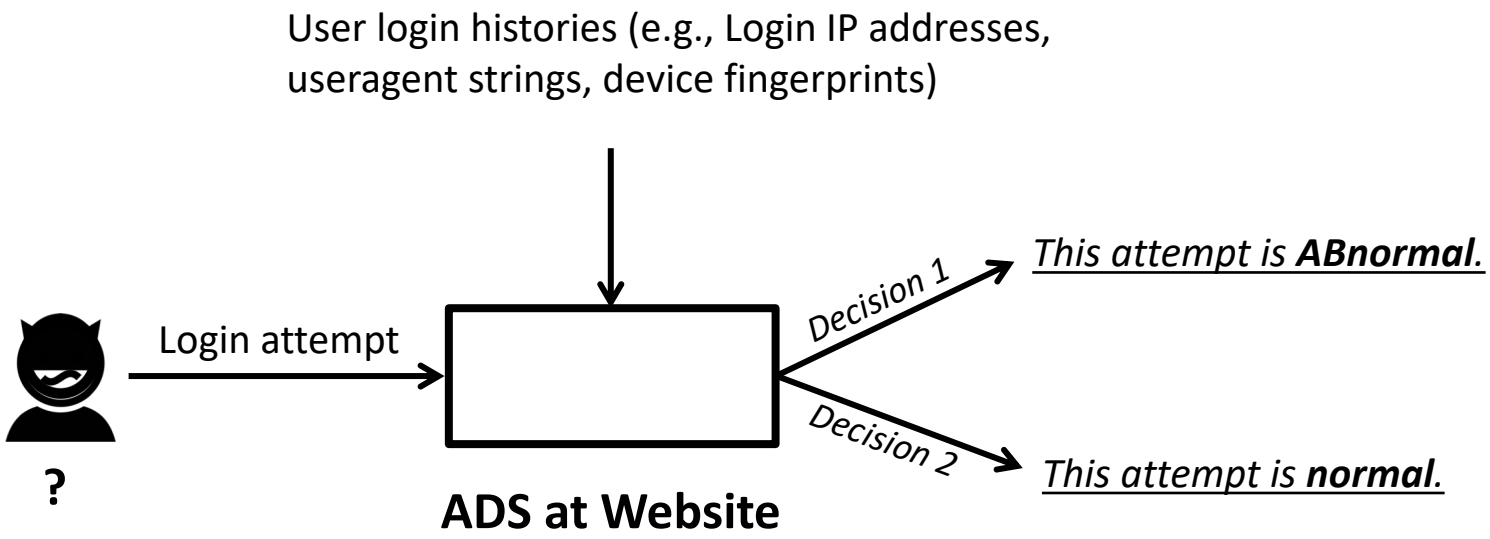
# Existing Approaches



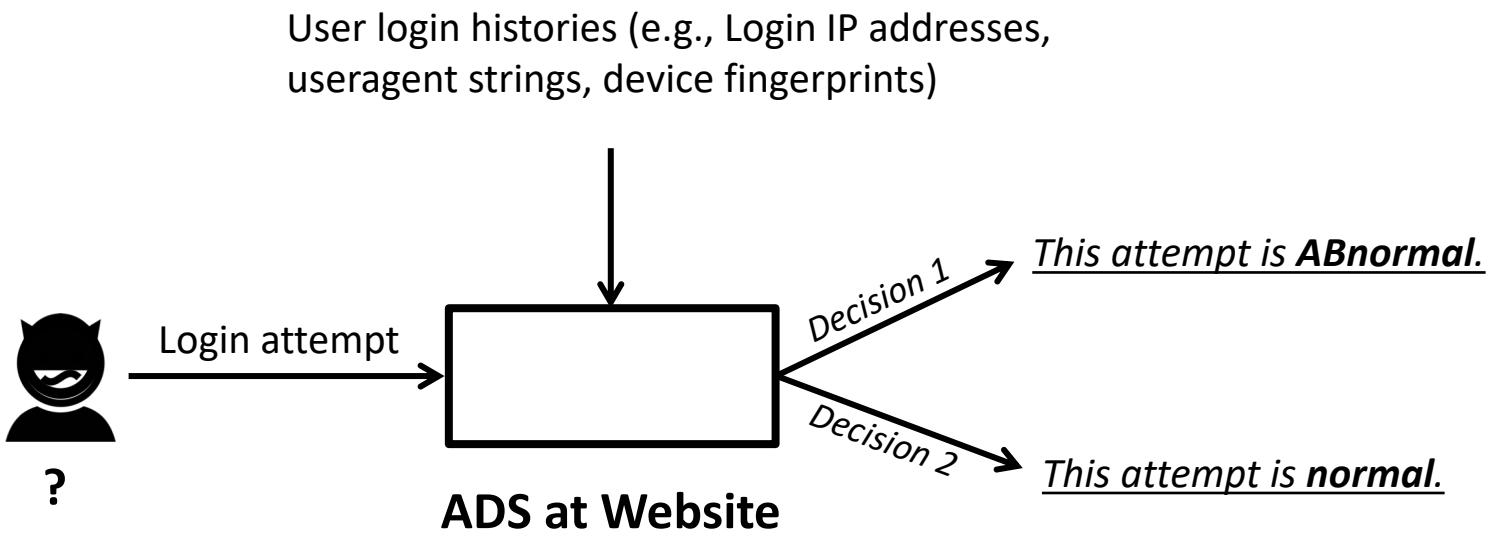
# Our Work


Detecting & cross-checking leaked passwords
Pre-attack
Attack
Post-attack
Adding more authentication factors: 2FA, MFA, etc.
Honey accounts, account activity monitoring, etc.
Leaked passwords
Credential stuffing
Account takeovers
Password reuse
Detecting & discouraging password reuse
Our work: detecting active credential stuffing across multiple websites

# Anomaly Detection Systems (ADS)



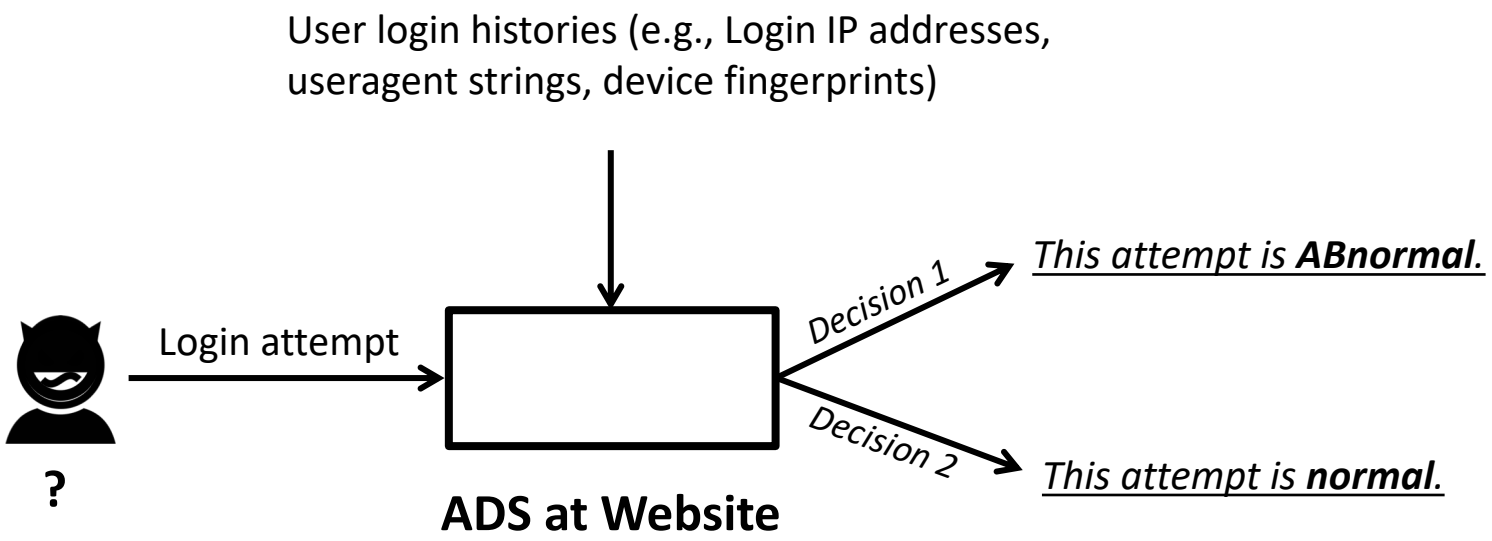
# Anomaly Detection Systems (ADS)




*Naïve ADS:*

- *Strange IPs = "abnormal"*
- *Strange devices = "abnormal"*

# Anomaly Detection Systems (ADS)



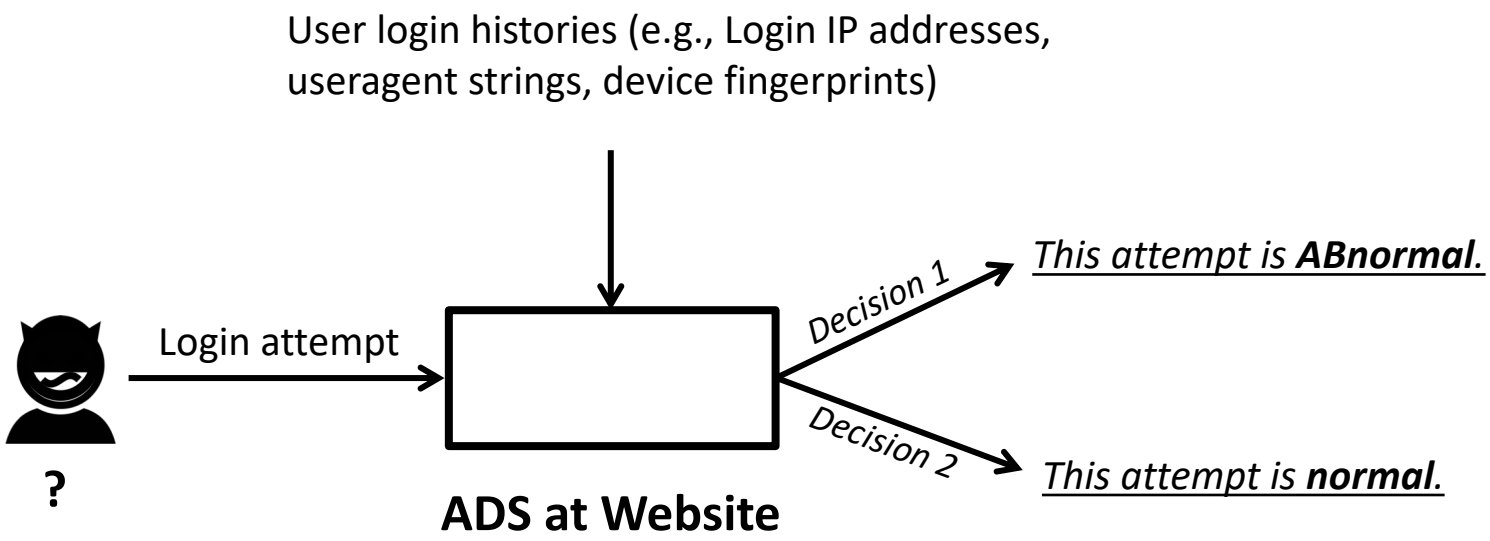

*More sophiscated ADS*:*

- *Multiple login features*
- *Attackers' different capability levels*

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*Freeman et al. [NDSS 2016]

# Anomaly Detection Systems (ADS)




*"Researching attacker"*:*

- *Hold users' correct passwords*
- *Try to access users' accounts from same countries of legitimate users*

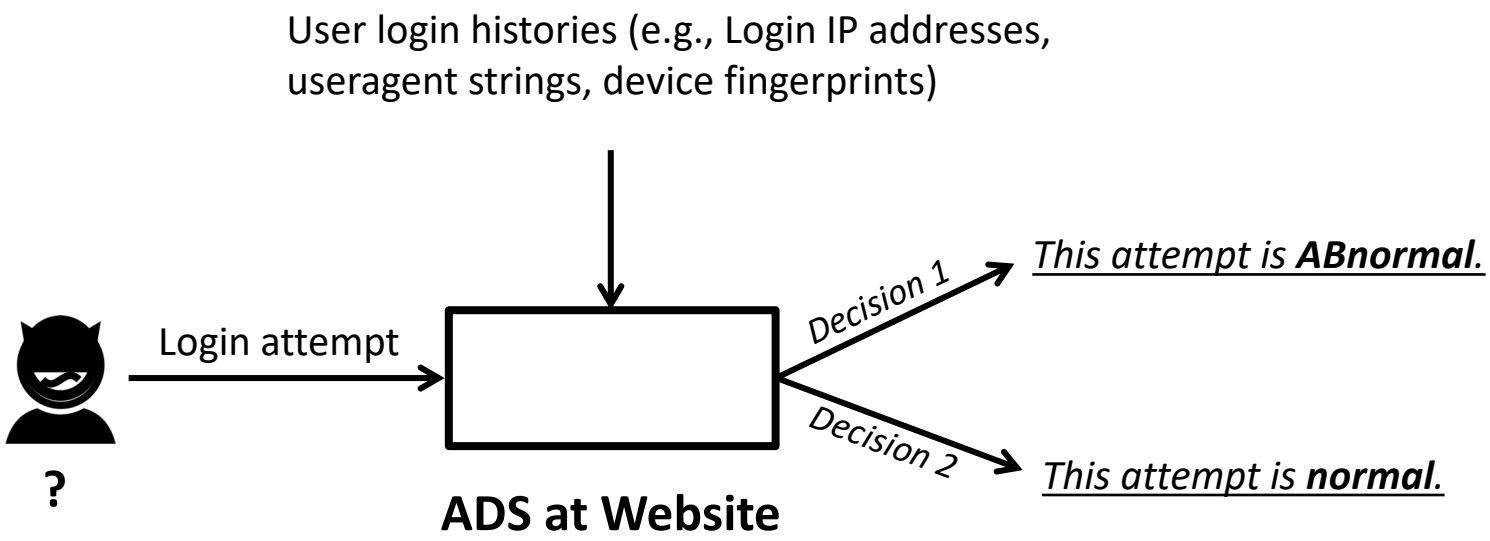
*"Phishing attacker"*:*

- *Hold users' correct passwords*
- *Try to access users' accounts from same countries with same browser user-agent strings of legitimate users*

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*Freeman et al. [NDSS 2016]

# Anomaly Detection Systems (ADS)




ADS:

- leverages users' login patterns (IPs, browser agentstrings, etc.)
- helps a website to distinguish malicious login attempts
- **NOT** an authentication factor that directly decides whether a login attempt is successful or not.

