

# TRIDENT: Towards Detecting and Mitigating Web-based Social Engineering Attacks

Zheng Yang<sup>†</sup>, Joey Allen<sup>†</sup>, Matthew Landen<sup>†</sup>, Roberto Perdisci<sup>‡</sup>, Wenke Lee<sup>†</sup>

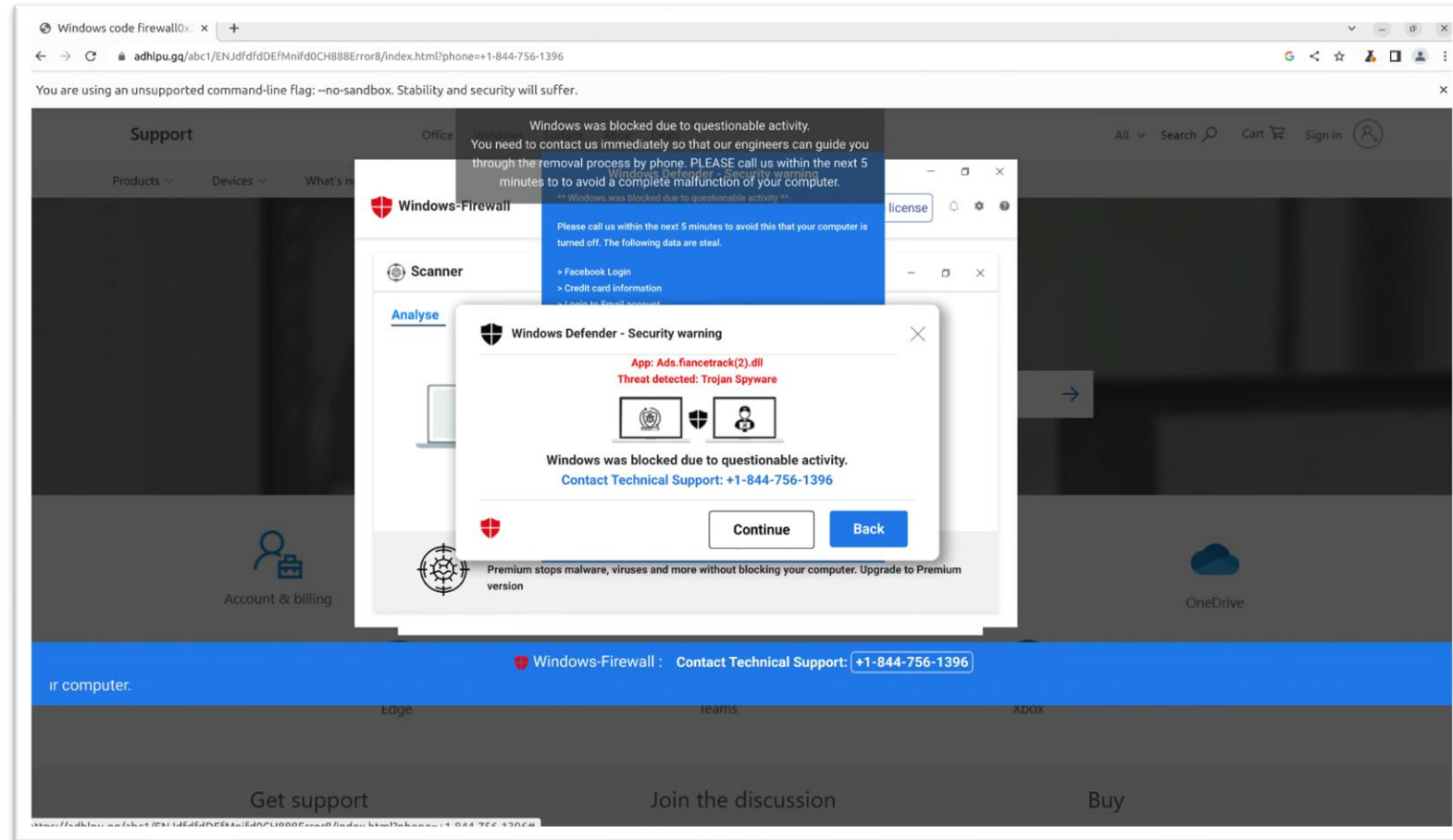
<sup>†</sup>Georgia Institute of Technology

<sup>‡</sup>University of Georgia

# Web-based Social Engineering Attacks

The image shows a simulated Windows desktop environment. A green alert box with a yellow warning icon and a red '1' in the top-left corner displays the text: "Your Mac is infected? Check your Mac for viruses". Below this, a white dialog box with a Windows logo icon contains the text: "Windows was blocked due to questionable activity. Contact Technical Support: +1-844-756-1396". At the bottom of the dialog are "Continue" and "Back" buttons. A blue footer bar at the bottom of the screen contains the text: "Windows-Firewall : Contact Technical Support: +1-844-756-1396". In the background, a "Windows Firewall" window is visible with the text: "You need to control... through the rem... minutes to...". Other visible elements include a "Scanner" window with an "Analyse" button, a "Windows" window with a shield icon, and a "Premium" advertisement for Windows Firewall. A black mouse cursor is positioned on the right side of the screen, pointing towards the green alert box.

# Prior State-of-the-art Solution: Tech-support Scam

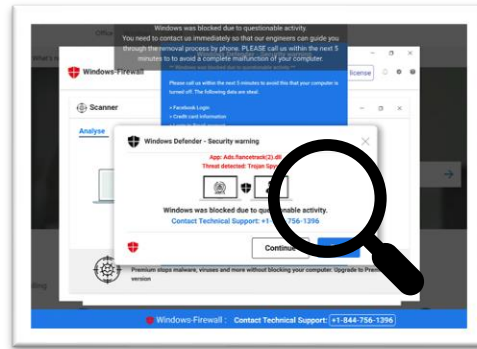


*Miramirkhani et al.*  
(NDSS'17)

# Prior State-of-the-art: Directly Detect SE Attacks

## Tech-support Scam

*Miramirkhani et al.*  
(NDSS'17)



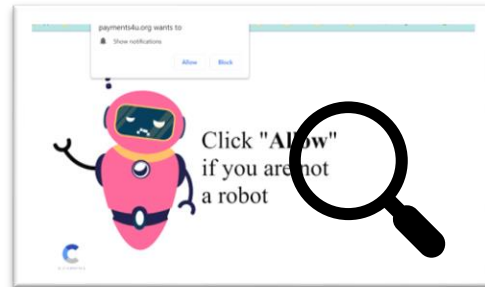
## Scareware

*Nelms et al.*  
(Usenix'16)



