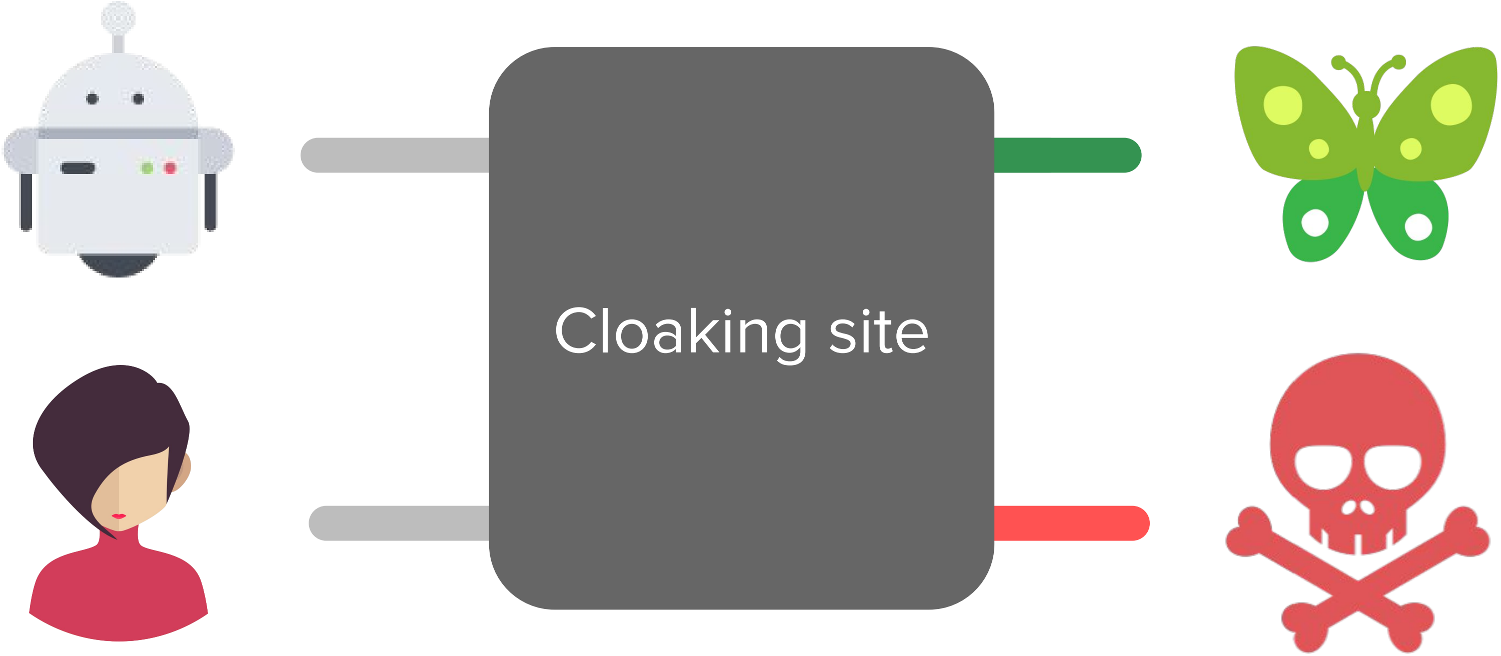
Cloak of Visibility: Detecting When Machines Browse a Different Web

Luca Invernizzi*, Kurt Thomas*, Alexandros Kapravelos†, Oxana Comanescu*, Jean-Michel Picod*, and Elie Bursztein*

* Google - Anti-fraud and abuse research

† North Carolina State University

Web cloaking



Cloaking site

Web cloaking

Search

Effective for Search Engine Optimization

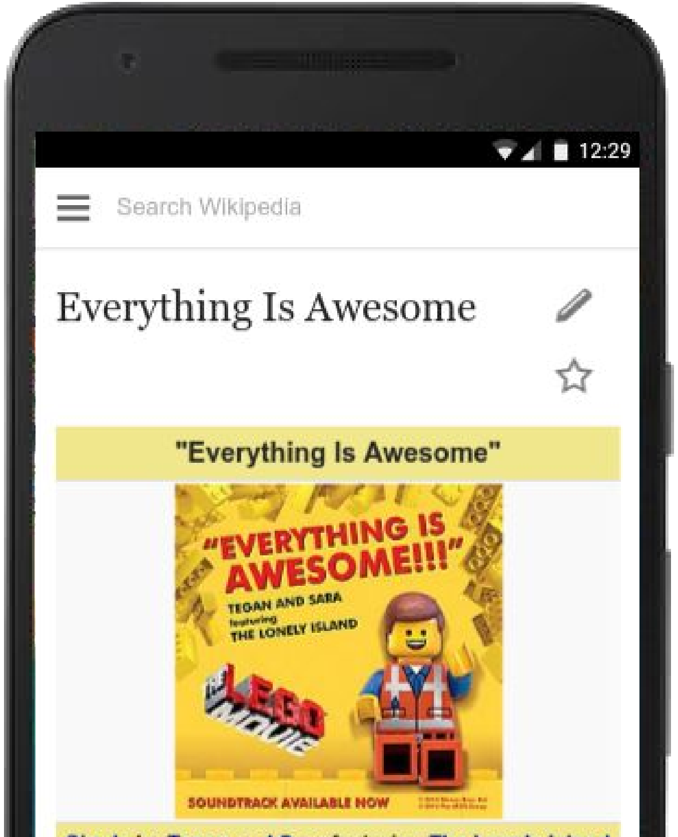
Ads

Effective to infringe policies

~~Malware~~

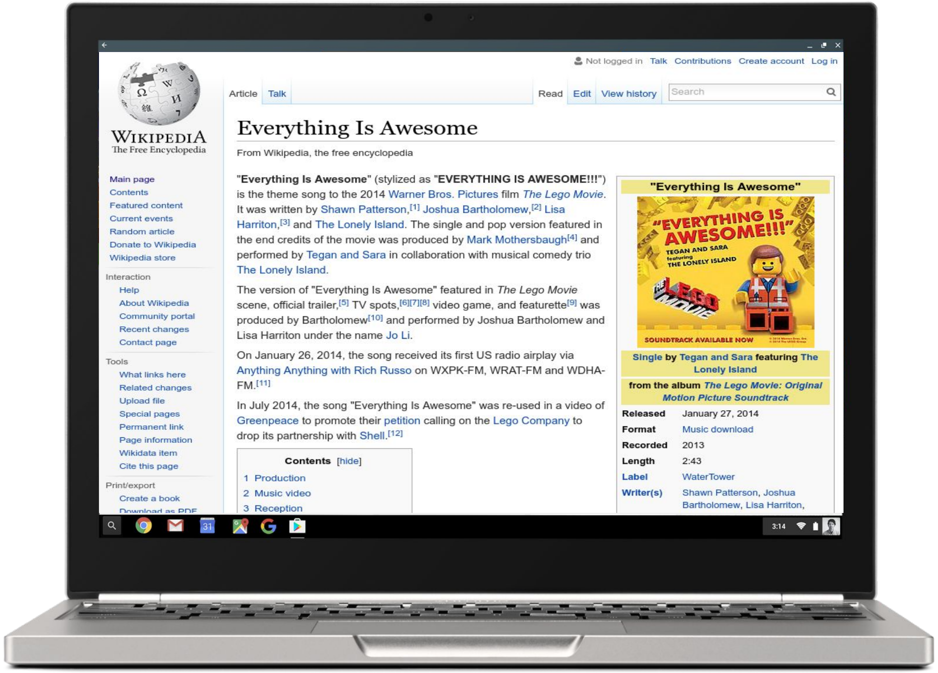
~~Effective to evade security crawlers~~

Responsive design vs cloaking



Everything Is Awesome

"Everything Is Awesome"



Everything Is Awesome

From Wikipedia, the free encyclopedia

"**Everything Is Awesome**" (stylized as "**EVERYTHING IS AWESOME!!!**") is the theme song to the 2014 Warner Bros. Pictures film *The Lego Movie*. It was written by Shawn Patterson,[1] Joshua Bartholomew,[2] Lisa Harriton,[3] and The Lonely Island. The single and pop version featured in the end credits of the movie was produced by Mark Mothersbaugh[4] and performed by Tegan and Sara in collaboration with musical comedy trio The Lonely Island.