# Evidence Trail from Credential Stuffing

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

**Credential Stuffer**

Websites where Alice has accounts


alice@yyy.com : *alicepwd0*
**ADS**

2FA


alice@yyy.com : *alicepwd*
**ADS**
**2FA**


alice@yyy.com : *alicepwd*
**ADS**

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Evidence Trail from Credential Stuffing

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

Websites where Alice
has accounts

**Credential Stuffer**

Login attempt with *c*


***≠ alicepwd***

alice@yyy.com : *alicepwd0*
**ADS**: *abnormal*

2FA

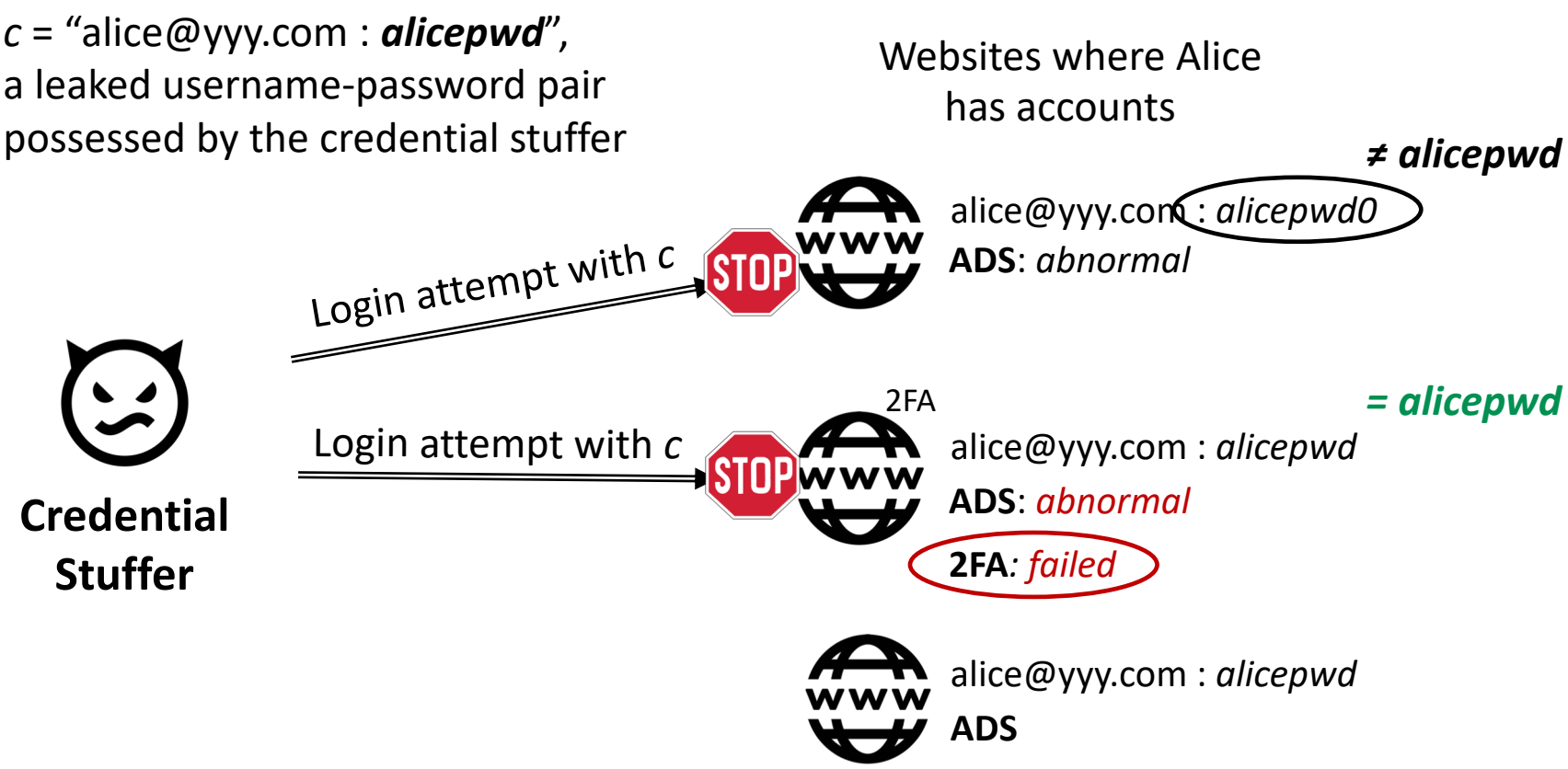
alice@yyy.com : *alicepwd*
**ADS**
**2FA**

alice@yyy.com : *alicepwd*
**ADS**

THE UNIVERSITY
*of* NORTH CAROLINA
*at* CHAPEL HILL

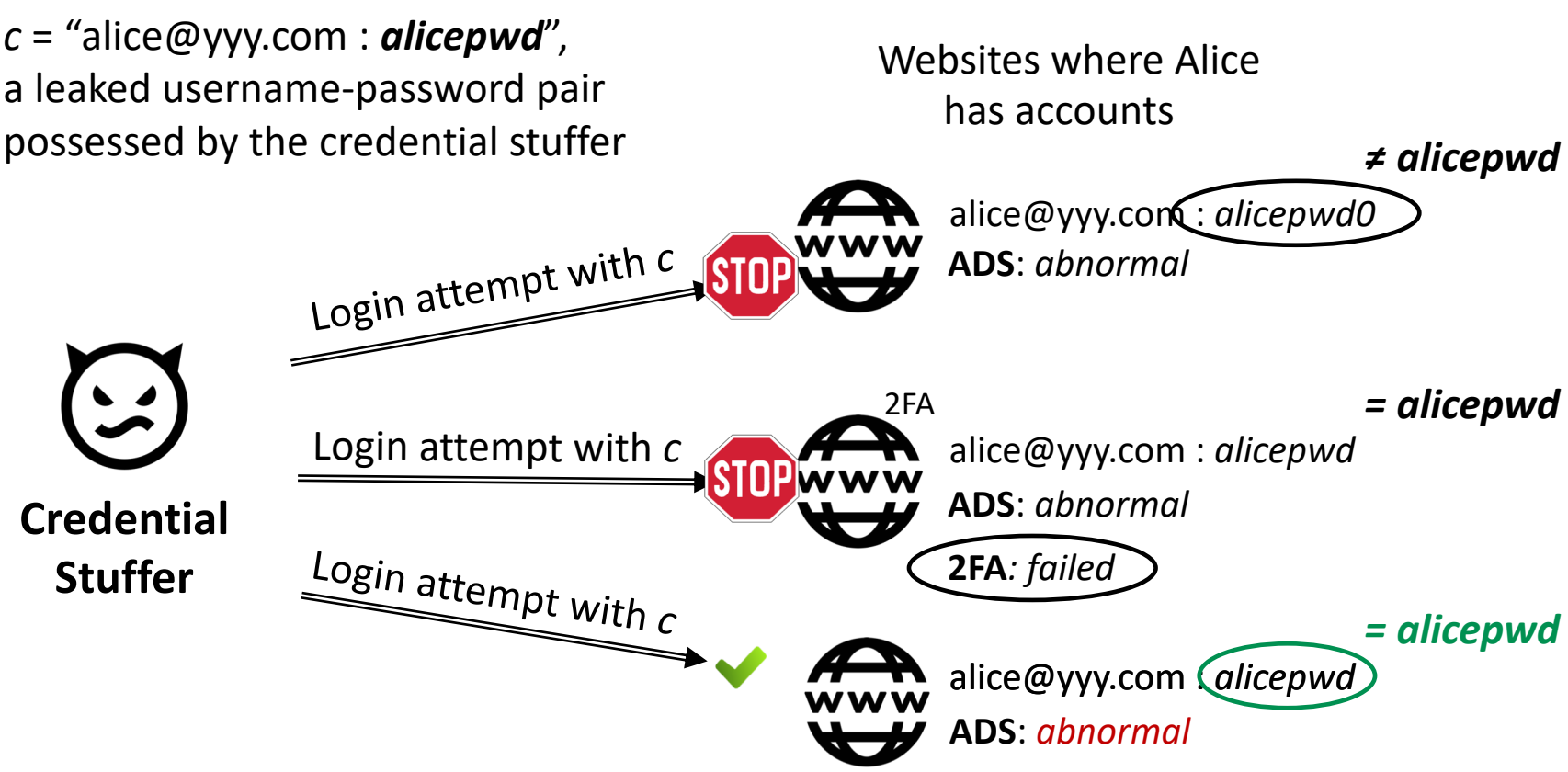
# Evidence Trail from Credential Stuffing

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer



# Evidence Trail from Credential Stuffing

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer



# Evidence Trail from Credential Stuffing

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

Websites where Alice
has accounts

***≠ alicepwd***

The "trail" left by credential stuffing attacks are those passwords submitted in abnormal login attempts that fail:

- Without 2FA
  - ADS reports "abnormal"; the submitted password is incorrect
- With 2FA:
  - ADS reports "abnormal"; the submitted password is incorrect
  - ADS reports "abnormal"; the submitted password is correct but 2FA fails

***pwd***

***pwd***



ADS: *abnormal*

# Our Framework

*c* = “alice@yyy.com : ***alicepwd***”,
a leaked username-password pair
possessed by the credential stuffer

**Credential Stuffer**

Websites where Alice has accounts


alice@yyy.com : *alicepwd0*
**ADS**

2FA


alice@yyy.com : *alicepwd*
**ADS**
**2FA**


alice@yyy.com : *alicepwd*
**ADS**

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

Websites where Alice
has accounts

**Credential Stuffer**

Login attempt with *c*


***≠ alicepwd***

alice@yyy.com : *alicepwd0*
**ADS**: *abnormal*
**SUSPICIOUS**: **{}**

2FA

alice@yyy.com : *alicepwd*
**ADS**
**2FA**

alice@yyy.com : *alicepwd*
**ADS**

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

Websites where Alice
has accounts

**Credential Stuffer**

Login attempt with *c*


alice@yyy.com : *alicepwd0*
**ADS**: *abnormal*
**SUSPICIOUS**: { ***alicepwd*** }

***≠ alicepwd***

collect

2FA

alice@yyy.com : *alicepwd*
**ADS**
**2FA**

alice@yyy.com : *alicepwd*
**ADS**

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

Websites where Alice
has accounts

alice@yyy.com : *alicepwd0*
**ADS**: *abnormal*
**SUSPICIOUS**: { ***alicepwd*** }

**Credential Stuffer**

Login attempt with *c*


2FA

alice@yyy.com : *alicepwd* ***= alicepwd***
**ADS**: *abnormal*
**2FA**: *failed*
**SUSPICIOUS**: { ***alicepwd*** }

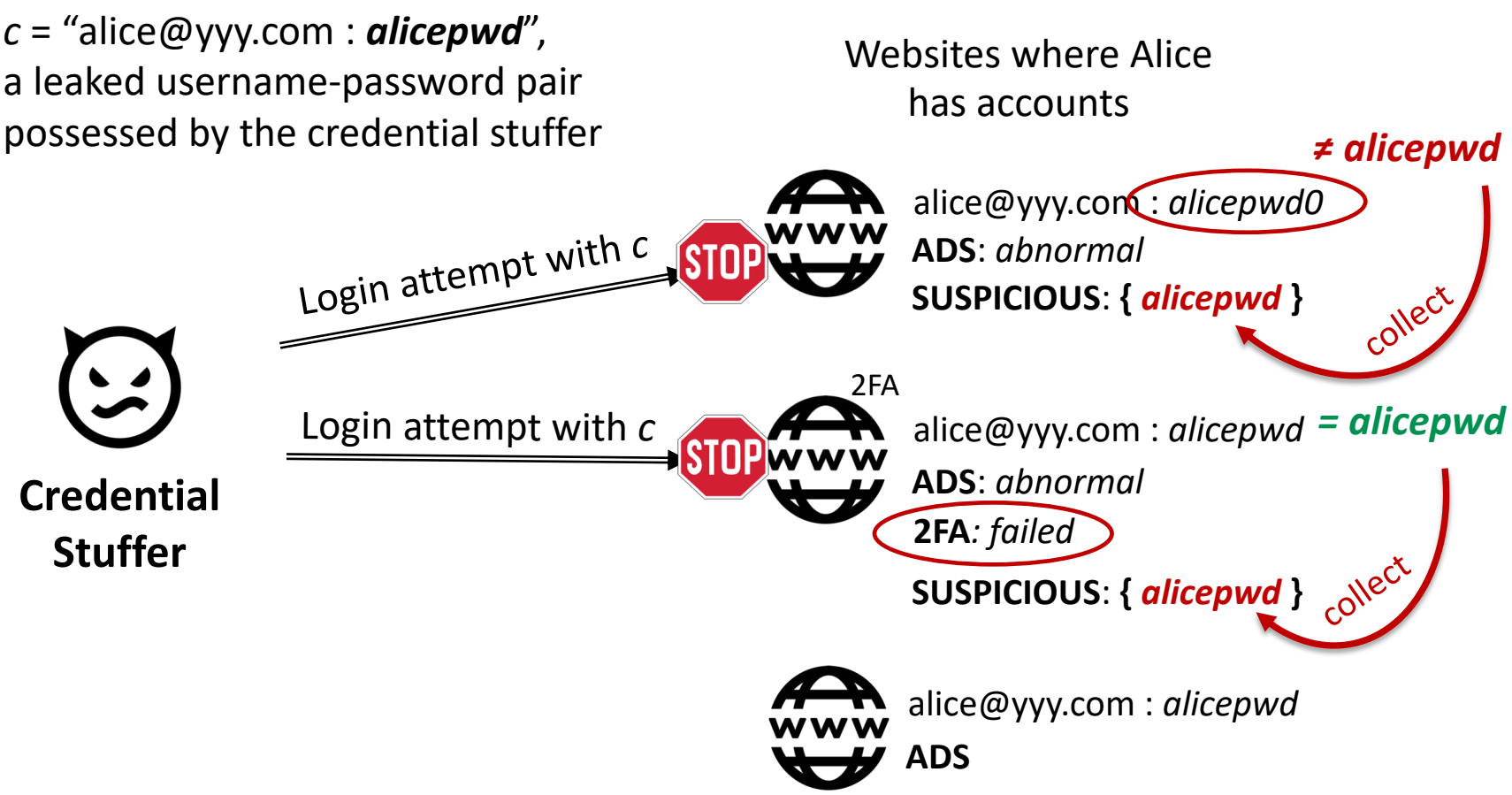
collect

alice@yyy.com : *alicepwd*
**ADS**

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer




***COLLECTING phase***

# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

**Credential Stuffer**

Login attempt with *c*

Websites where Alice has accounts

alice@yyy.com : *alicepwd0*
**ADS**: *abnormal*
**SUSPICIOUS**: { ***alicepwd*** }

2FA

alice@yyy.com : *alicepwd*
**ADS**: *abnormal*
**2FA**: *failed*
**SUSPICIOUS**: { ***alicepwd*** }