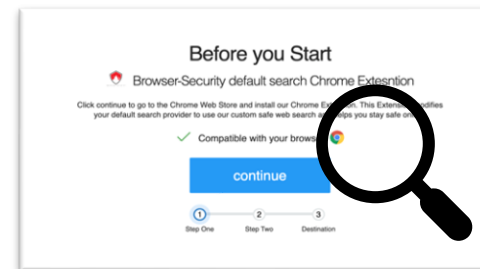
## Notification Spam

*Subramani et al.*  
(ACM IMC'20)



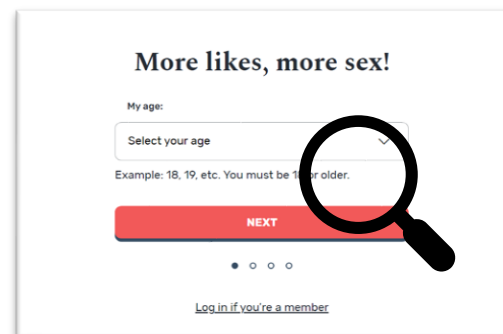
## Unwanted-software Download

*DeKoven et al.*  
(Usenix'17)



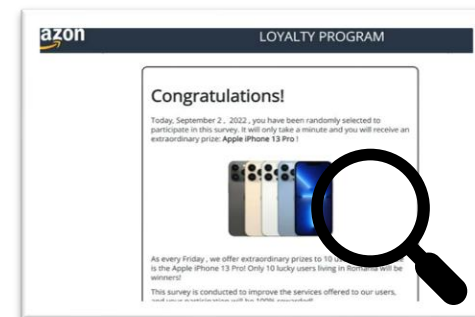
## Dating Scam

*Suarez-Tangil et al.*  
(TIFS'19)



## Prize Scam

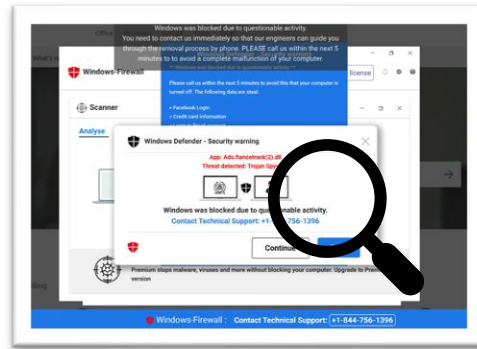
*Kharraz et al.*  
(IEEE S&P'18)



# Prior State-of-the-art: Directly Detect SE Attacks

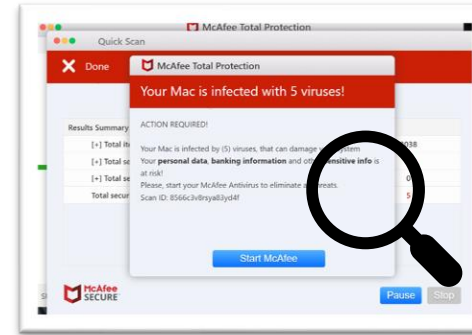
## Tech-support Scam

Miramirkhani et al.  
(NDSS'17)



## Scareware

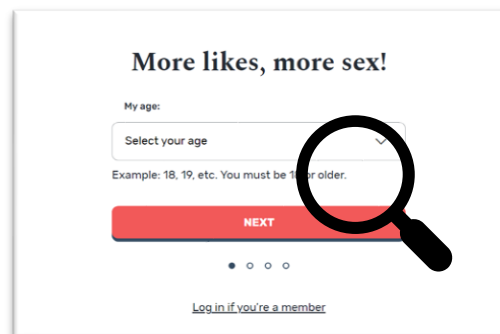
Nelms et al.  
(Usenix'16)



Directly detecting social engineering websites/attacks is unrealistic and not scalable!

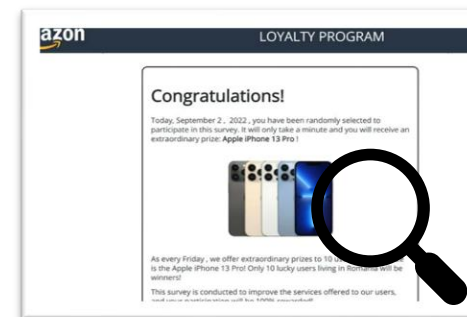
## Dating Scam

Suarez-Tangil et al.  
(TIFS'19)



## Prize Scam

Kharraz et al.  
(IEEE S&P'18)



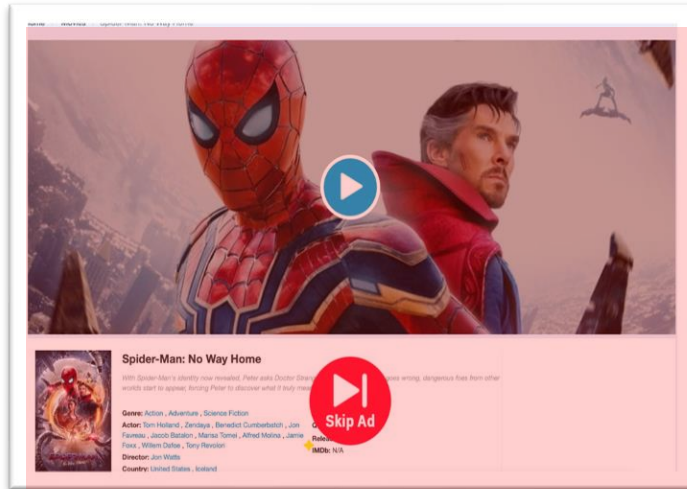
# Key Idea: Indirectly Detect and Block SE Attacks



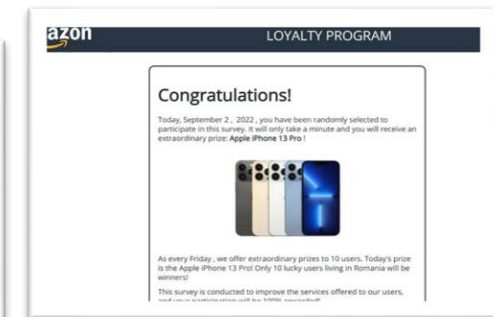
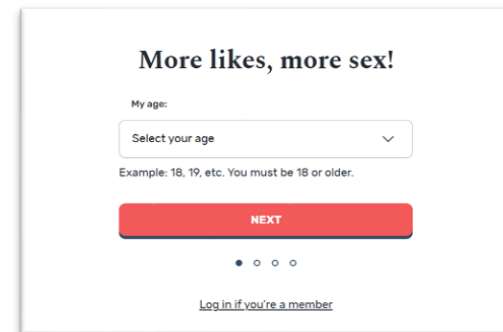
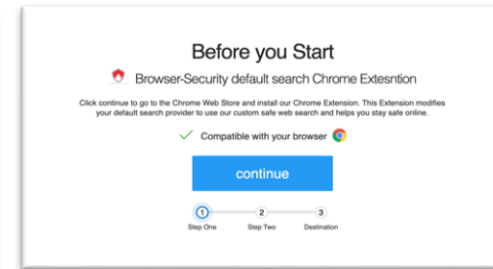
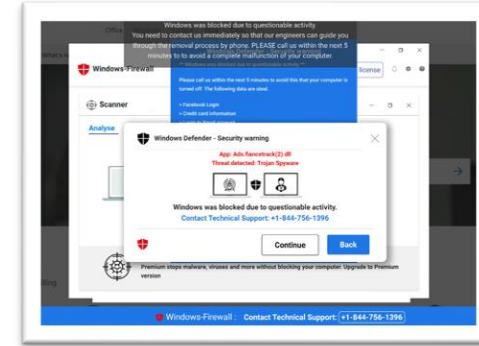
Low-tier Ad Networks