The version of "Everything Is Awesome" featured in *The Lego Movie* scene, official trailer,[5] TV spots,[6][7][8] video game, and featurette[9] was produced by Bartholomew[10] and performed by Joshua Bartholomew and Lisa Harriton under the name Jo Li.

On January 26, 2014, the song received its first US radio airplay via Anything Anything with Rich Russo on WXPK-FM, WRAT-FM and WDHA-FM.[11]

In July 2014, the song "Everything Is Awesome" was re-used in a video of Greenpeace to promote their petition calling on the Lego Company to drop its partnership with Shell.[12]

Contents [hide]

1 Production
2 Music video
3 Reception

"Everything Is Awesome"

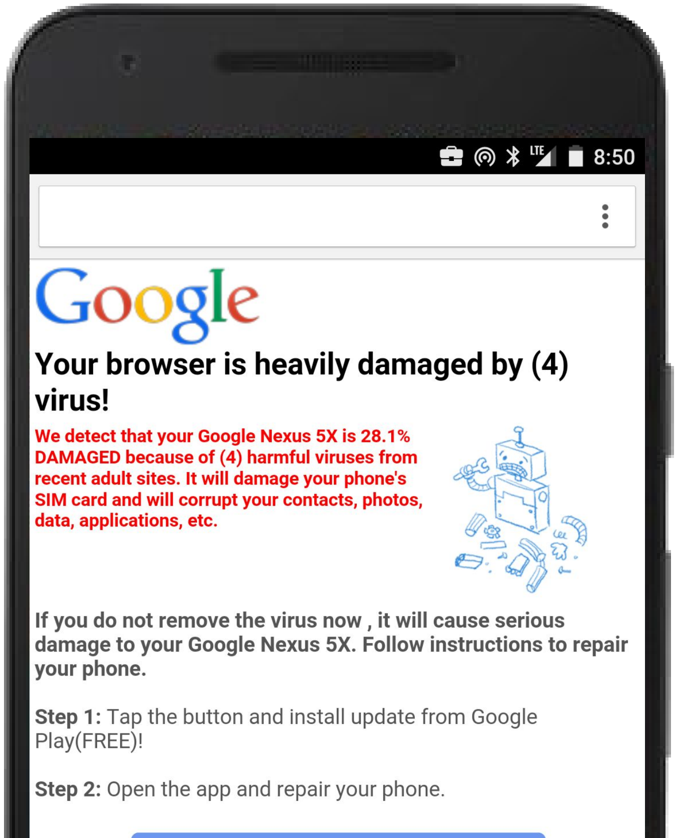
Single by Tegan and Sara featuring The Lonely Island

from the album *The Lego Movie: Original Motion Picture Soundtrack*

Released	January 27, 2014
Format	Music download
Recorded	2013
Length	2:43
Label	WaterTower
Writer(s)	Shawn Patterson, Joshua Bartholomew, Lisa Harriton,

This is **not cloaking**.

Responsive design vs cloaking



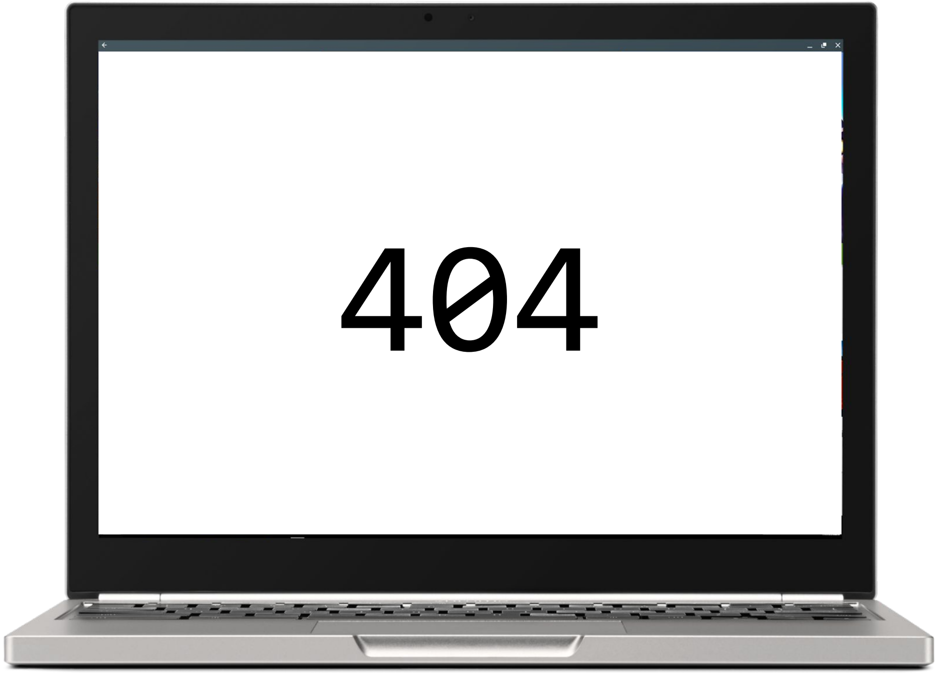
Your browser is heavily damaged by (4) virus!

We detect that your Google Nexus 5X is 28.1% DAMAGED because of (4) harmful viruses from recent adult sites. It will damage your phone's SIM card and will corrupt your contacts, photos, data, applications, etc.

If you do not remove the virus now , it will cause serious damage to your Google Nexus 5X. Follow instructions to repair your phone.

Step 1: Tap the button and install update from Google Play(FREE)!

Step 2: Open the app and repair your phone.



404

This is **cloaking**.