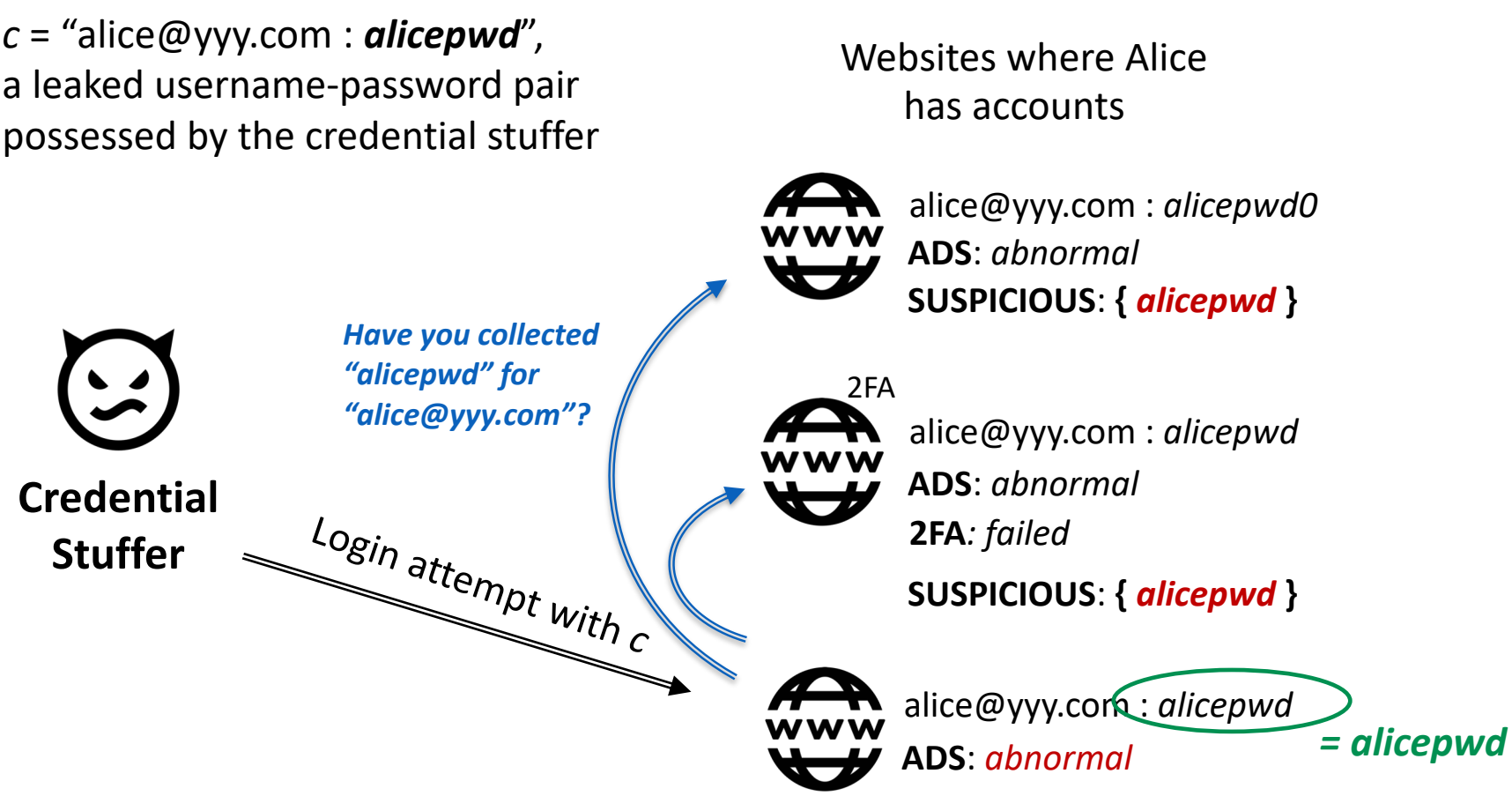
alice@yyy.com : *alicepwd*
**ADS**: *abnormal*

***= alicepwd***

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Our Framework

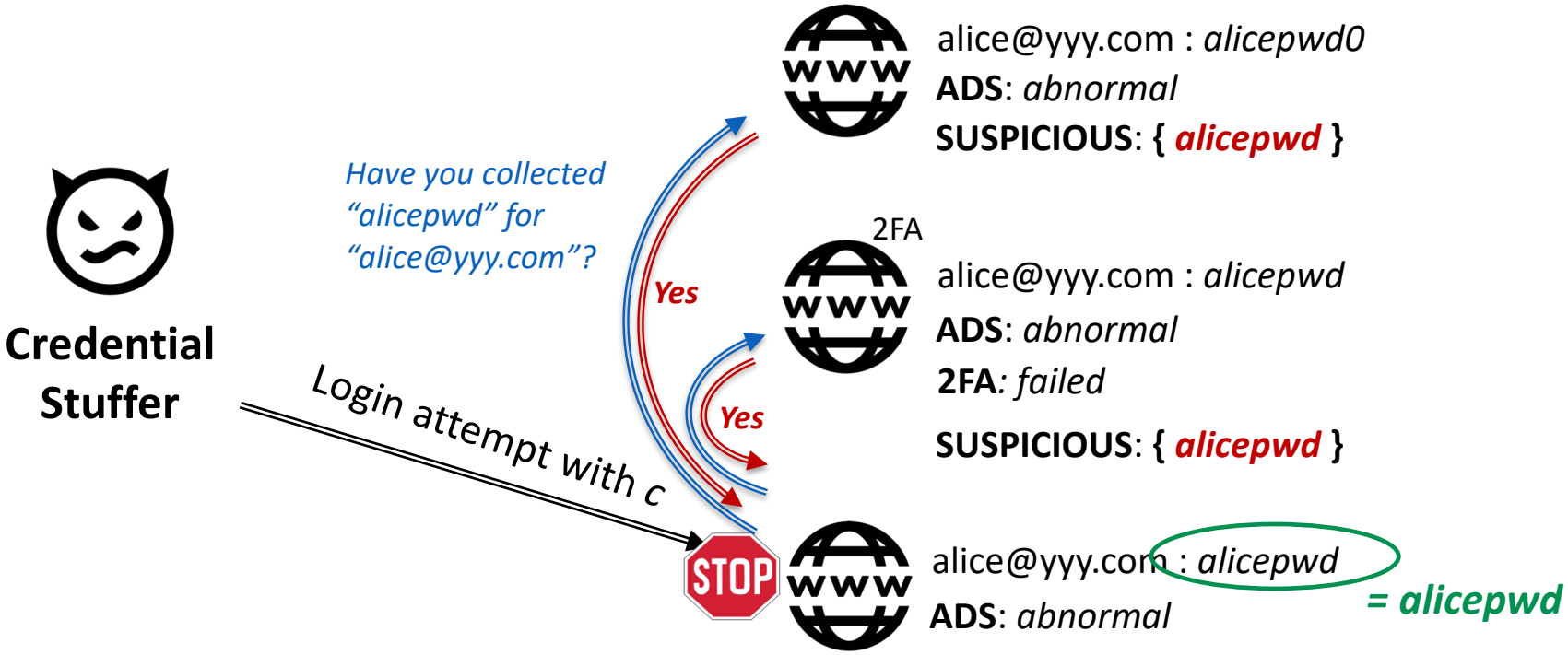
*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer



# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer

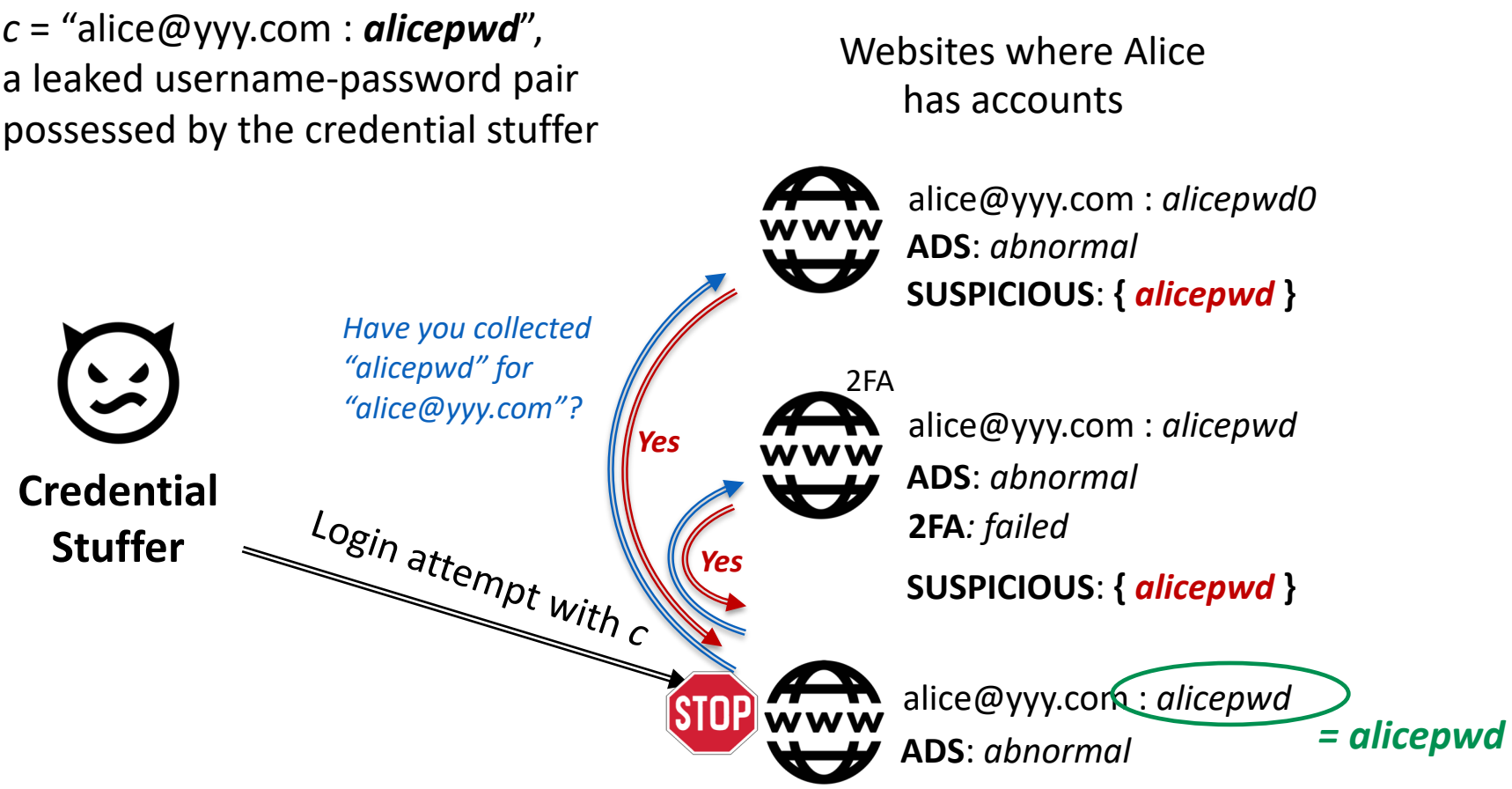
Websites where Alice
has accounts




A positive detection happens when the number of received positive responses is **>=** a pre-set threshold ("**attack width**").

# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair
possessed by the credential stuffer



# Our Framework

*c* = "alice@yyy.com : ***alicepwd***",
a leaked username-password pair

Websites where Alice

Two important questions:

- **False detection rate (FDR)**
  - *What if a (forgetful) user "guesses" her own passwords at her accounts?*
- **True detection rate (TDR)**
  - *What if a credential stuffer tries to circumvent detection by trying a smart attack strategy?*




alice@yyy.com : *alicepwd*
**ADS**: *abnormal*

***= alicepwd***



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

***COUNTING phase***

# Conservatively Estimating FDR & TDR

- ***A forgetful user as a MDP*:***
  - ***Maximizing the probability of triggering a false detection (false detection rate)***



* MDP: Markov decision process

# Conservatively Estimating FDR & TDR

- A forgetful user as a MDP*:
  - Maximizing the probability of triggering a false detection (false detection rate)

- ***A credential stuffer as a MDP*:***
  - ***Minimizing the probability of getting detected while maximizing the number of account takeovers (true detection rate)***



* MDP: Markov decision process

# Conservatively Estimating FDR & TDR

***Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users*

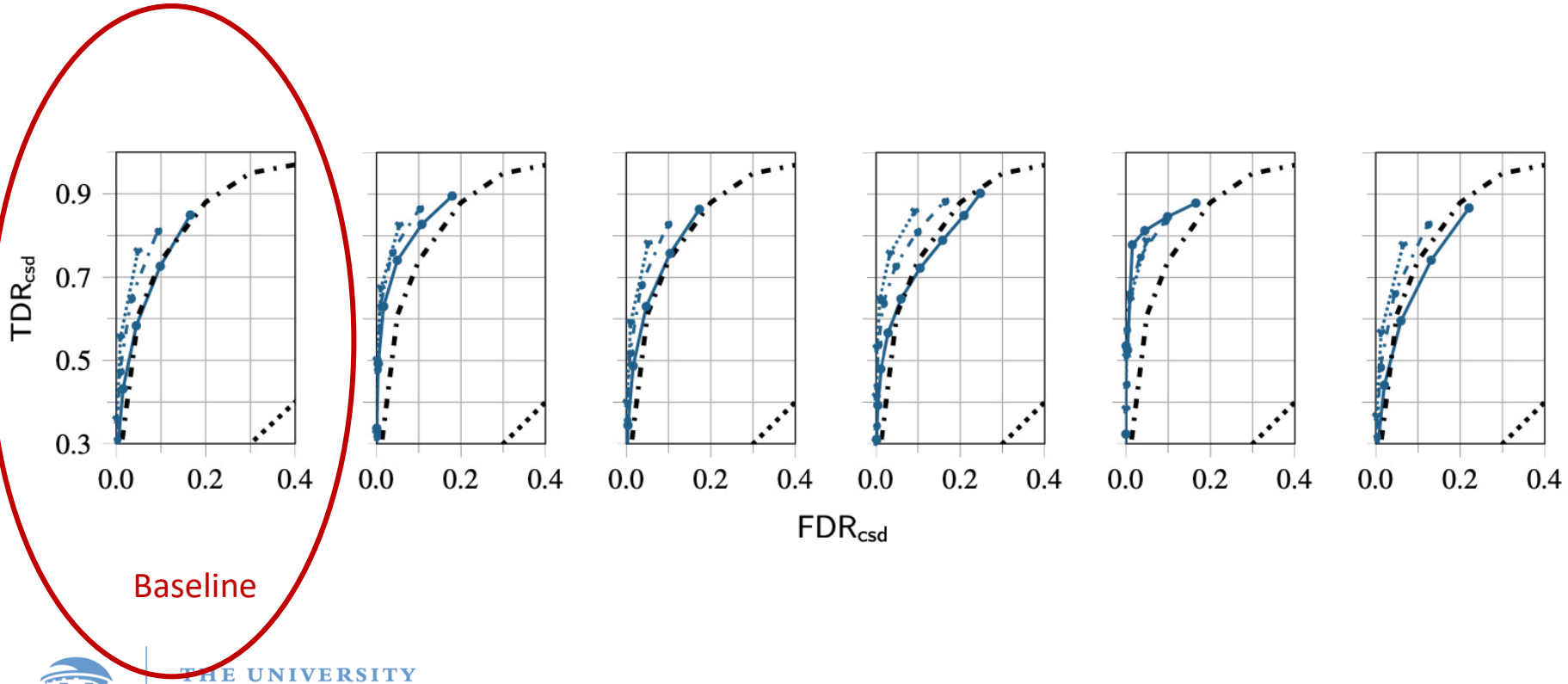
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- ***Default (baseline) setting****: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them*



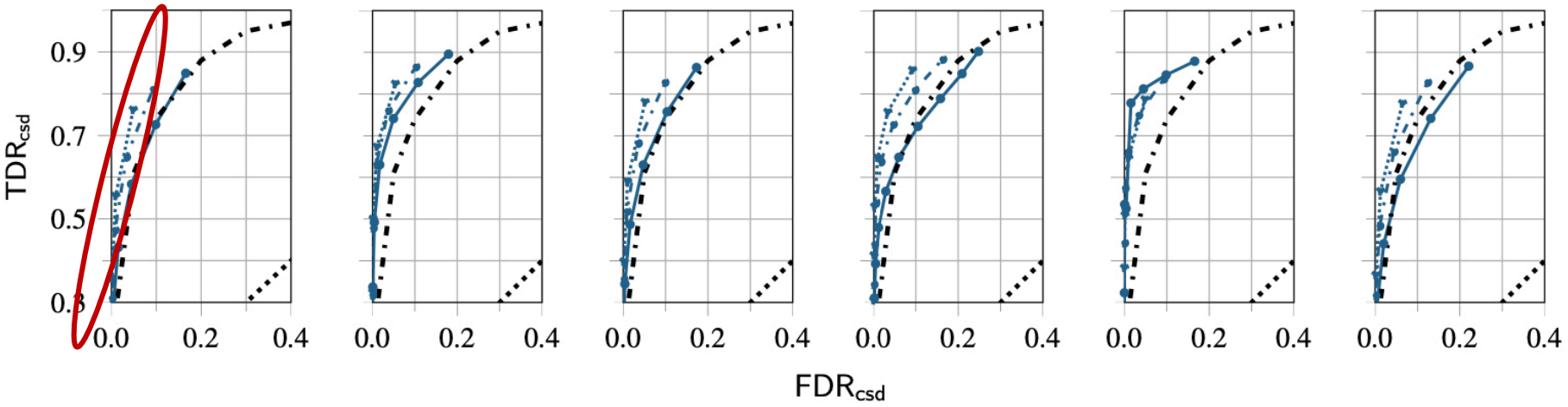

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- **Default (baseline) setting**: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them
- ***Blue curves**: each for a different ADS threshold in the collecting phase*