Navigation Blocked



Ad Publishers



SE Attacks



# Generic Ad Blockers Are Ineffective

## Blocklist-based Solution

- Brave Shield [49]
  - Ad-blocking module for Brave Browser.
  - **14.74% false negative rate** on 1,479 social engineering attacks.

## ML-based Solution

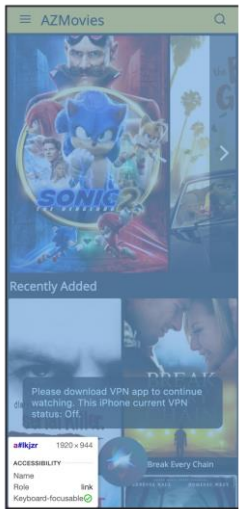
- AdGraph [19]

Model	Accuracy	Precision	Recall	F1
Original	90.52%	88.32%	88.33%	88.32%
Retrained	83.25%	80.12%	81.65%	80.88%
SE-Ads	81.51%	71.34%	75.33%	73.28%

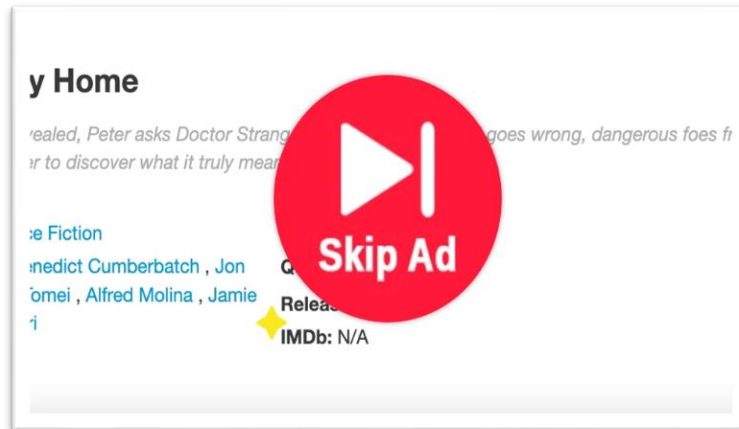
Social engineering “ads” can evade state-of-the-art ad blocking tools easily!

# Tech Challenges: SE Ads Are Not Traditional Ads

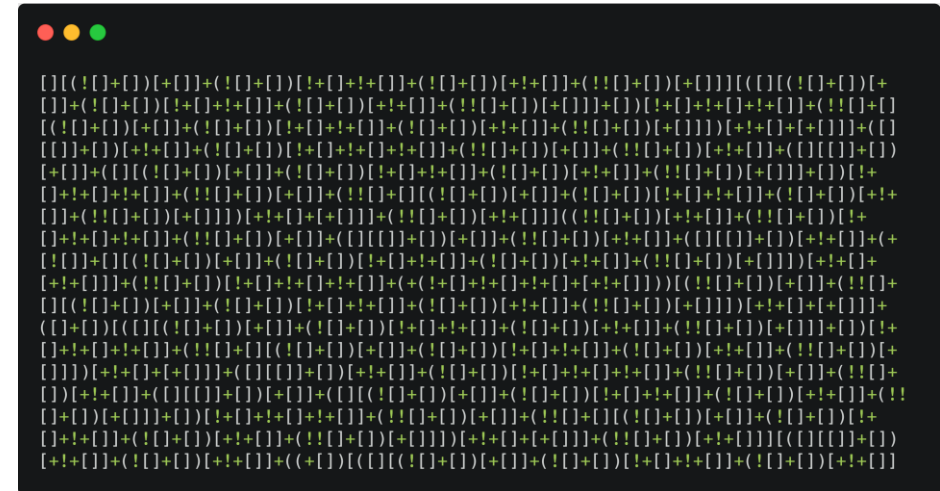
- Invisible on the DOM or misleading content.
- Obfuscated JavaScript code from low-tier ad networks.
- Frequently updated URLs.



Invisible



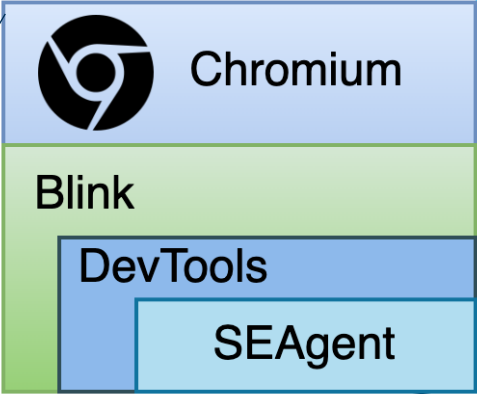
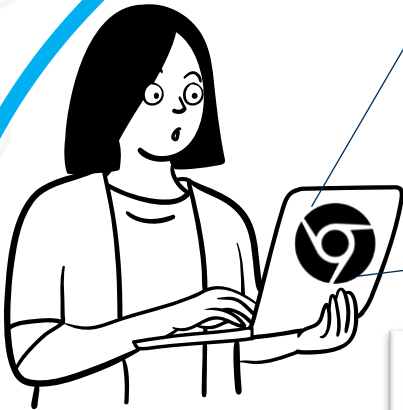
Misleading Content