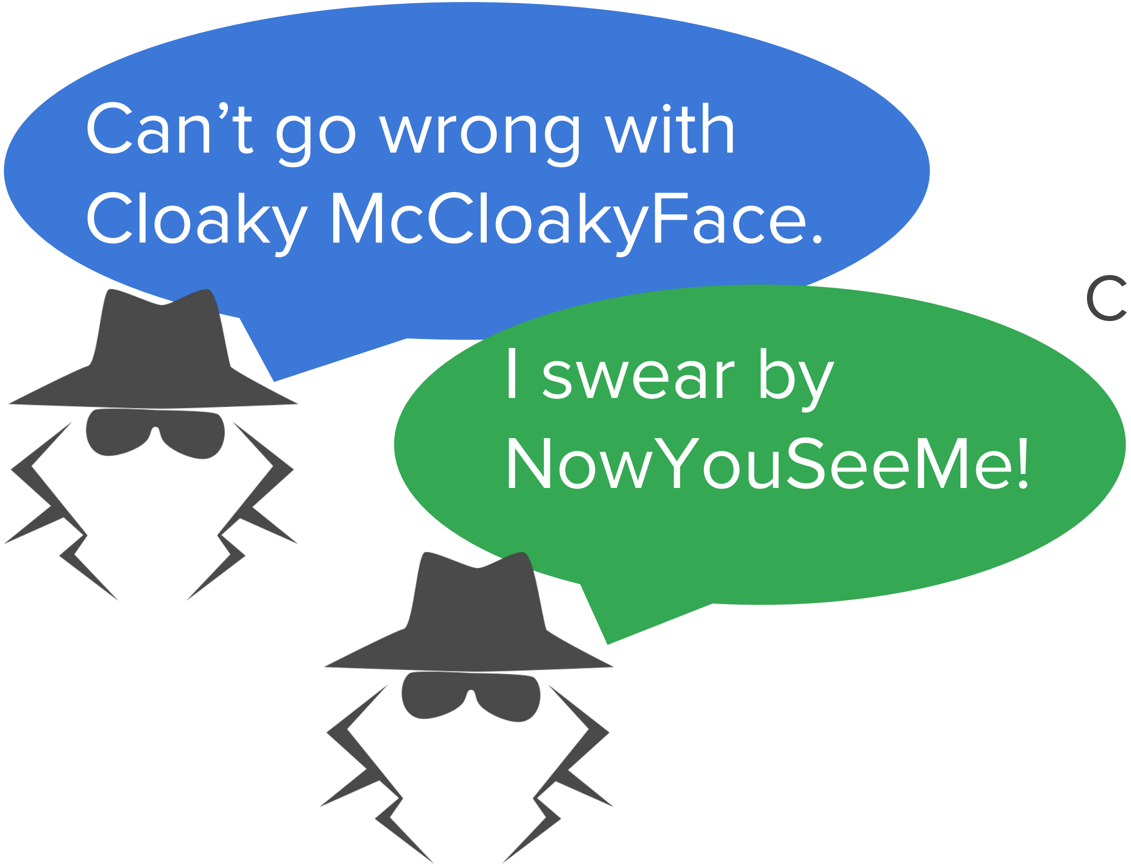
Research goals

Keep up with arms race

Identify trends

Explore alternatives

Blackmarket Investigation

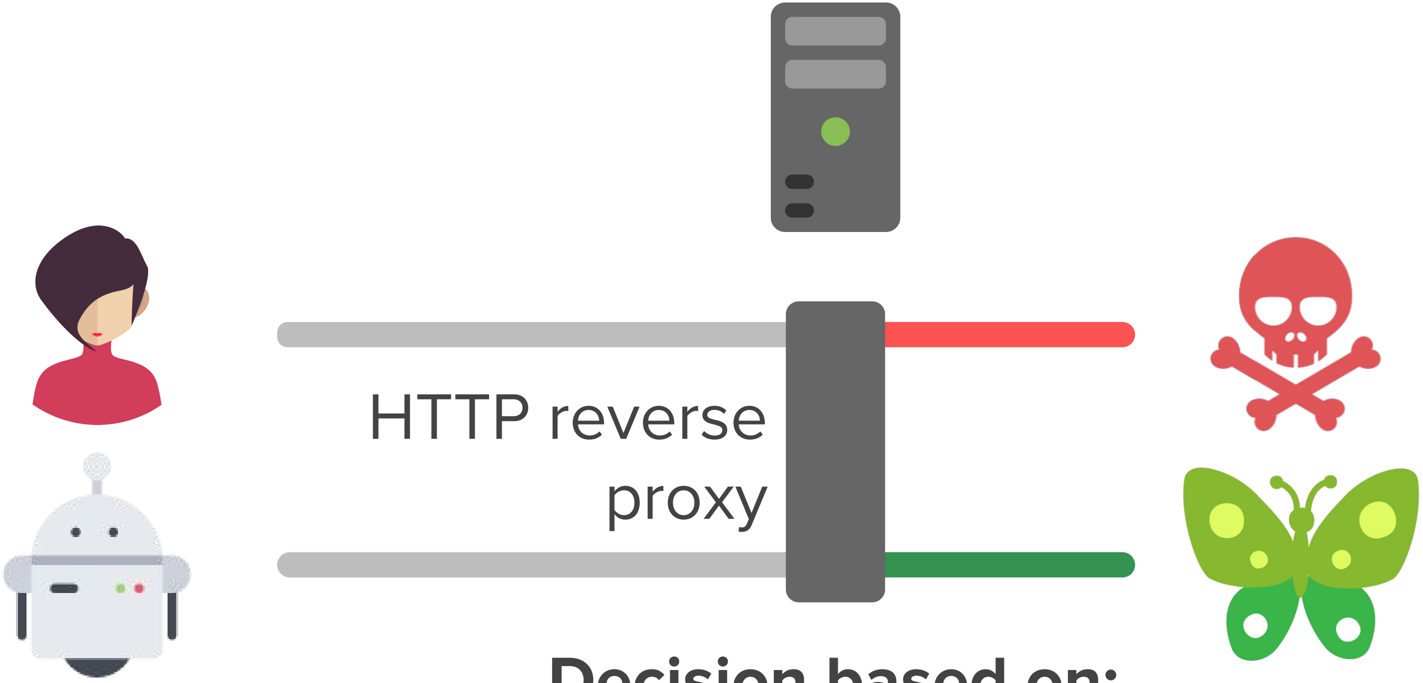


Acquired

Top 10

Cloaking software samples

$3500+ cloaking software



HTTP reverse proxy

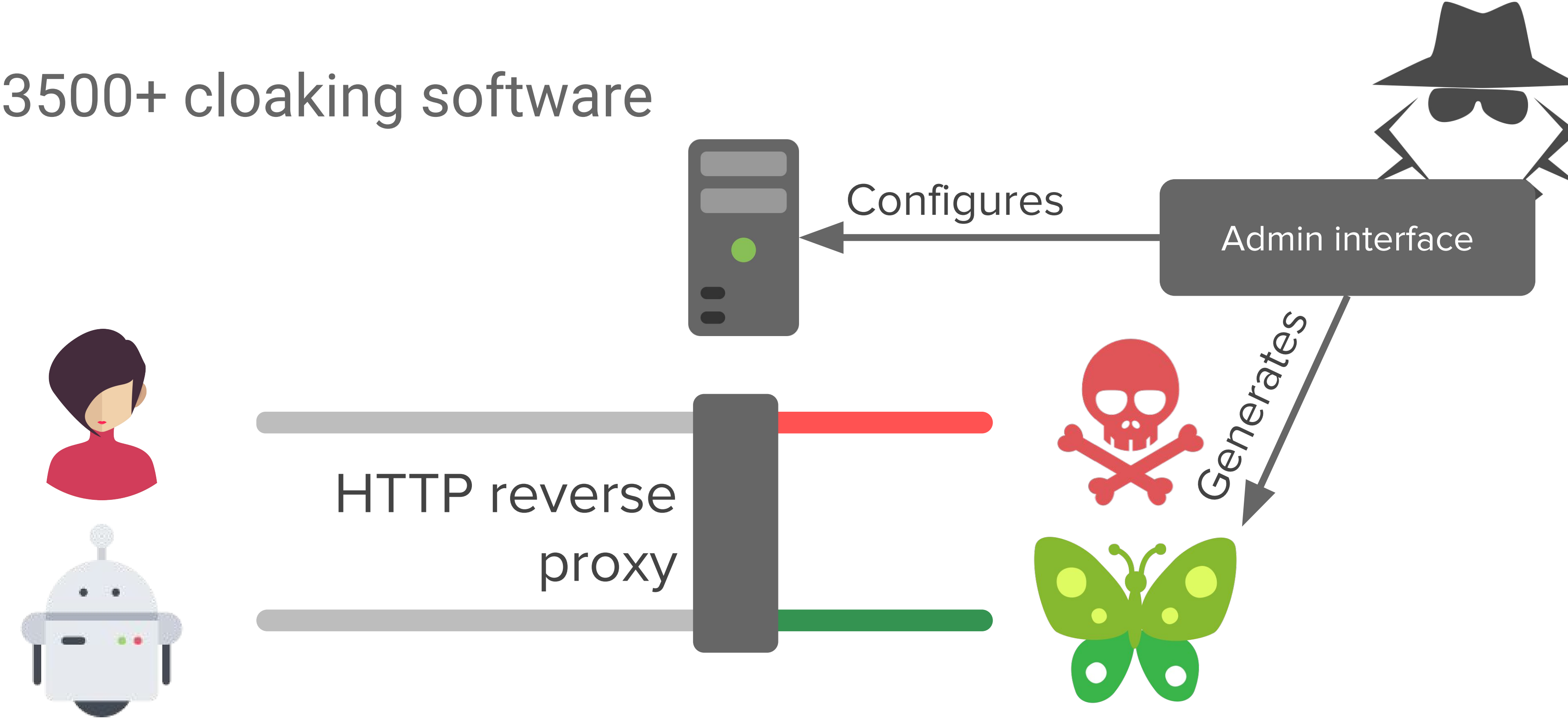
Decision based on:

Network

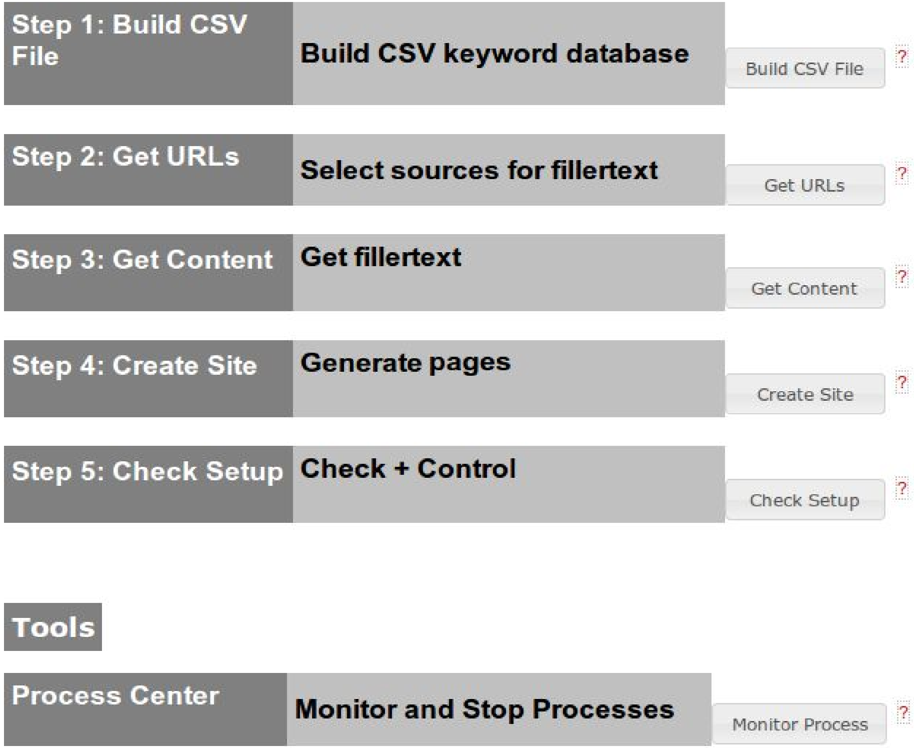
Browser

Browsing context

$3500+ cloaking software



Admin interface