Baseline

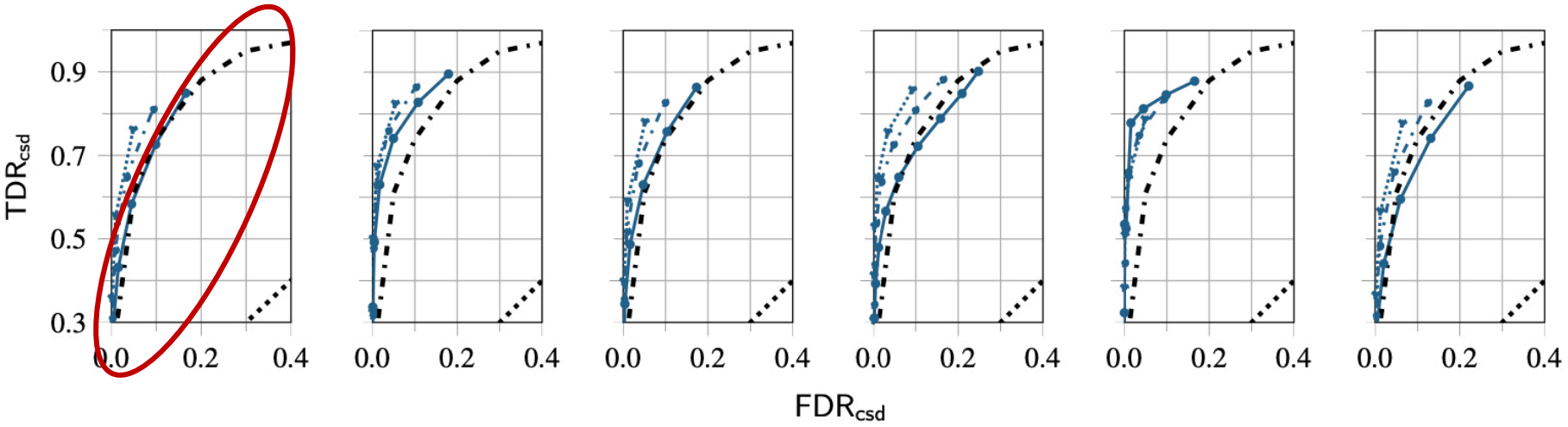
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- **Default (baseline) setting**: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them
- **Blue curves**: each for a different ADS threshold in *the collecting phase*
- ***Black, dashed curves**: corresponding ADS's accuracy in detecting abnormal logins*




Baseline

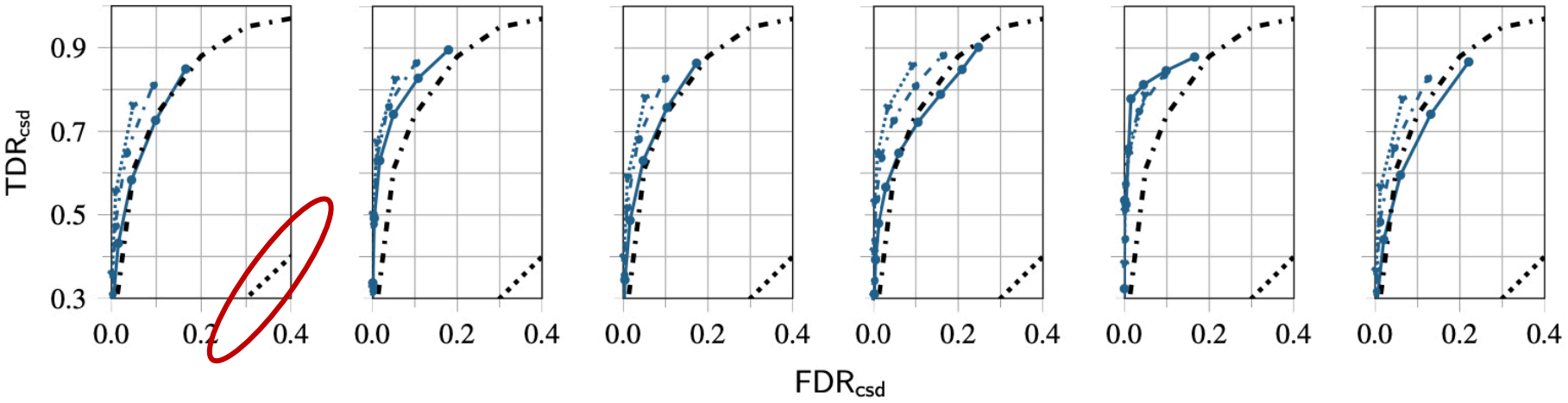
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- **Default (baseline) setting**: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them
- **Blue curves**: each for a different ADS threshold in *the collecting phase*
- **Black, dashed curves**: corresponding ADS's accuracy in detecting suspicious logins
- ***Black, dotted lines**: random guessing*




Baseline

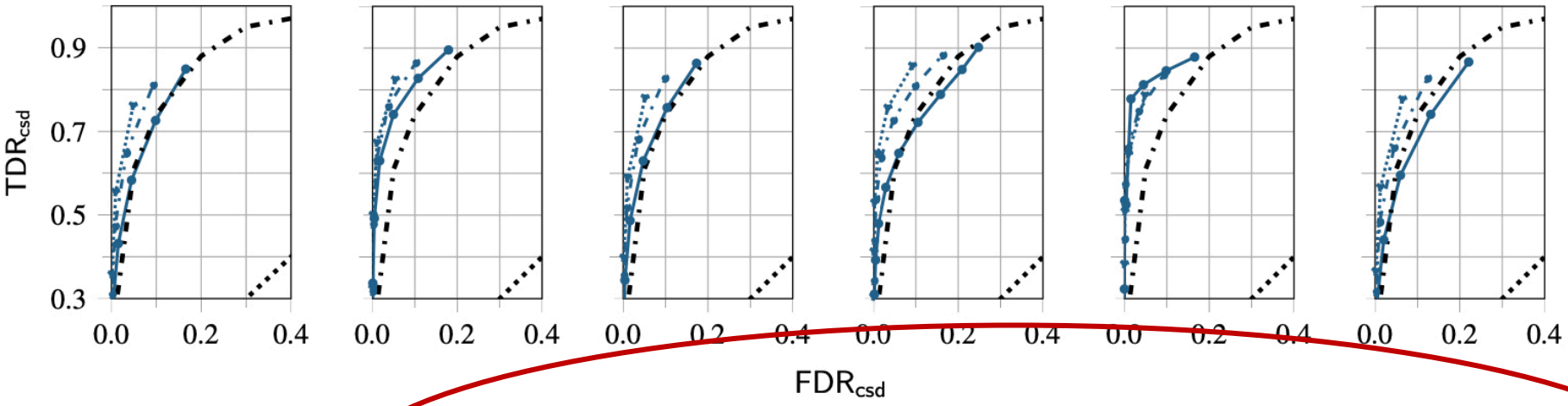
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- **Default (baseline) setting**: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them
- **Blue curves**: each for a different ADS threshold in *the collecting phase*
- **Black, dashed curves**: corresponding ADS's accuracy in detecting suspicious logins
- **Black, dotted lines**: random guessing




Baseline

Less pwd reuse

# of pwds +1

# of accnts + 10

# of 2FA + 5

Higher ADS detection rates in the *counting phase*

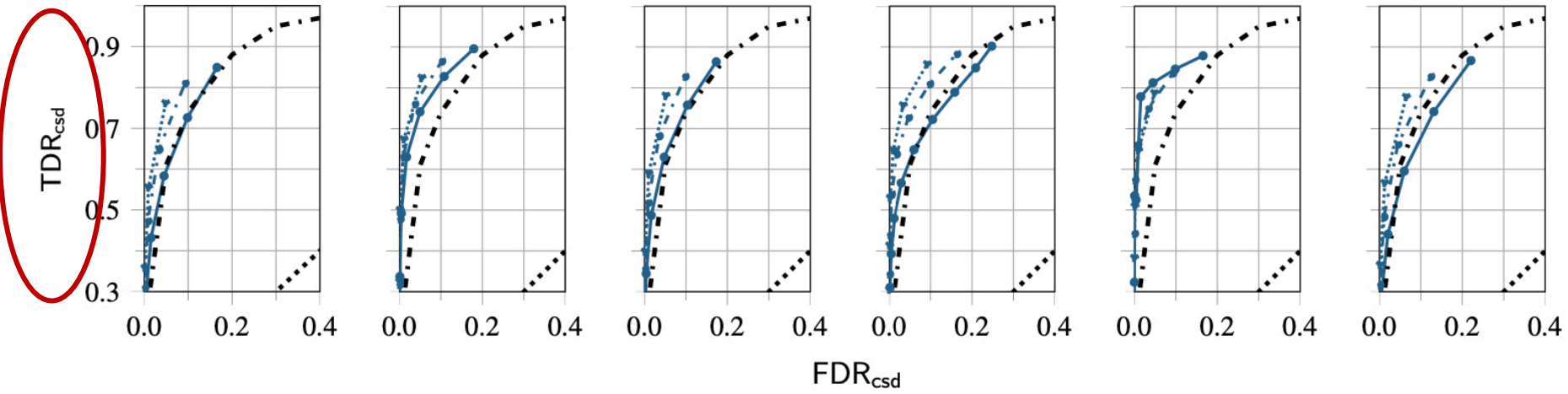
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- **Default (baseline) setting**: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them
- **Blue curves**: each for a different ADS threshold in *the collecting phase*
- **Black, dashed curves**: corresponding ADS's accuracy in detecting suspicious logins
- **Black, dotted lines**: random guessing




Baseline | Less pwd reuse | # of pwds +1 | # of accnts + 10 | # of 2FA + 5 | Higher ADS detection rates in the *counting phase*

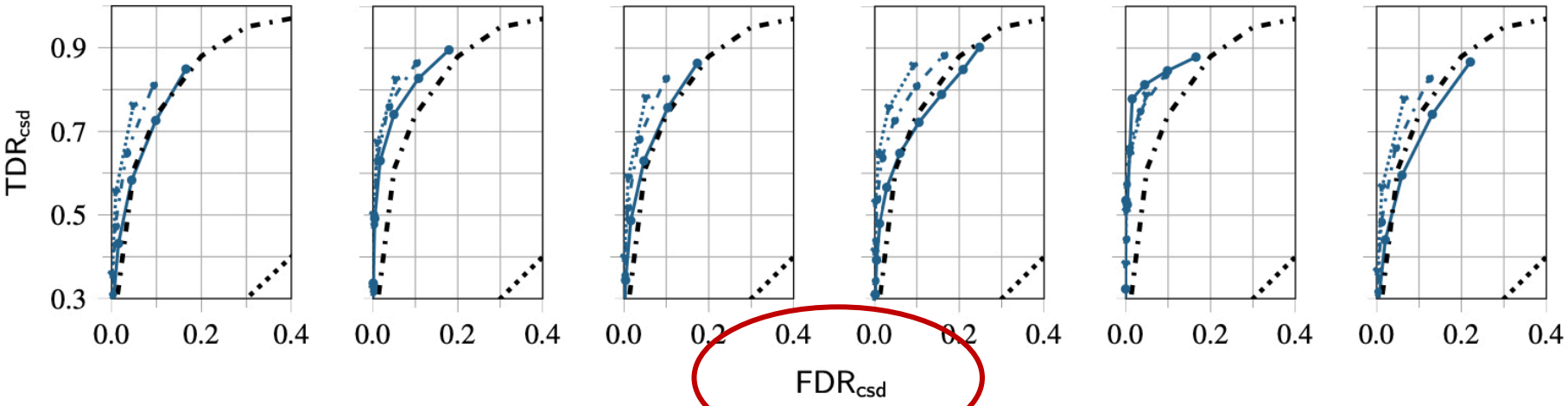
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- **Default (baseline) setting**: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them
- **Blue curves**: each for a different ADS threshold in *the collecting phase*
- **Black, dashed curves**: corresponding ADS's accuracy in detecting suspicious logins
- **Black, dotted lines**: random guessing




Baseline

Less pwd reuse

# of pwds +1

# of accnts + 10

# of 2FA + 5

Higher ADS detection rates in the *counting phase*

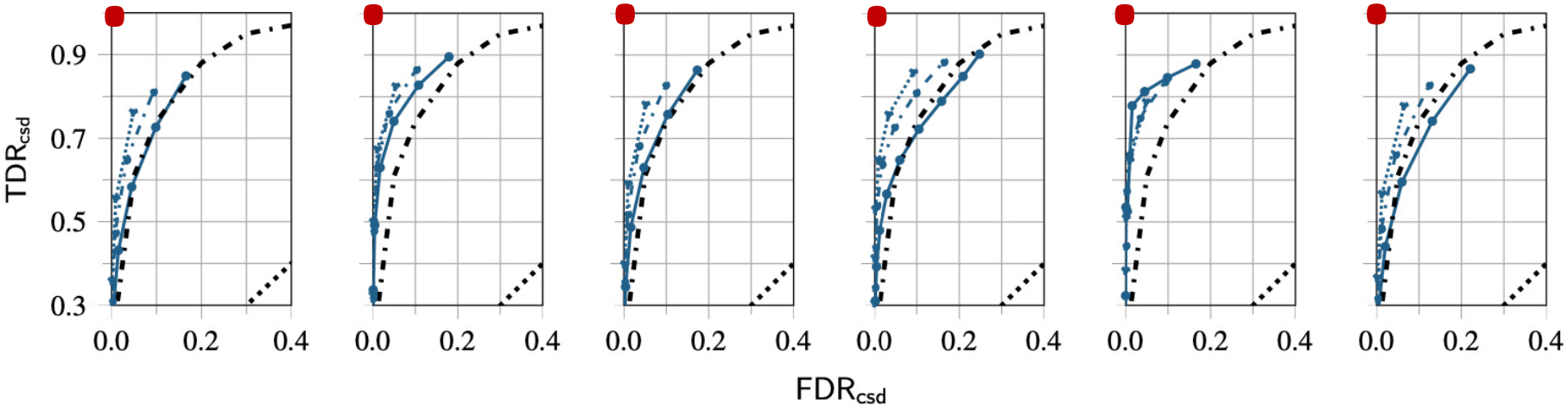
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Phishing attackers***: valid passwords from same countries with same browser user-agent strings of legitimate users

- **Default (baseline) setting**: some level of password reuse in a set of 4 distinct passwords across 10 accounts (one per site) with no 2FA deployed among them
- **Blue curves**: each for a different ADS threshold in *the collecting phase*
- **Black, dashed curves**: corresponding ADS's accuracy in detecting suspicious logins
- **Black, dotted lines**: random guessing