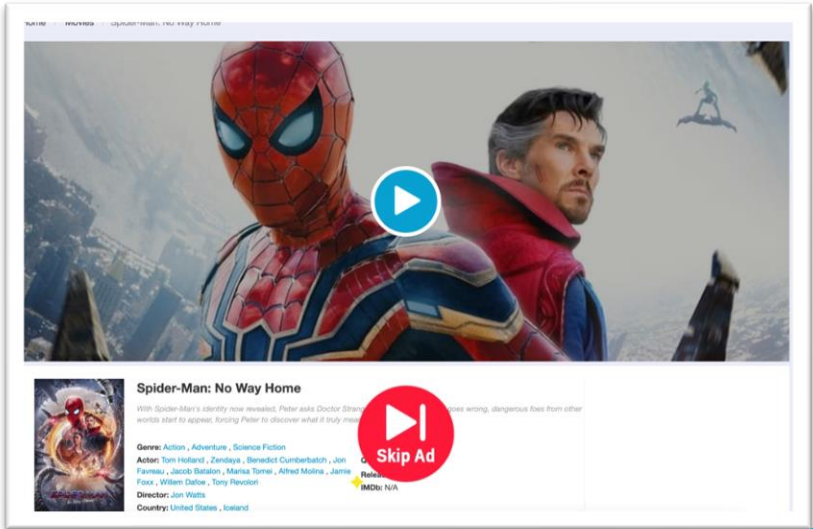
Obfuscated Code



# Design Overview



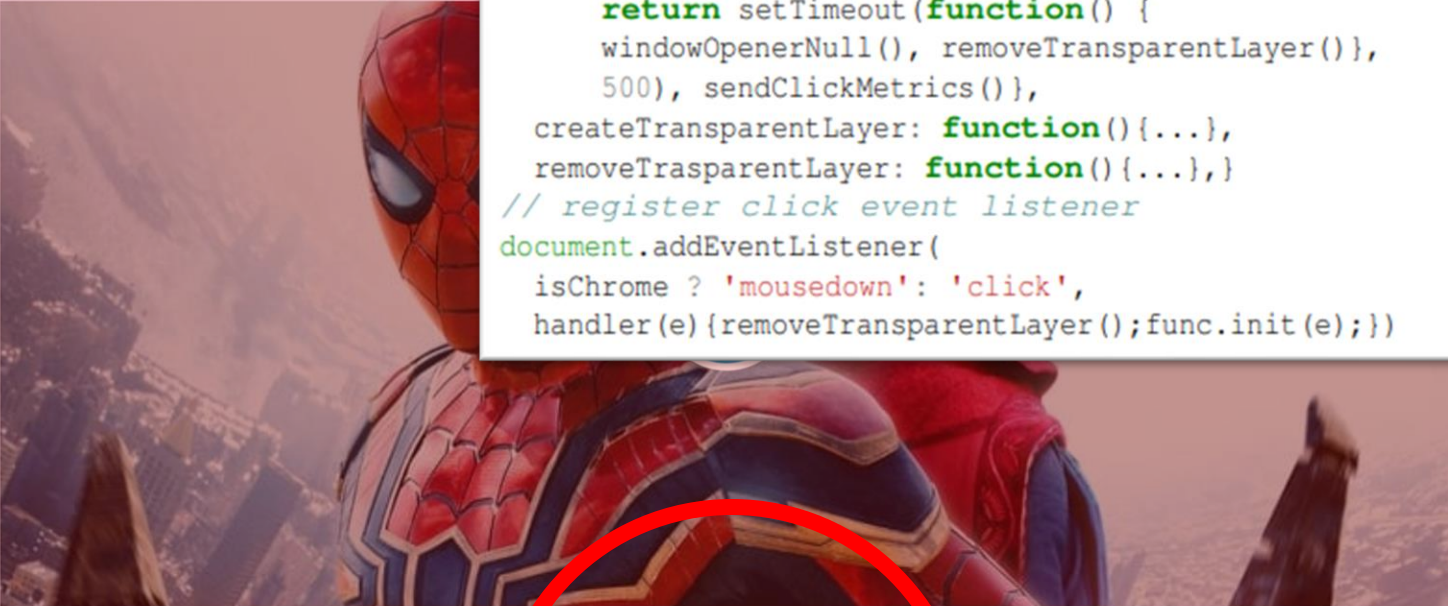
Navigation 



SE Attacks

# Design: Monitor JavaScript Actions

Home / Movies / Spider-Man: No Way Home



```
var func = {init: function(event) {  
    return setTimeout(function() {  
        windowOpenerNull(), removeTransparentLayer(),  
        500), sendClickMetrics()),  
    createTransparentLayer: function(){...},  
    removeTrasparentLayer: function(){...},  
    // register click event listener  
    document.addEventListener(  
        isChrome ? 'mousedown': 'click',  
        handler(e){removeTransparentLayer();func.init(e);})
```

**Spider-Man: No Way Home**

With Spider-Man's identity now revealed, Peter asks Doctor Strange for help. When a spell goes wrong, dangerous foes from other worlds start to appear, forcing Peter to discover what it truly means to be Spider-Man.

**Genre:** Action , Adventure , Science Fiction

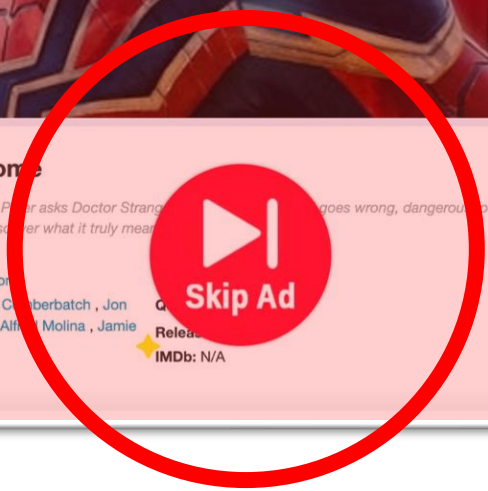
**Actor:** Tom Holland , Zendaya , Benedict Cumberbatch , Jon Favreau , Jacob Batalon , Marisa Tomei , Alfred Molina , Jamie Foxx , Willem Dafoe , Tony Revolori