Step 1: Build CSV File	**Build CSV keyword database**	Build CSV File	?
Step 2: Get URLs	**Select sources for fillertext**	Get URLs	?
Step 3: Get Content	**Get fillertext**	Get Content	?
Step 4: Create Site	**Generate pages**	Create Site	?
Step 5: Check Setup	**Check + Control**	Check Setup	?

Tools

Process Center	**Monitor and Stop Processes**	Monitor Process	?

Input

keywords => http://money.site

Features

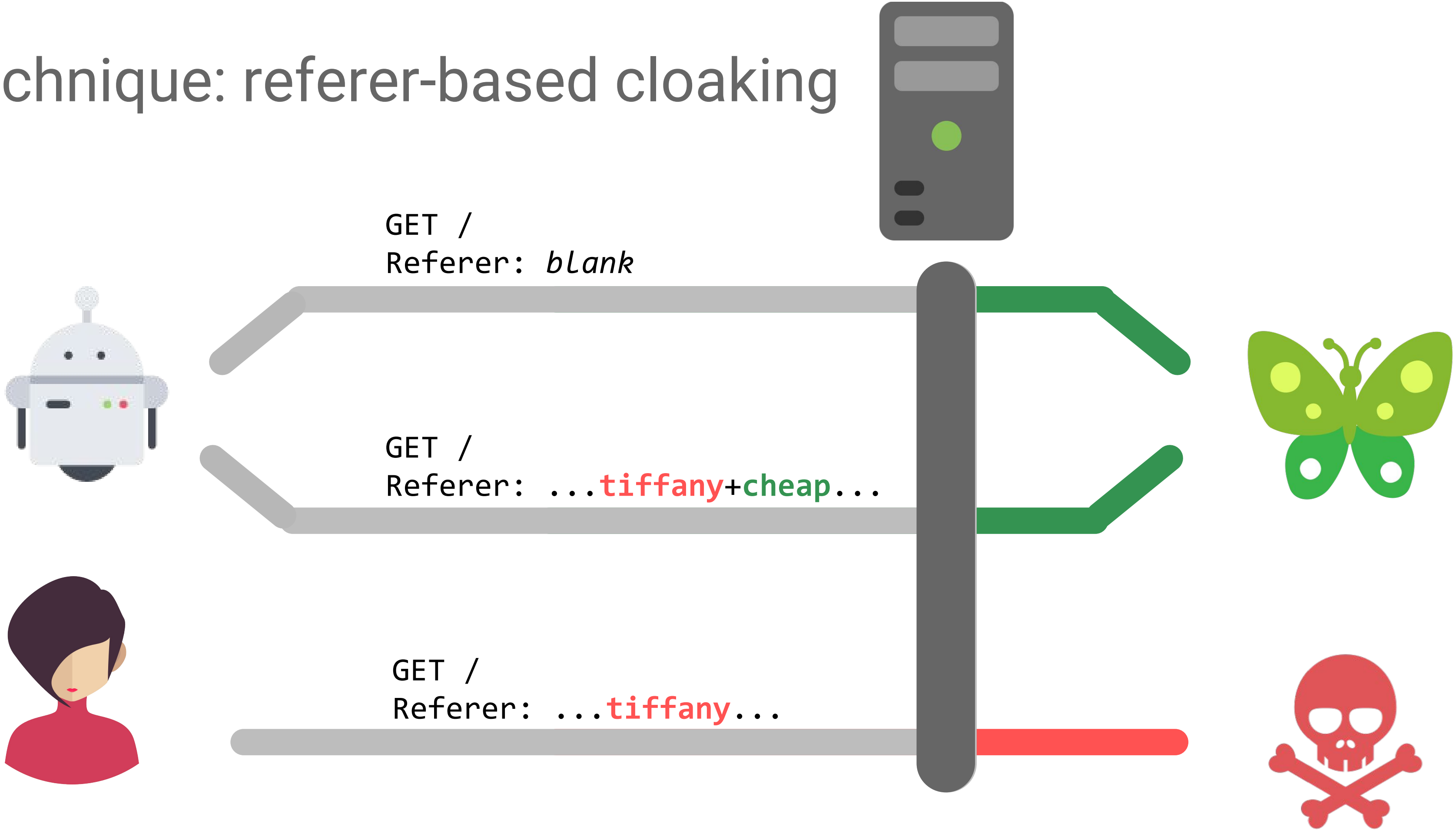
- Find similar sites through SERPs
- Content/Template spinning
- Drip-feeding