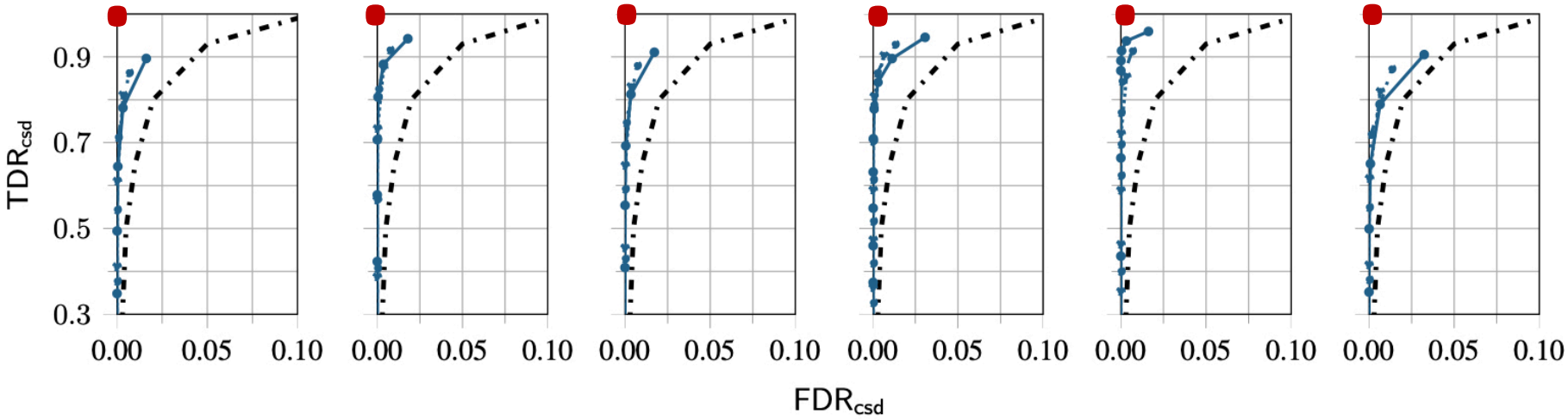

Baseline | Less pwd reuse | # of pwds +1 | # of accnts + 10 | # of 2FA + 5 | Higher ADS detection rates in the *counting phase*

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Freeman et al. (NDSS 2016)

# Conservatively Estimating FDR & TDR

**Researching attackers***: valid passwords from same countries of legitimate users.




| Baseline | Less pwd reuse | # of pwds +1 | # of accnts + 10 | # of 2FA + 5 | Higher ADS detection rates in the *counting phase* |
|---|---|---|---|---|---|

* Freeman et al. (NDSS 2016)

Other features of our framework:



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*c* = “alice@yyy.com : ***alicepwd***”

Other features of our framework:

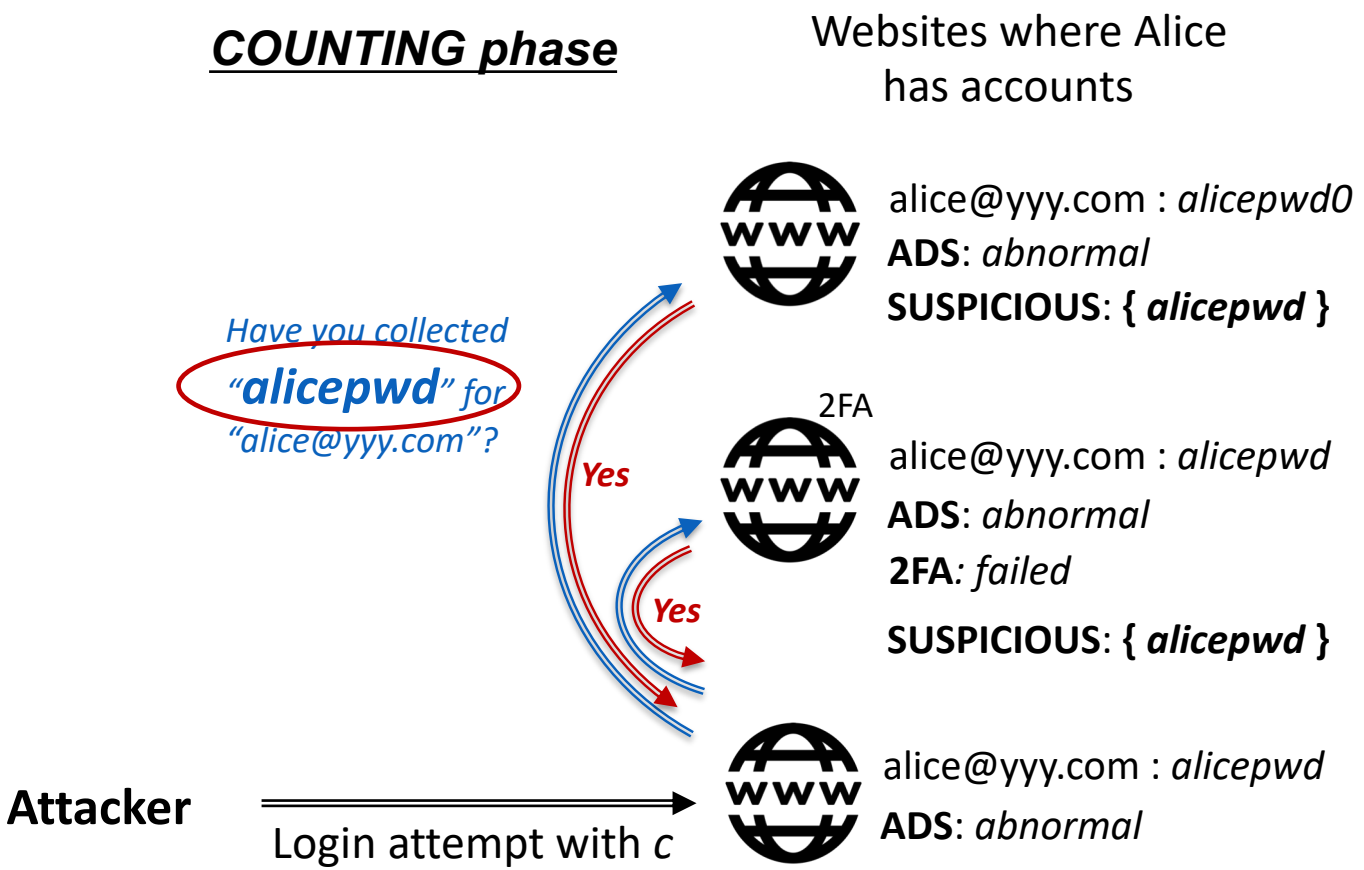
- ***Account security***



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*c* = “alice@yyy.com : ***alicepwd***”

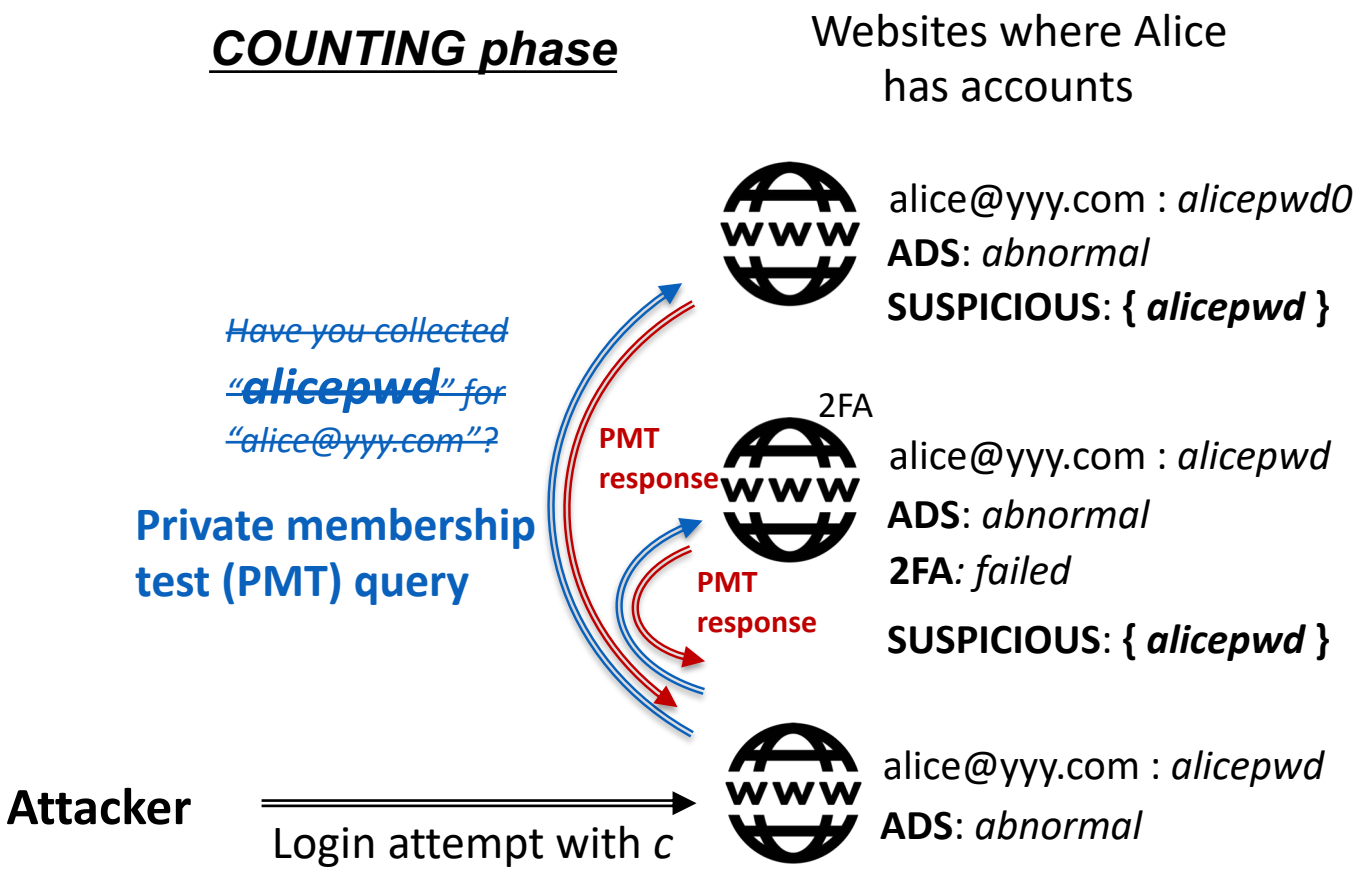
# Account Security






*c* = "alice@yyy.com : ***alicepwd***"

# Account Security



## Other features of our framework:

- Account security
  - ***A new one-round two-party private membership test (PMT) protocol***



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*c* = “alice@yyy.com : ***alicepwd***”

## Other features of our framework:

- Account security
  - A new one-round two-party private membership test (PMT) protocol
- ***Directory***



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*c* = "alice@yyy.com : ***alicepwd***"

## Other features of our framework:

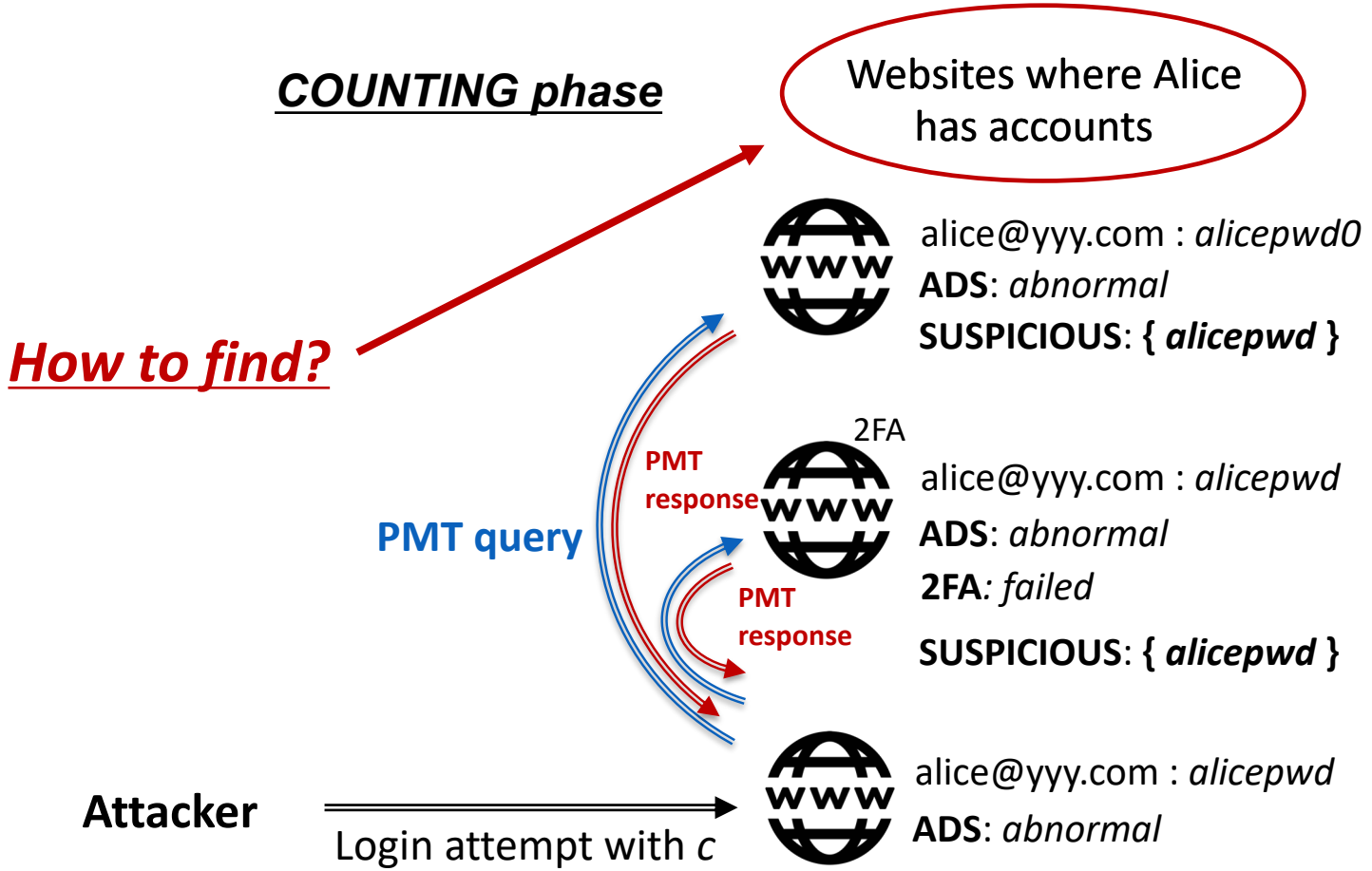
- Account security
  - A new one-round two-party private membership test (PMT) protocol
- Directory
  - ***A "look-up table" that maintains where a user has accounts***



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*c* = "alice@yyy.com : ***alicepwd***"