**Director:** Jon Watts

**Country:** United States , Iceland

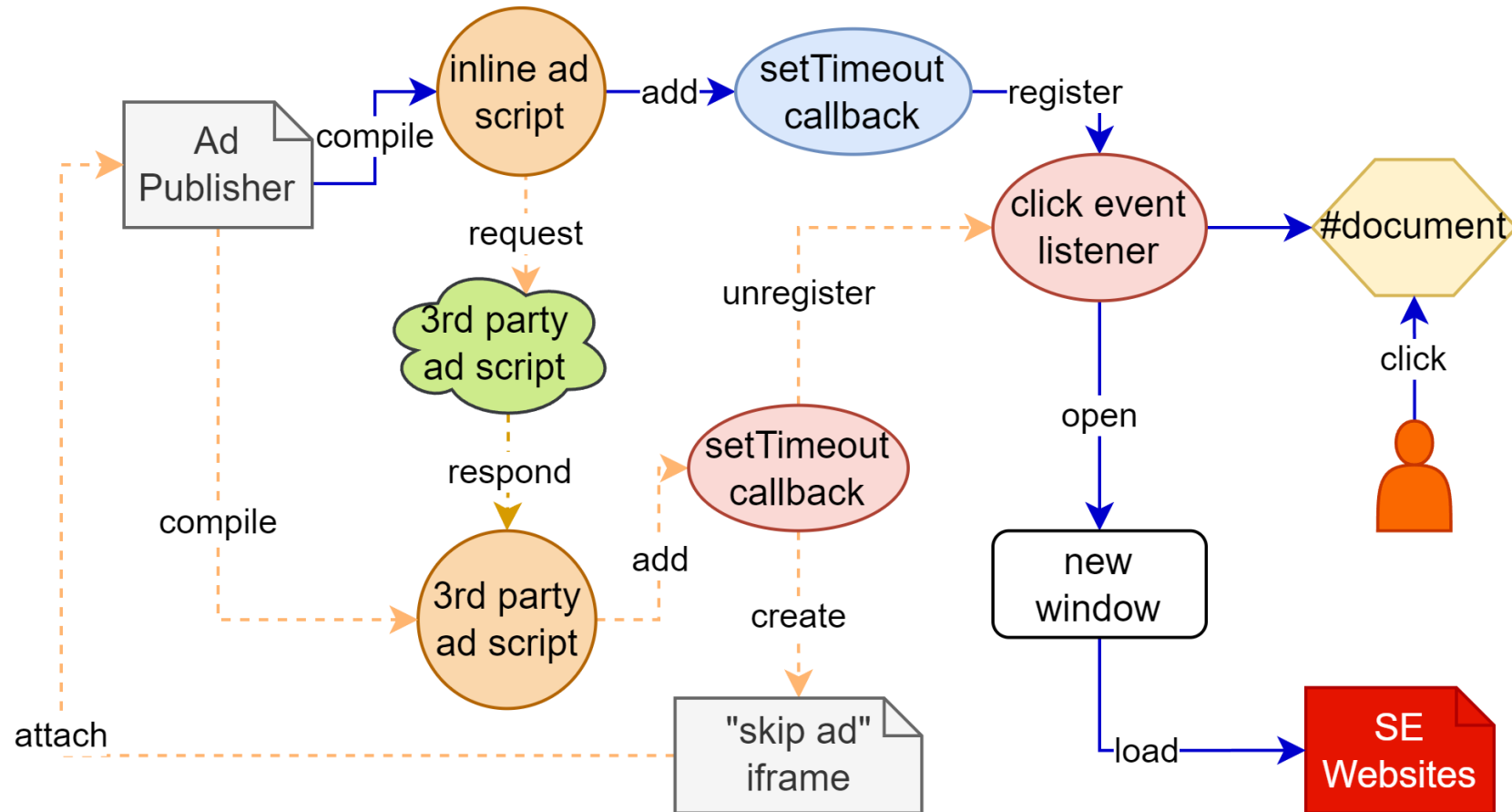
**Release Date:** ...

**IMDb:** N/A

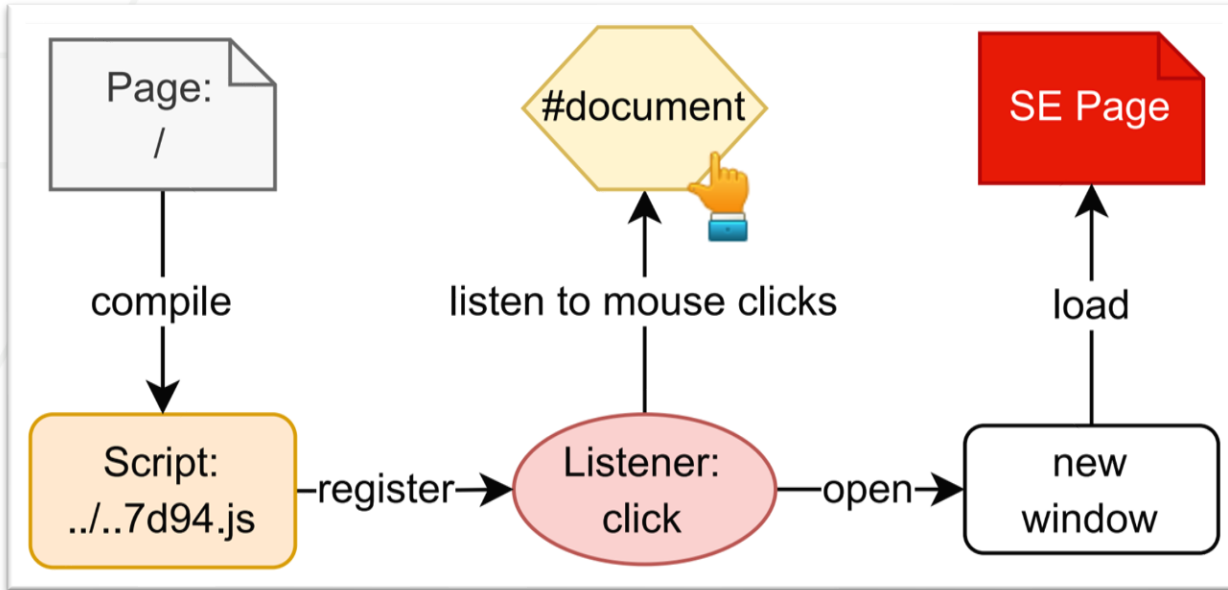


- JavaScript function calls
  - Callbacks
  - Event listeners
  - ...
- DOM manipulation
  - Create/modify/remove nodes
  - Open new tabs
  - ...
- Network communications
  - Request resources
  - Navigation requests
  - ...