Added services

- Plagiarism detection
- SERP ranking

Cloaking techniques

Technique: referer-based cloaking



```
GET /
Referer: *blank*
```

```
GET /
Referer: ...tiffany+cheap...
```

```
GET /
Referer: ...tiffany...
```

Technique: IP blacklisting

51m

Blacklisted IPs

30

Security companies

3

Proxy networks

983

Subnets

2

Hacking collectives

122

Entities: companies, universities, registrars

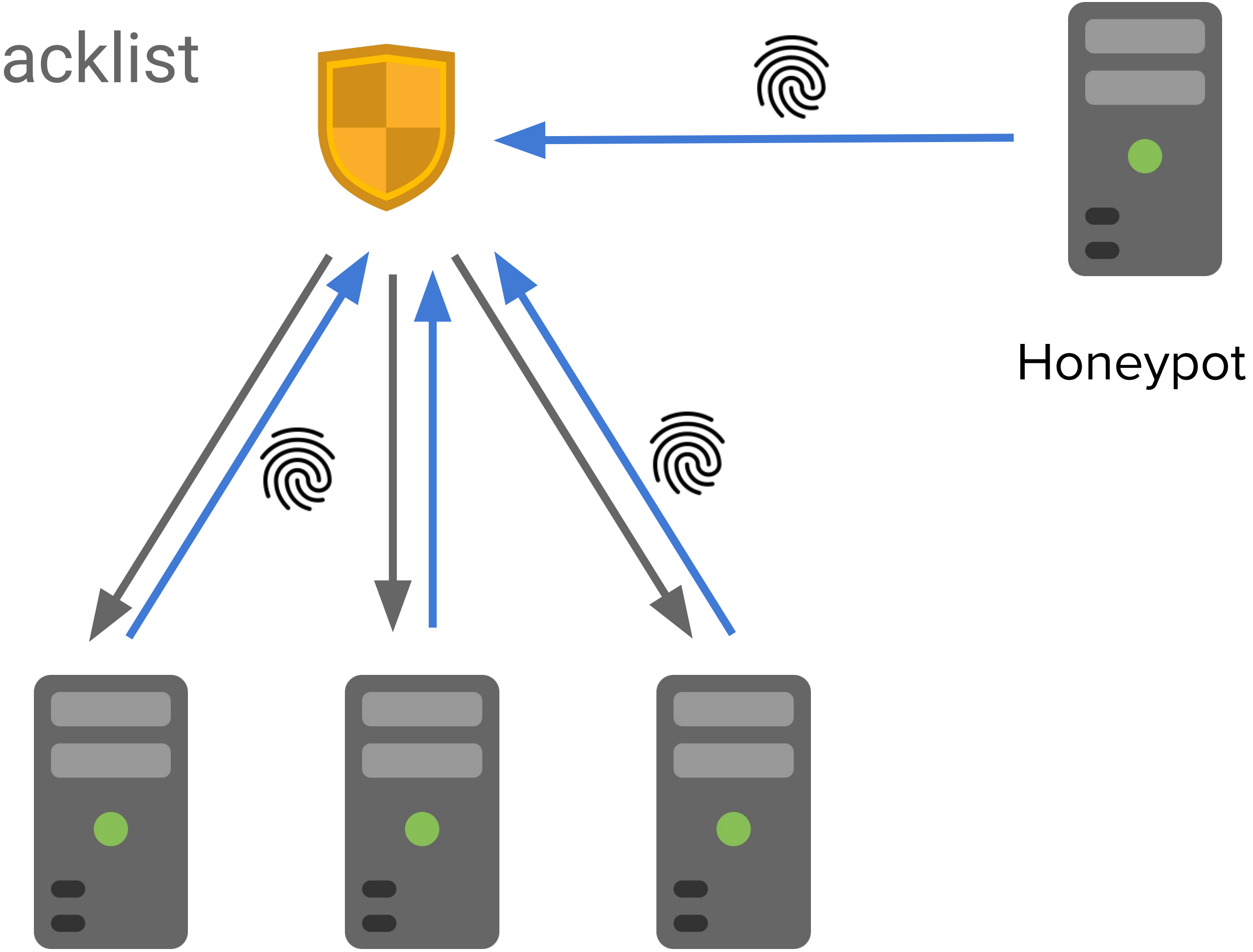
Crowdsourced blacklist

50k

Blacklisted IPs

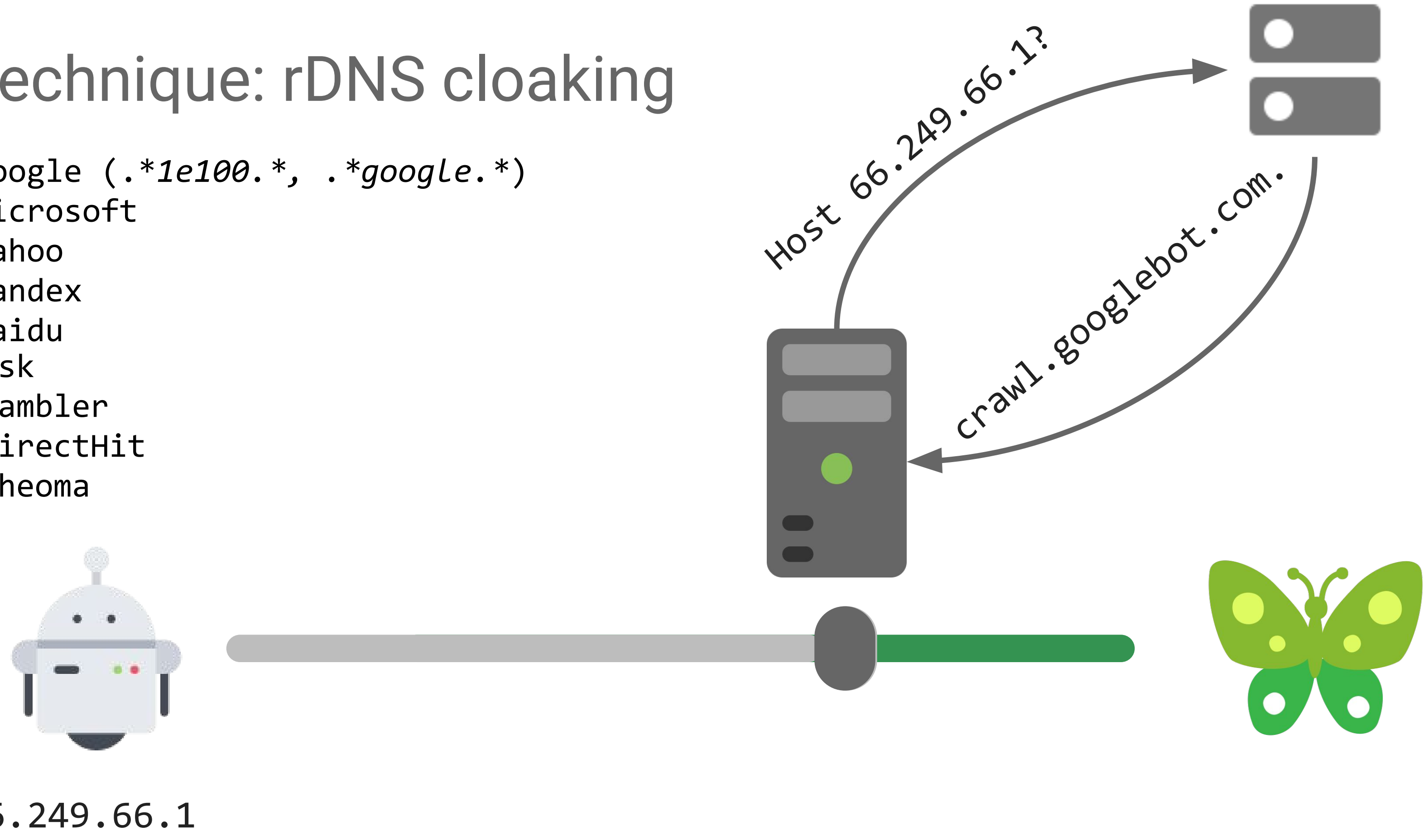
$350+

Subscription

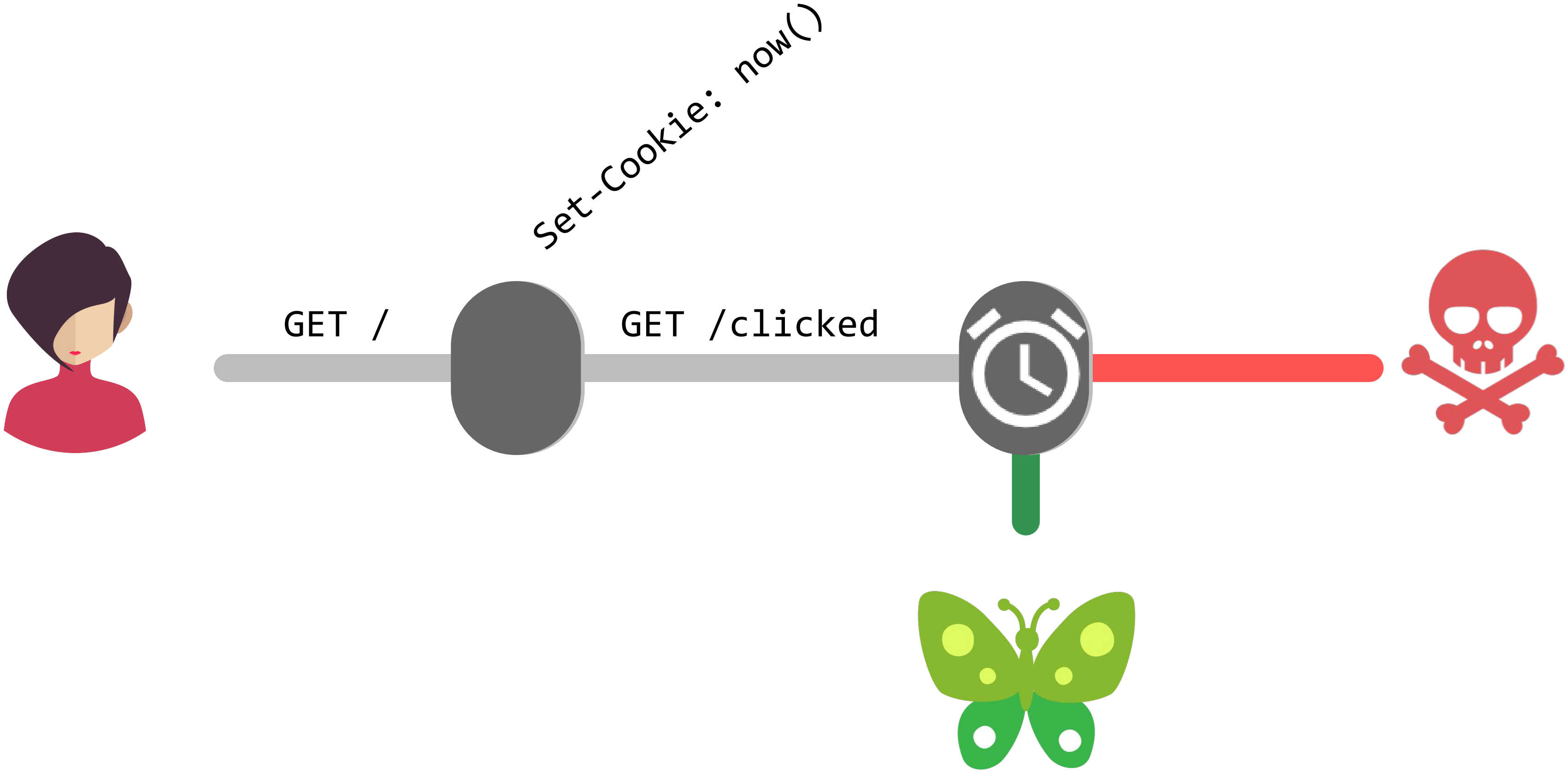


Technique: rDNS cloaking

Google (.*1e100.*, .*google.*)
Microsoft
Yahoo
Yandex
Baidu
Ask
Rambler
DirectHit
Theoma



Technique: browsing pattern cloaking



GET /

Set-Cookie: now()