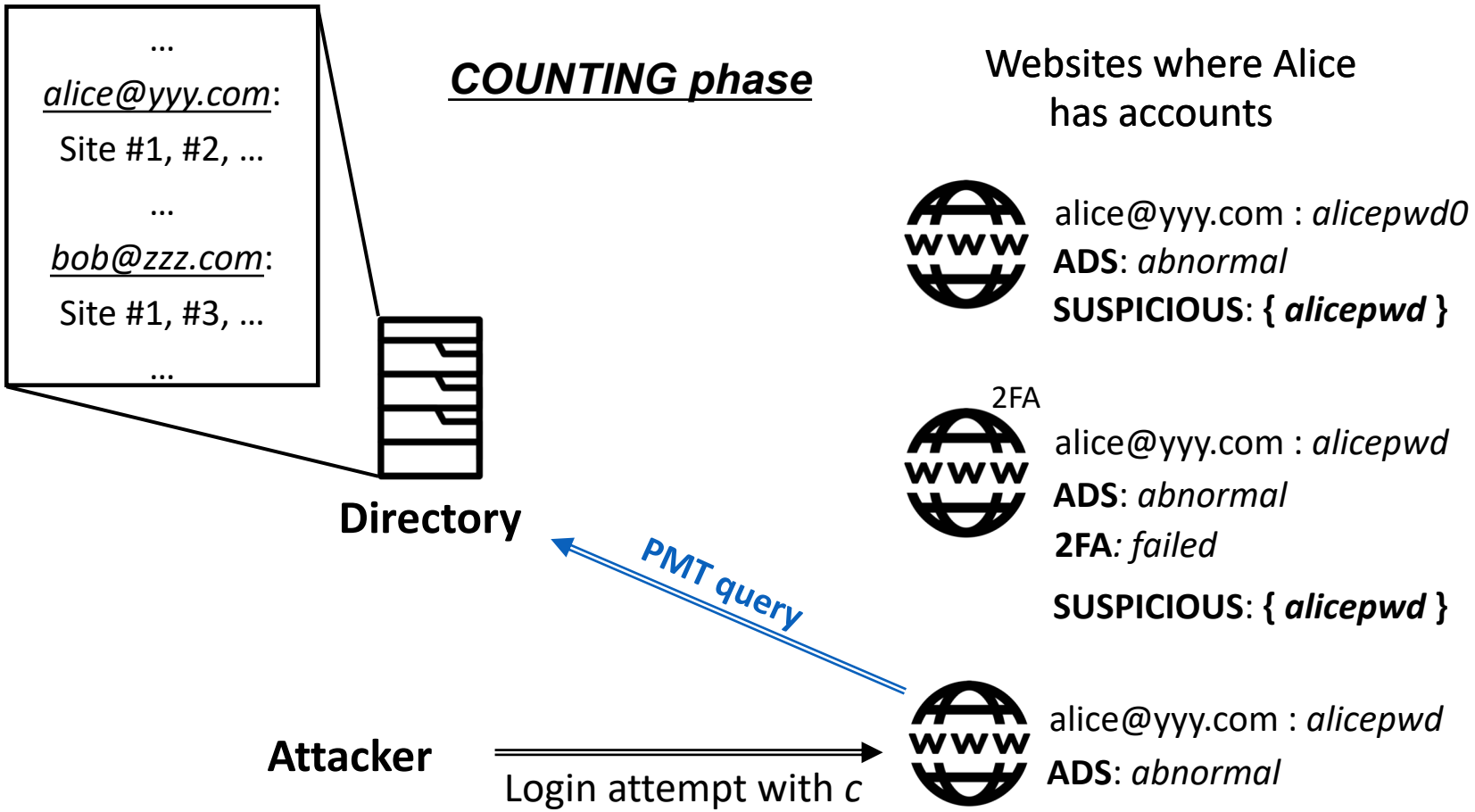
# Directory






$c$ = "alice@yyy.com : ***alicepwd***"

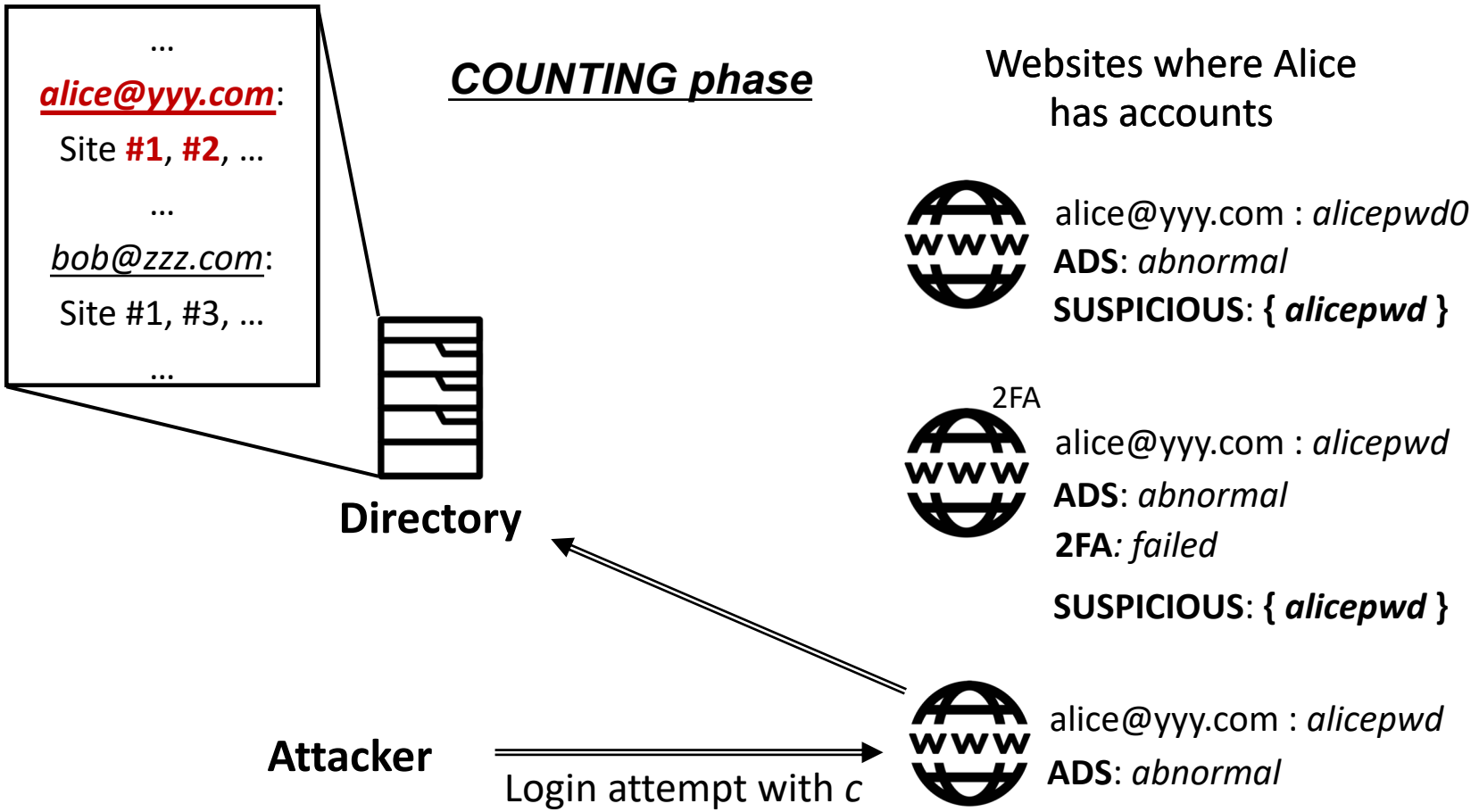
# Directory






$c$ = "alice@yyy.com : ***alicepwd***"

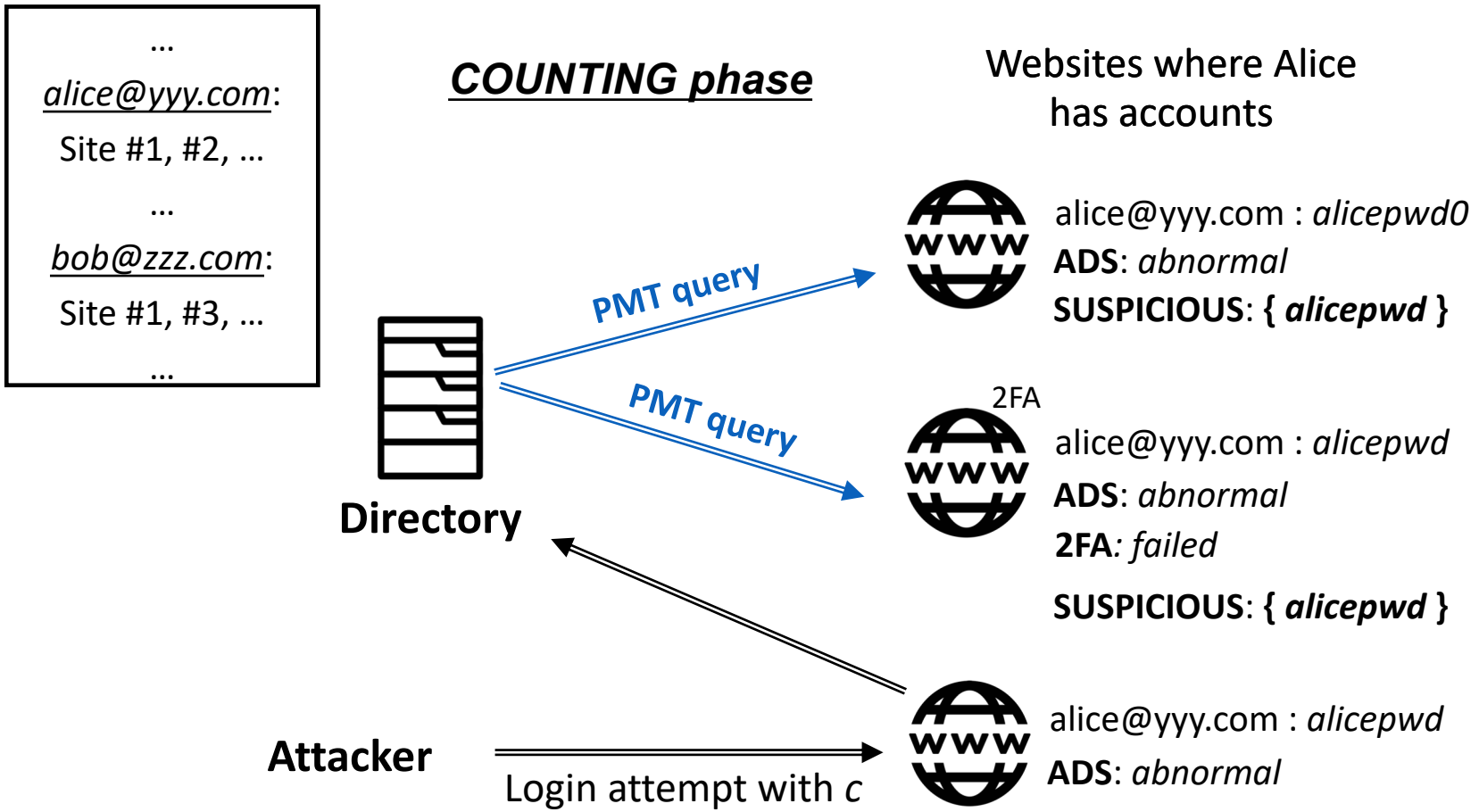
# Directory






$c$ = "alice@yyy.com : ***alicepwd***"

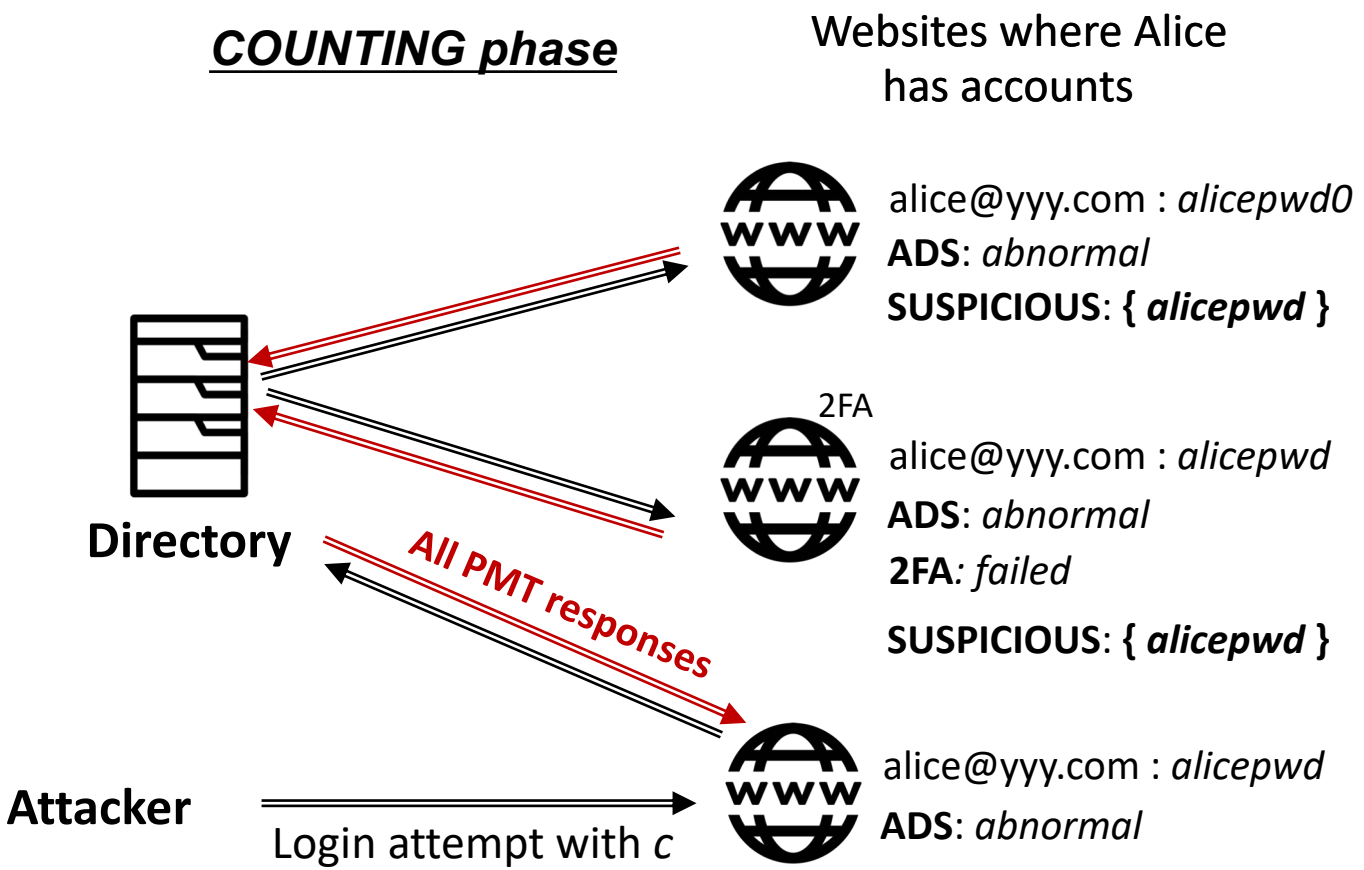
# Directory






$c$ = "alice@yyy.com : ***alicepwd***"

# Directory






*c* = "alice@yyy.com : ***alicepwd***"

## Other features of our framework:

- Account security
  - A new one-round two-party private membership test (PMT) protocol
- Directory
  - A “look-up table” that maintains where a user has accounts
- ***Login privacy***



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

*c* = “alice@yyy.com : ***alicepwd***”