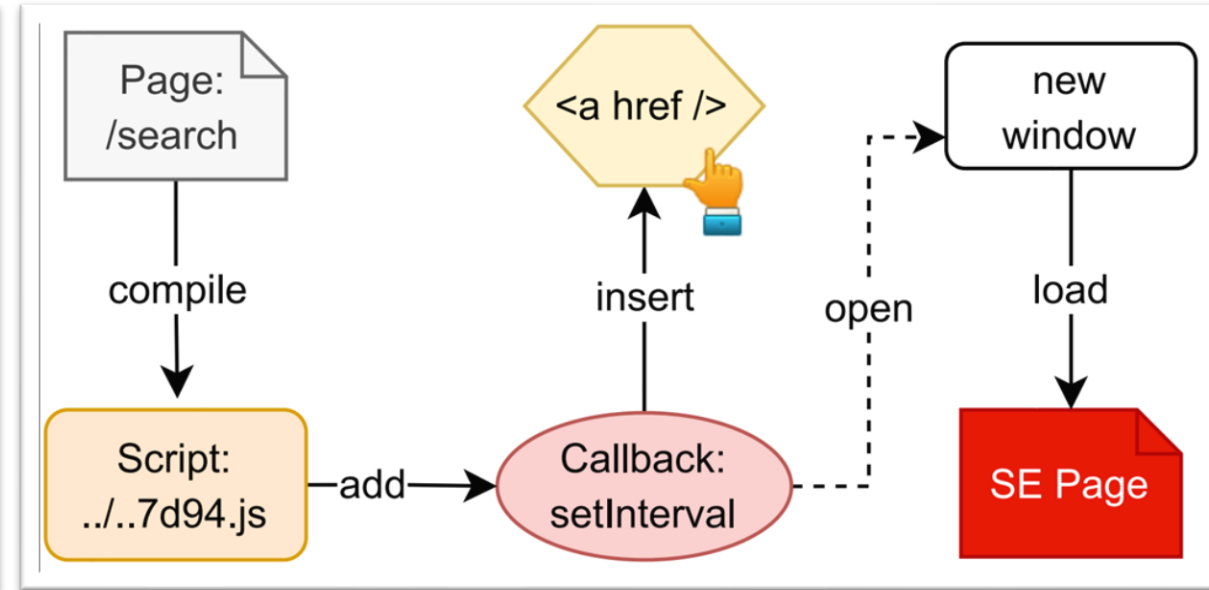
# Design: Web Action History Graph



# Design: Navigation Initiator



Initiated by A Mouse Event Listener



Initiated by Anchor Links

# Design: Social Engineering Features

## Property Features

- Describe what the code is.

script

execution context  
script type  
requestor  
...

## Action Features

- Monitor what the code does.

script

register

callback

open

window

## Consequence Features

- Observe what the code causes.

window

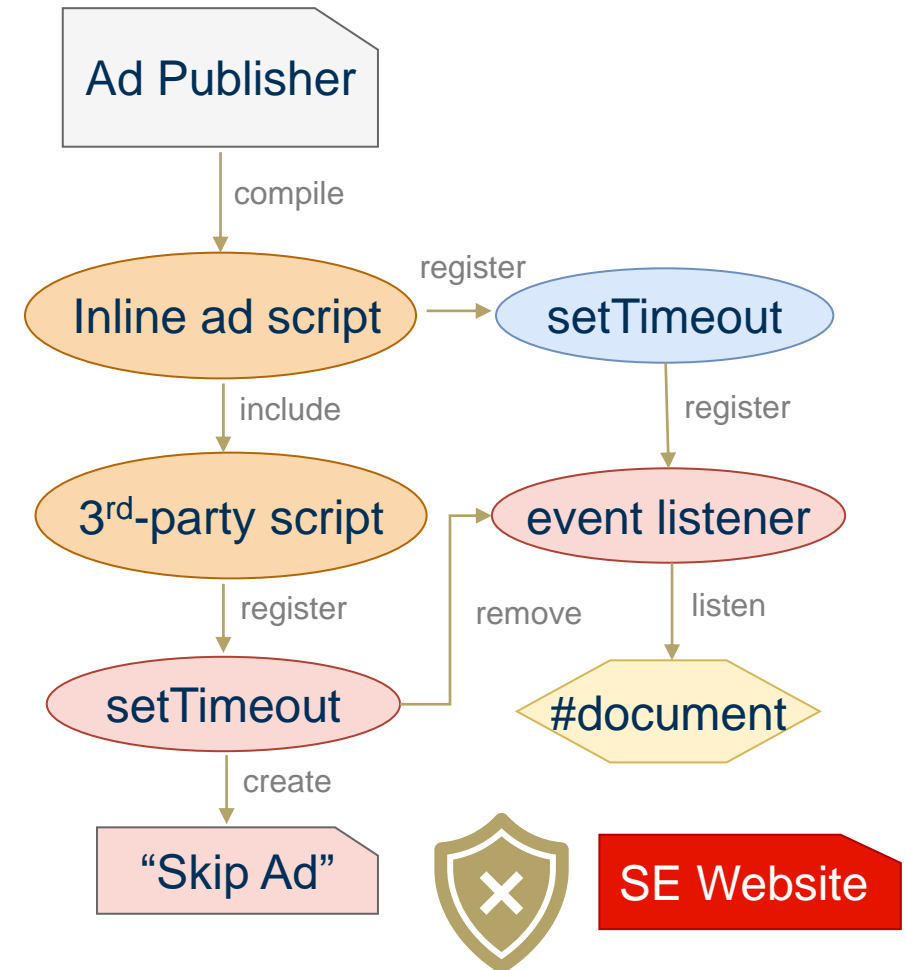
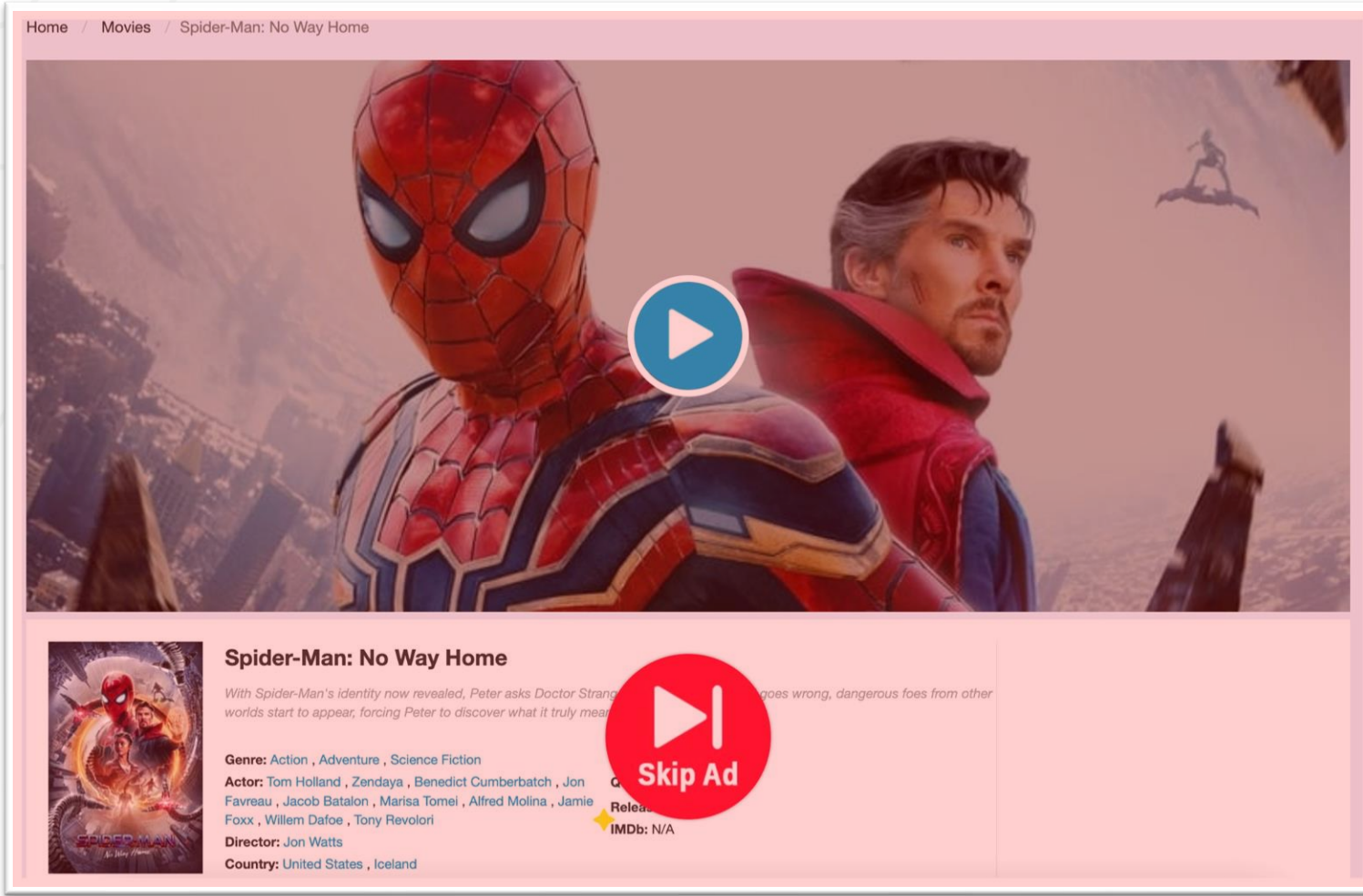
load