GET /clicked

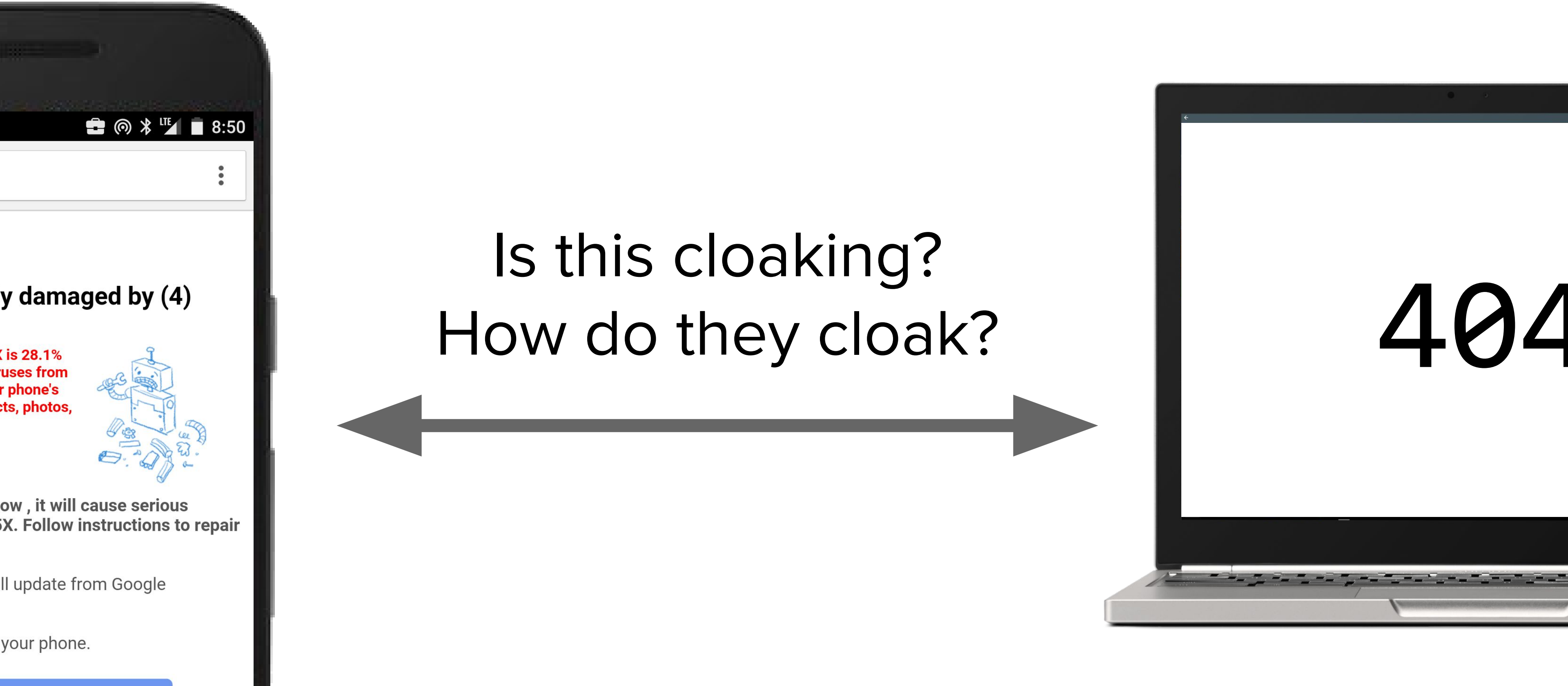
More techniques

Geolocation: country, city, carrier level.

Flash/JS support & fingerprints

User-Agent

Prevalence and dominant techniques



y damaged by (4)

X is 28.1%
uses from
r phone's
cts, photos,

ow , it will cause serious
5X. Follow instructions to repair

ll update from Google

your phone.

Is this cloaking?
How do they cloak?

Browser farm

Pretend Google bots

User-Agent: GoogleBot

Referer: blank

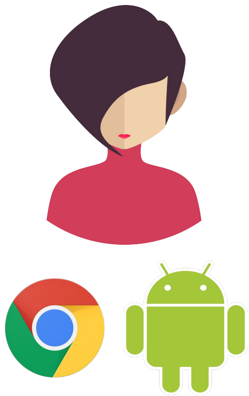
Google IP

Simple honey clients

User-Agent: Chrome

Referer: blank, or simple

Cloud provider IPs



Realistic honey clients

User-Agent: Chrome

Referer: context-aware

Residential and mobile IPs

Features

	HTML	Image
Syntactic	Content similarity	Screenshot similarity
Semantic	Topic similarity	Screenshot topic similarity

Classification

82%
True positive rate

.9%
False positive rate

95k labeled samples
75k legitimate websites (Alexa) + 20k cloaked storefronts

Prevalence

4.9%

Cloaking pages in **Google AdWords**, for health and software ads.

11.7%

Cloaking pages in **Google Search**, for luxury storefronts keywords.

Traditional techniques: only IP, Referer, and User-Agent

Search: 1 out of 5
Ads: 1 out of 4

Current techniques: JavaScript support

Search: Half

Ads: 1 out of 4

Current techniques: wait for click

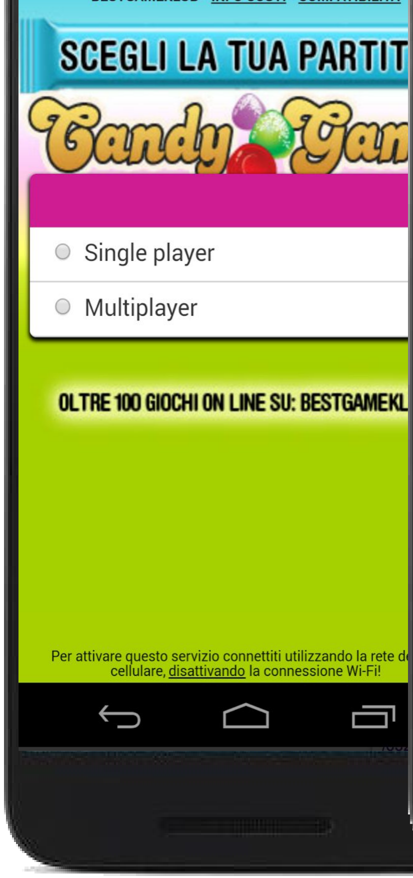
Search: 1 out of 10

Ads: 1 out of 5

Delivery: same-page cloaking

Search: 1 out of 5

Ads: **2 out of 3**



SCEGLI LA TUA PARTIT

Candy Gam

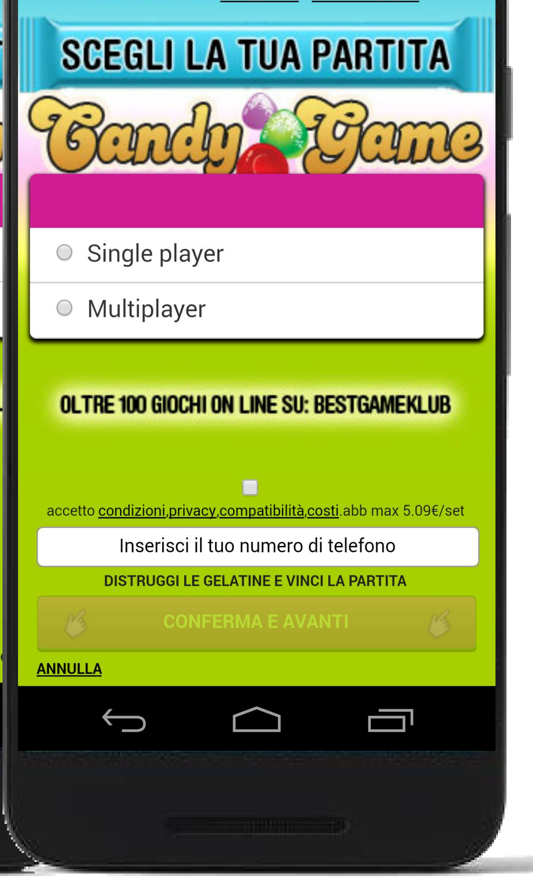
○ Single player

○ Multiplayer

OLTRE 100 GIOCHI ON LINE SU: BESTGAMEKL

Per attivare questo servizio connettiti utilizzando la rete de cellulare, disattivando la connessione Wi-Fi!