## Other features of our framework:

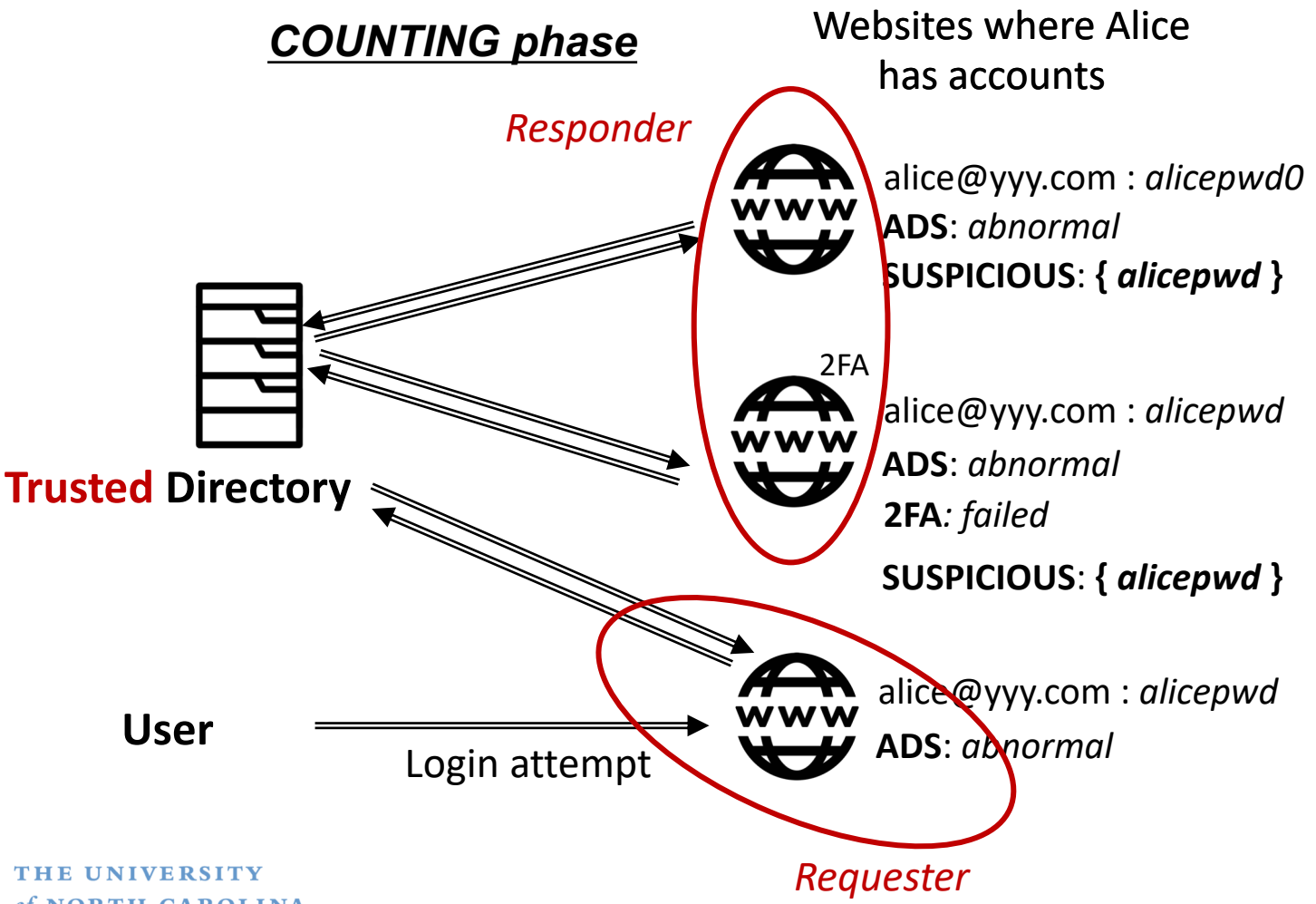
- Account security
  - A new one-round two-party private membership test (PMT) protocol
- Directory
  - A "look-up table" that maintains where a user has accounts
- Login privacy
  - ***Trusted directory for login privacy***





*c* = "alice@yyy.com : ***alicepwd***"

# Login Privacy



## Other features of our framework:

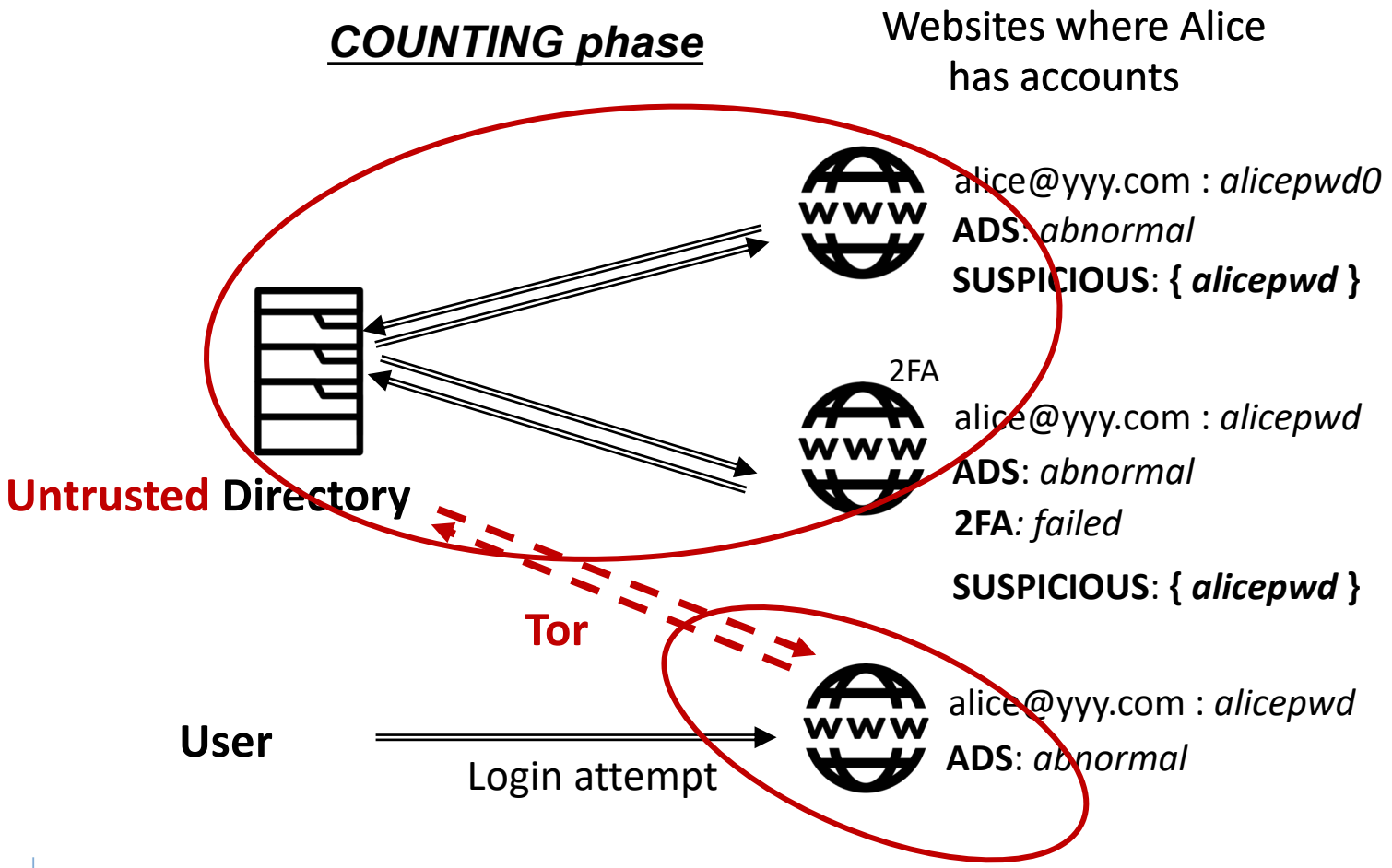
- Account security
  - A new one-round two-party private membership test (PMT) protocol
- Directory
  - A “look-up table” that maintains where a user has accounts
- Login privacy
  - Trusted directory for login privacy
  - ***Untrusted directory for login privacy***



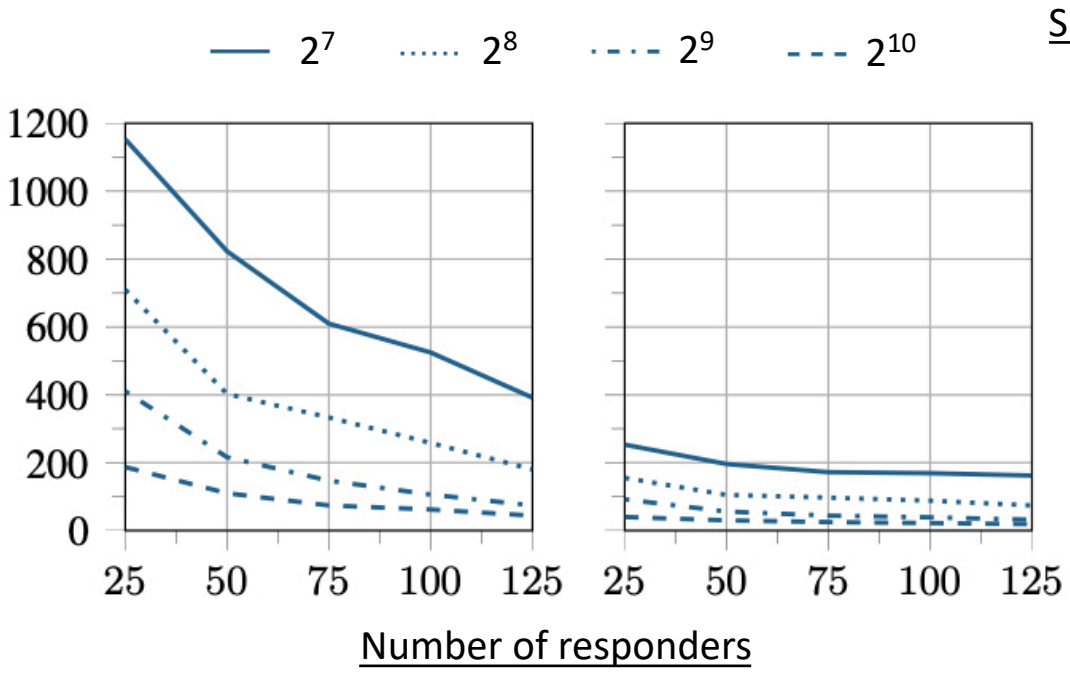


*c* = “alice@yyy.com : ***alicepwd***”

# Login Privacy



# Scalability

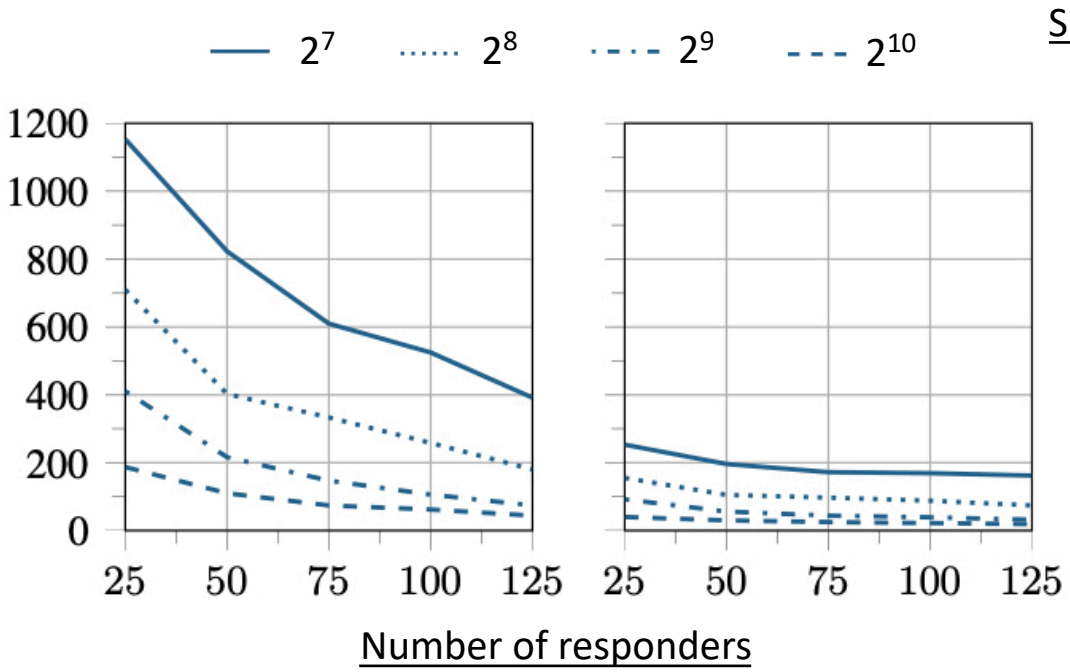



*Trusted* directory for *login privacy* (Qualifying response: <= 5s)

*Untrusted* directory for *login privacy* (Qualifying response: <= 8s)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Scalability

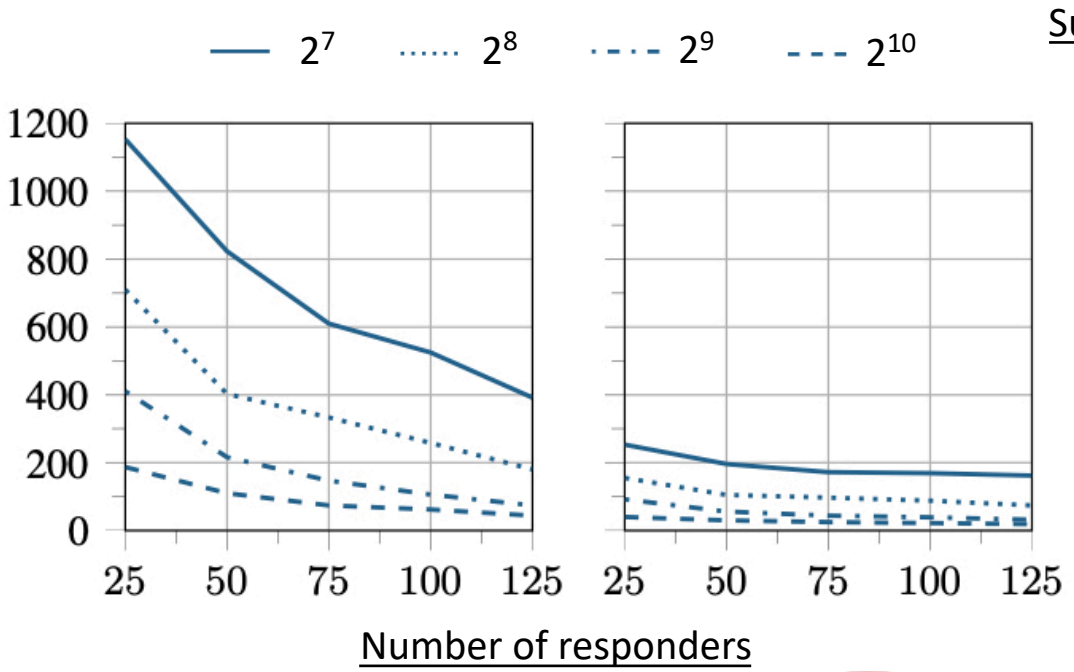



*Trusted* directory for *login privacy* (Qualifying response: <= 5s)

*Untrusted* directory for *login privacy* (Qualifying response: <= 8s)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Scalability