webpage

# of redirects  
domains



# Design: Block The Navigation



TRIDENT



# Evaluation

Oct. 2021

Started crawling **100,000+ ad publishers using low-tier ad networks to distribute** social engineering ads/attacks.

Oct. 2022

- Crawled another batch of testing data.
- Achieved **97.37% accuracy** and **97.81% F1 score**.

Jan. 2022

- Collected **259,487 navigation events**.
- **1,479** were labeled as **social engineering attacks**.
- Obtained **92.63% accuracy** and **93.37% F1 score** with a Random Forest classifier with 10-fold cross validation.

# Evaluation: Training with A Diverse Dataset

- Training in Jan. 2022
  - Covered more than 10 low-tier ad networks (e.g., AdSterra, PopCash) and top-tier ad networks (e.g., Google, Facebook, Amazon).
  - Found 6 types of social engineering attacks with a semi-auto labeling technique.
    - 857 Unwanted-software Download
    - 222 Dating Scam
    - 177 Prize Scam
    - 148 Push Notification Spam
    - 51 Scareware
    - 24 Tech-support Scam

# Evaluation: Performance Over Time

- Testing in Oct. 2022
  - 2.57% false positive rate.
    - Inject DOM elements for benign purpose such as AddThis.
    - Inject social engineering ads, but do not take the user to social engineering websites.
    - Inject social engineering ads and take the user to adult websites which do not launch social engineering attacks immediately when labeling.
  - 0.13% false negative rate.
    - Only 1 case that embedded a malicious link as an image in the first party website. That link leads to a malicious software download website.
  - **Detected** social engineering attacks distributed by **two unseen low-tier ad networks**.
    - PopAds – 2 SE attacks out of 296 navigation events.
    - PopMyAds – 2 SE attacks out of 349 navigation events.



# Why The Performance Went Up?

New-tab Nav.	Same-tab Nav.	Accuracy	Precision	Recall	F-1 Score
100%	0%	87.76%	86.69%	89.31%	87.98%
90%	10%	88.30%	86.09%	91.68%	88.80%
<b>50%</b>	<b>50%</b>	<b>92.63%</b>	<b>90.63%</b>	<b>96.28%</b>	<b>93.37%</b>
0%	100%	99.76%	99.78%	99.43%	99.60%
Random Sampling		99.36%	99.14%	99.59%	98.17%
No Sampling		97.69%	89.71%	76.39%	82.52%