Uncloaked



SCEGLI LA TUA PARTITA

Candy Game

○ Single player

○ Multiplayer

OLTRE 100 GIOCHI ON LINE SU: BESTGAMEKLUB

□

accetto condizioni,privacy,compatibilità,costi.abb max 5.09€/set

Inserisci il tuo numero di telefono

DISTRUGGI LE GELATINE E VINCI LA PARTITA

CONFERMA E AVANTI

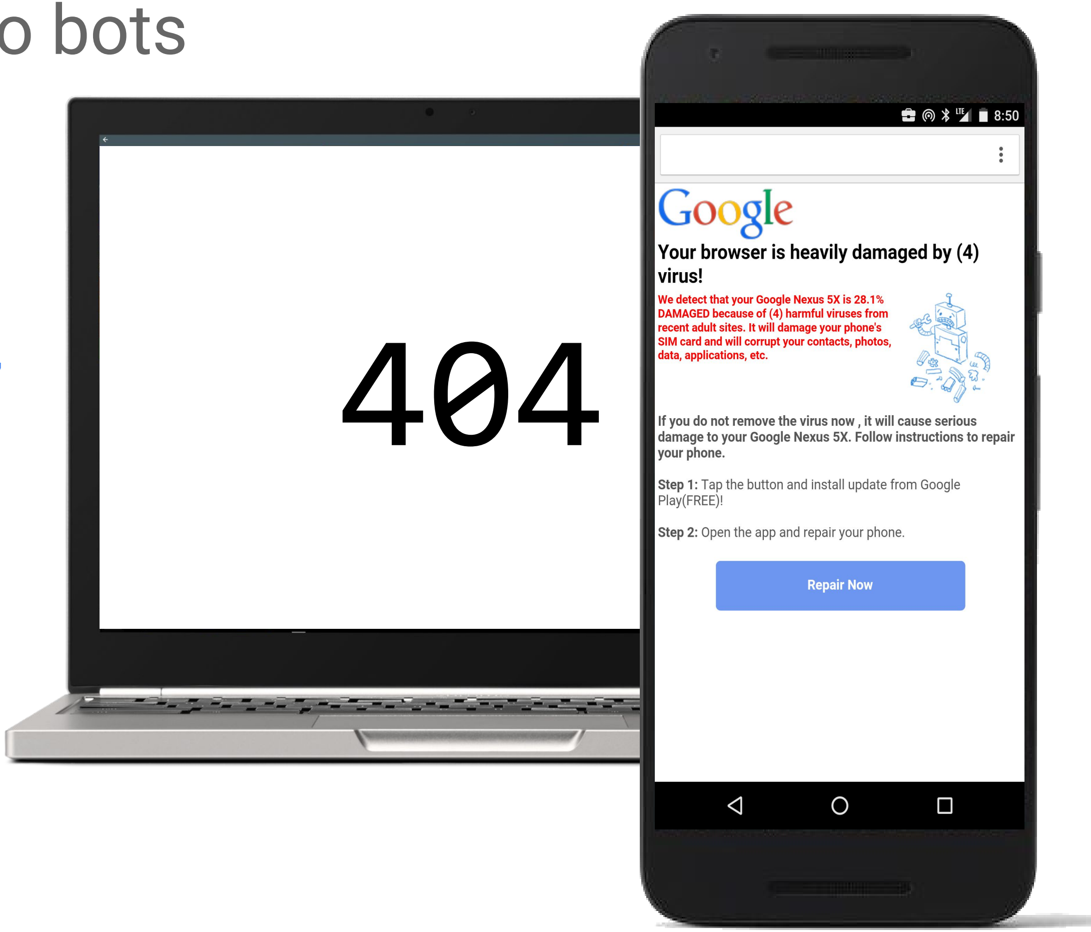
ANNULLA

Cloaked

Delivery: 40x/50x errors to bots

Search: 1 out of 7

Ads: 1 out of 8



Your browser is heavily damaged by (4) virus!

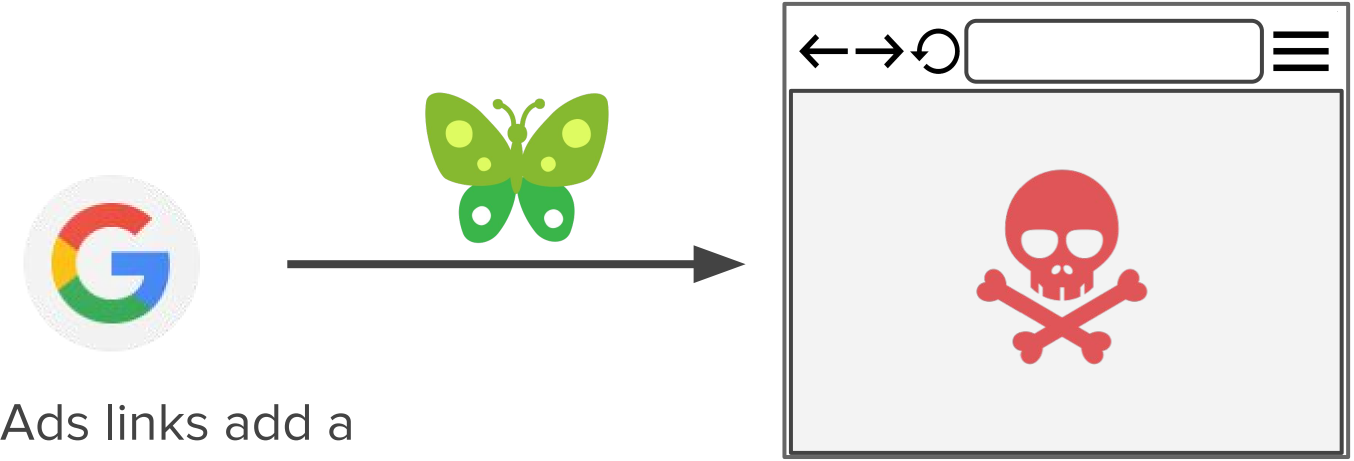
We detect that your Google Nexus 5X is 28.1% DAMAGED because of (4) harmful viruses from recent adult sites. It will damage your phone's SIM card and will corrupt your contacts, photos, data, applications, etc.

If you do not remove the virus now , it will cause serious damage to your Google Nexus 5X. Follow instructions to repair your phone.

Step 1: Tap the button and install update from Google Play(FREE)!

Step 2: Open the app and repair your phone.

Future: client-side detection



Search/Ads links add a parameter with the **topics** found by the bot.

Check that the page matches the **same topics.**

Takeaways

Prevalence

5% of ads and 12% of search results for cloaking-prone keywords cloak.

Techniques

IP/User-Agent/Referer only gets ⅕ of cloaking.

Moving forward

Client side, semantic features needed for hard cases.

Thank you!

Luca Invernizzi
invernizzi@google.com