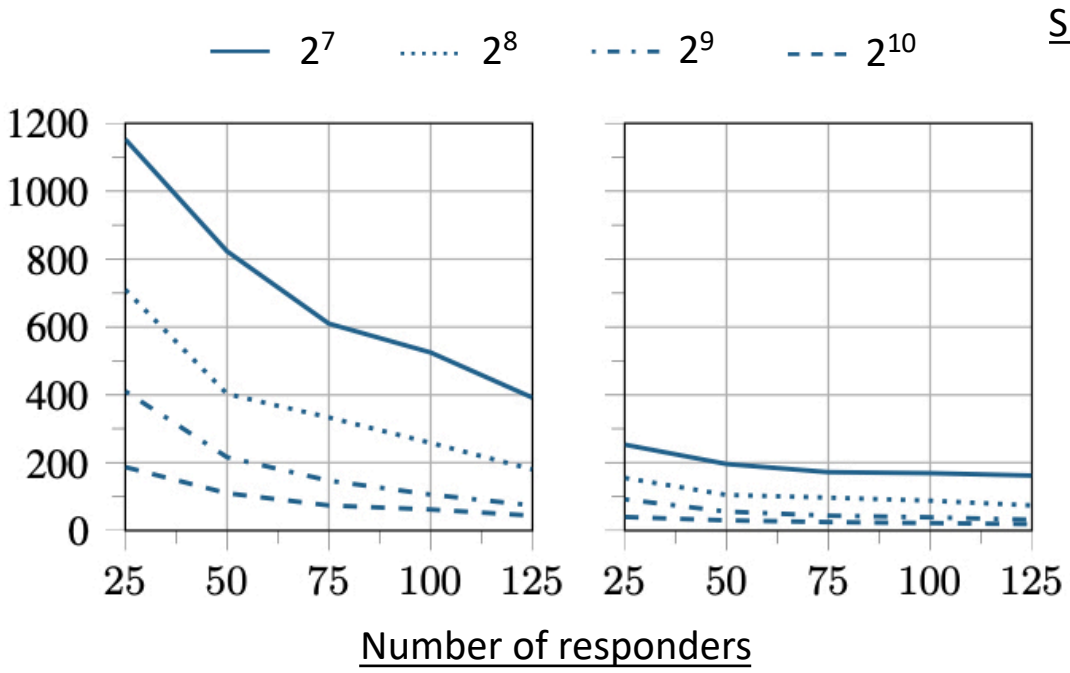



*Trusted* directory for *login privacy* (Qualifying response: <= 5s)

*Untrusted* directory for *login privacy* (Qualifying response: <= 8s)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Scalability

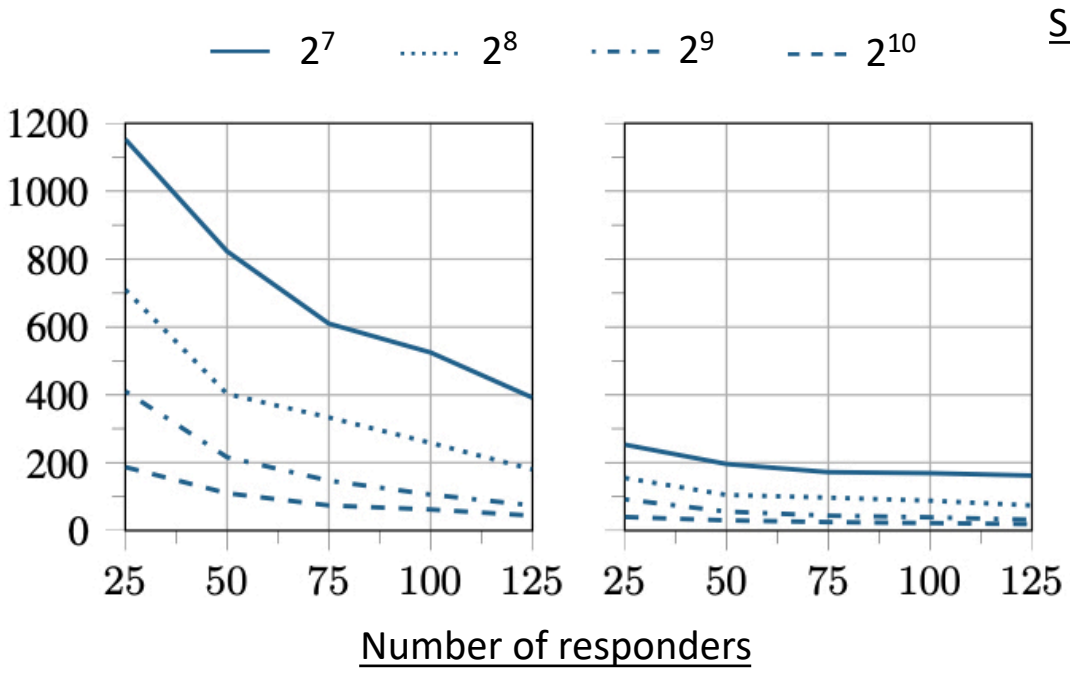



*Trusted* directory for *login privacy* (Qualifying response: <= 5s)

*Untrusted* directory for *login privacy* (Qualifying response: <= 8s)

Response time measured at the requester

# Scalability

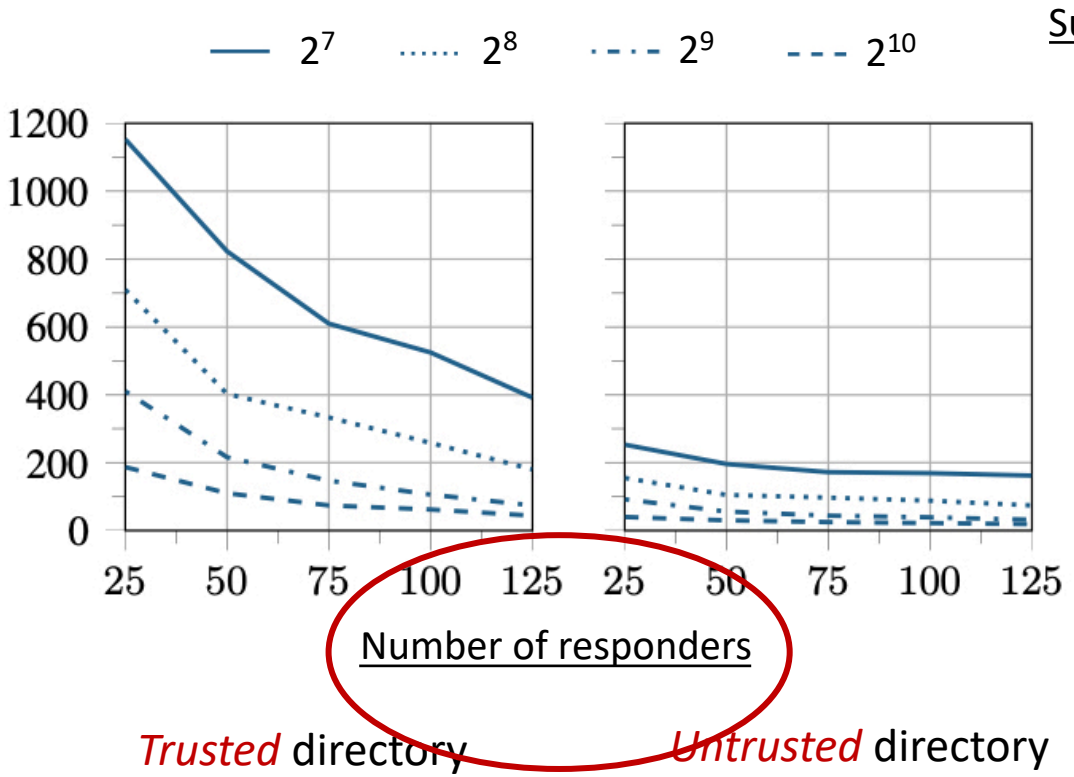



*Trusted* directory for *login privacy* (Qualifying response: <= 5s)

*Untrusted* directory for *login privacy* (Qualifying response: <= 8s)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Scalability

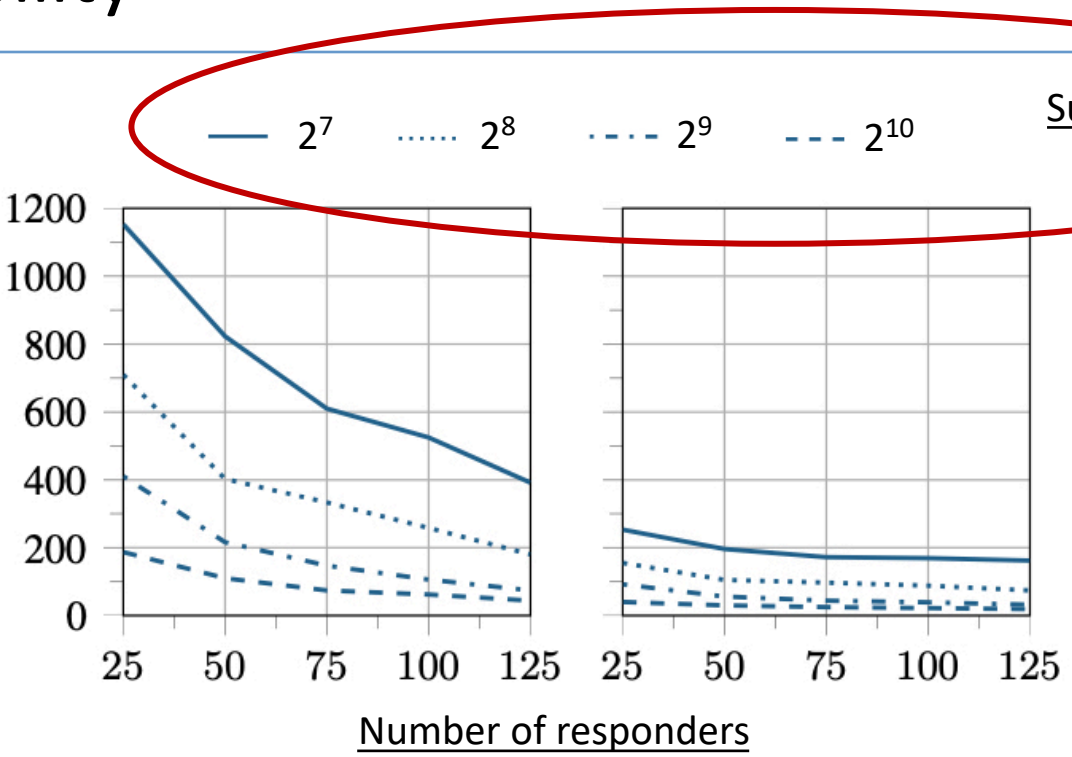



*Trusted* directory for *login privacy* (Qualifying response: <= 5s)

*Untrusted* directory for *login privacy* (Qualifying response: <= 8s)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Scalability




*Trusted* directory for *login privacy* (Qualifying response: <= 5s)

*Untrusted* directory for *login privacy* (Qualifying response: <= 8s)

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Scalability

| | Credential-stuffing login attempts per day | Proportion that succeed | Proportion of all login attempts |
|---|---|---|---|
| Airline | 1.4 Million | 1.00% | 60% |
| Hotel | 4.3 Million | 1.00% | 44% |
| Retail | 131.5 Million | 0.50% | 91% |
| Consumer banking | 232.2 Million | 0.05% | 58% |

Table: Credential stuffing estimates for four major U.S. industries*

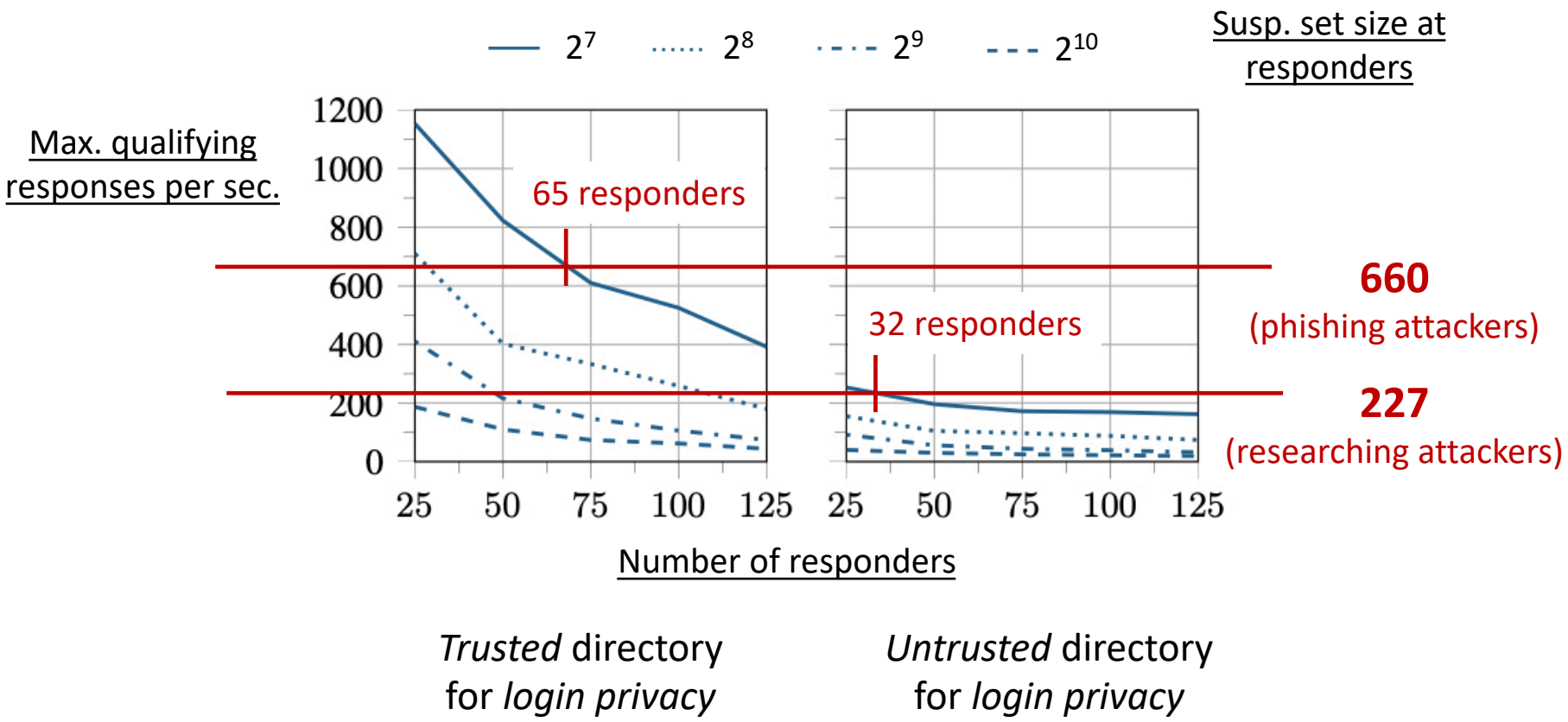
*Total number of PMT queries per second:*

- *If ADS false & true detection rates are 0.30 & 0.95 (against phishing attackers):* ***660***
- *If ADS false & true detection rates are 0.10 & 0.99 (against researching attackers):* ***227***

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

* Shape Security, "2018 Credential spill report"

# Scalability



# Summary

- ***A framework to detect credential stuffing***

# Summary

- A framework to detect credential stuffing
  - ***Leverages ADS and evidence trail left by credential stuffing***

# Summary

- A framework to detect credential stuffing
  - Leverages ADS and evidence trail left by credential stuffing
  - ***Account security achieved by a novel PMT protocol***

# Summary

- A framework to detect credential stuffing
  - Leverages ADS and evidence trail left by credential stuffing
  - Account security achieved by a novel PMT protocol
  - ***Login privacy enforced by the directory or by Tor***

# Summary

- A framework to detect credential stuffing
  - Leverages ADS and evidence trail left by credential stuffing
  - Account security achieved by a novel PMT protocol
  - Login privacy enforced by the directory or by Tor
- ***First to detect active credential stuffing attacks across multiple websites***

# Summary

- A framework to detect credential stuffing
  - Leverages ADS and evidence trail left by credential stuffing
  - Account security achieved by a novel PMT protocol
  - Login privacy enforced by the directory or by Tor
- First to detect active credential stuffing attacks across multiple websites
- ***Even a minimal-infrastructure deployment of our framework should support the combined login load experienced by four major sectors of the U.S economy***

# Thank you!

Coby Wang
Email: *kwang@cs.unc.edu*