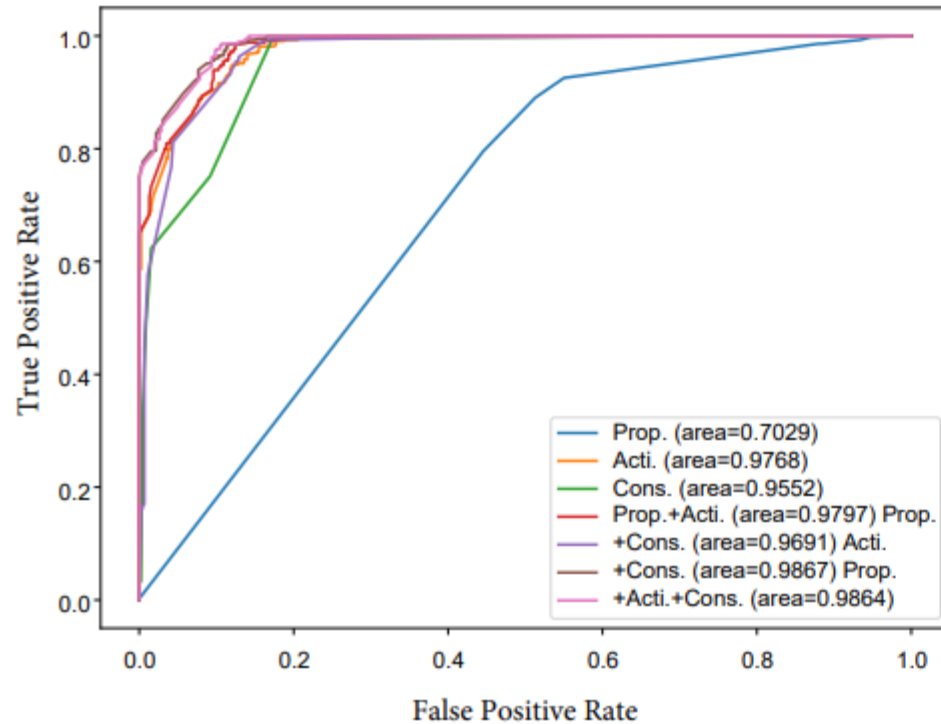
Poor performance

Overfitting

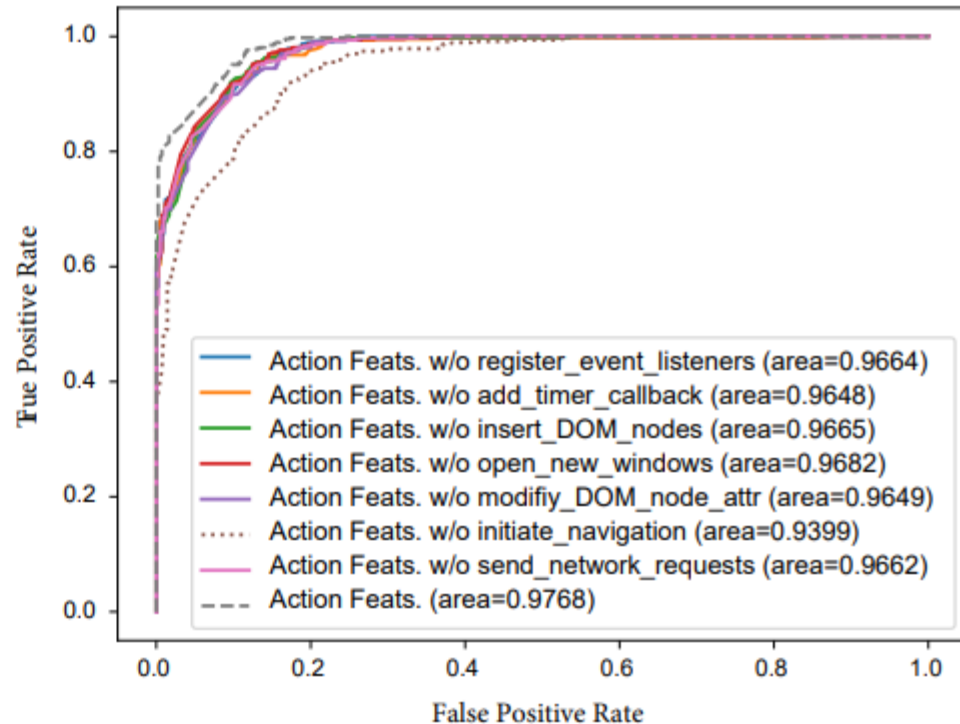
We choose a conservative model for focusing on data points near the borderline. More details are available in the paper.

# Feature Importance

- Evaluated by the Leave-One-Group-Out approach
  - Action + Consequence feature groups perform the best (AUC=0.9867)
  - Property + Action + Consequence features groups perform more robust (AUC= 0.9864).



Feature Importance by Groups



Feature Importance within Action Feature Group

# Evaluation: Evasion Attempts

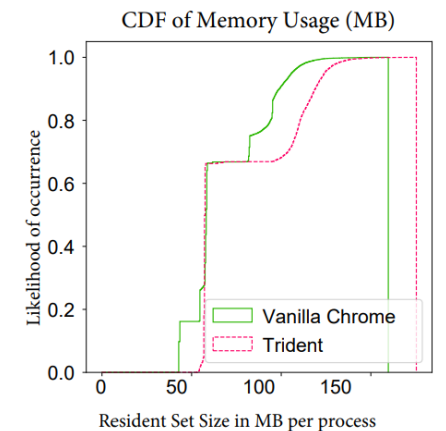
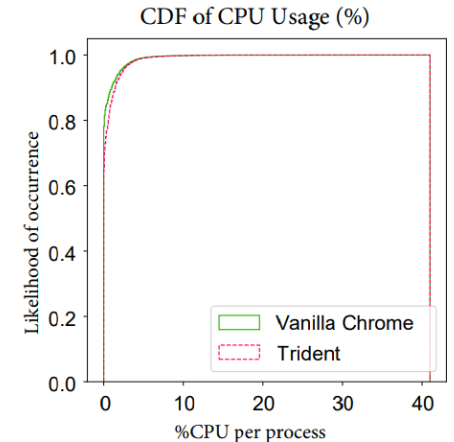
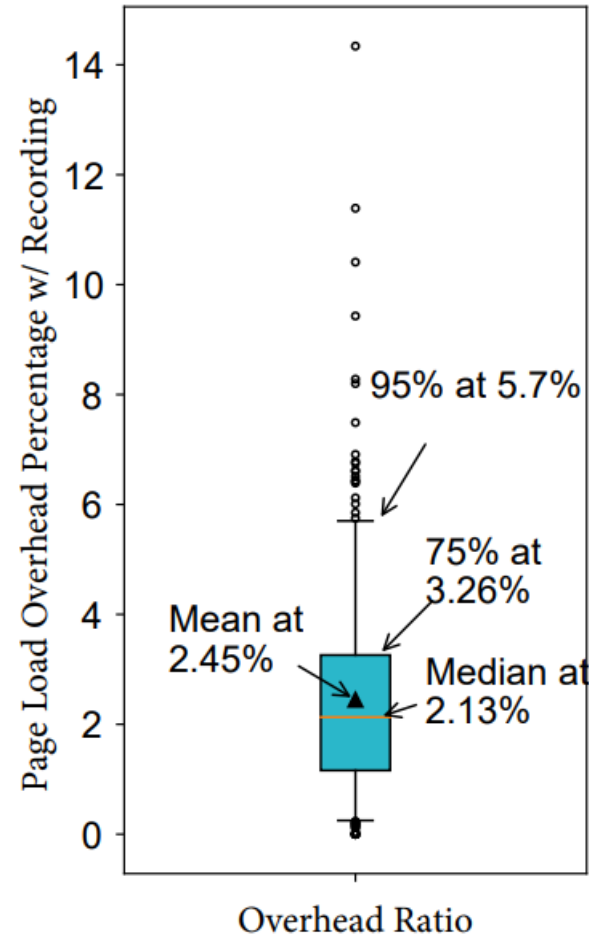
- Include the malicious script as the **first-party script**.
- Put the malicious script as **an inline script (implying first-party)**.
- Directly take the user to social engineering websites **without redirects**.
- **Behave as benign** scripts while stealing clicks.

Approaches	Evasion Rate
First-party script (Fst.Pty.)	2.13%
Inline script (Inl.)	5.11%
No redirects (NoRdr.)	3.62%
NoRdr. + Fst.Pty.	2.56%
NoRdr. + Inl. + Fst.Pty.	9.17%
Do not request external resources	1.49%
Do not add callbacks	1.49%
Do not attach iframes	1.92%
Do not modify node attributes	1.70%



# Runtime Overhead

- Event monitoring agent implemented in Chrome DevTools Protocol with <800 lines of C++ code.
- 2.13% runtime overhead when browsing the Internet.
- Negligible resource overhead.



# Conclusion

- A novel **online** system for **indirectly** detecting and blocking social engineering attacks.
- **92.63% accuracy**, which **outperforms the state-of-the-art** generic ad-blocking tools **by more than 10%** with negligible runtime overhead.
- **Robust to evasion attempts.**

# Q & A

Zheng Yang  
[ianyang@gatech.edu](mailto:ianyang@gatech.edu)  
<https://ian.yang.bio>