RSA Conference2016
San Francisco | February 29 – March 4 | Moscone Center

Connect to Protect

SESSION ID: MASH-R05

This Doesn't End Well: The TLD Explosion

Chris Larsen
Architect
WebPulse Threat Research Lab
Blue Coat
@bc_maware_guy

Daniel Hardman
Senior Research Engineer
WebPulse Threat Research Lab
Blue Coat

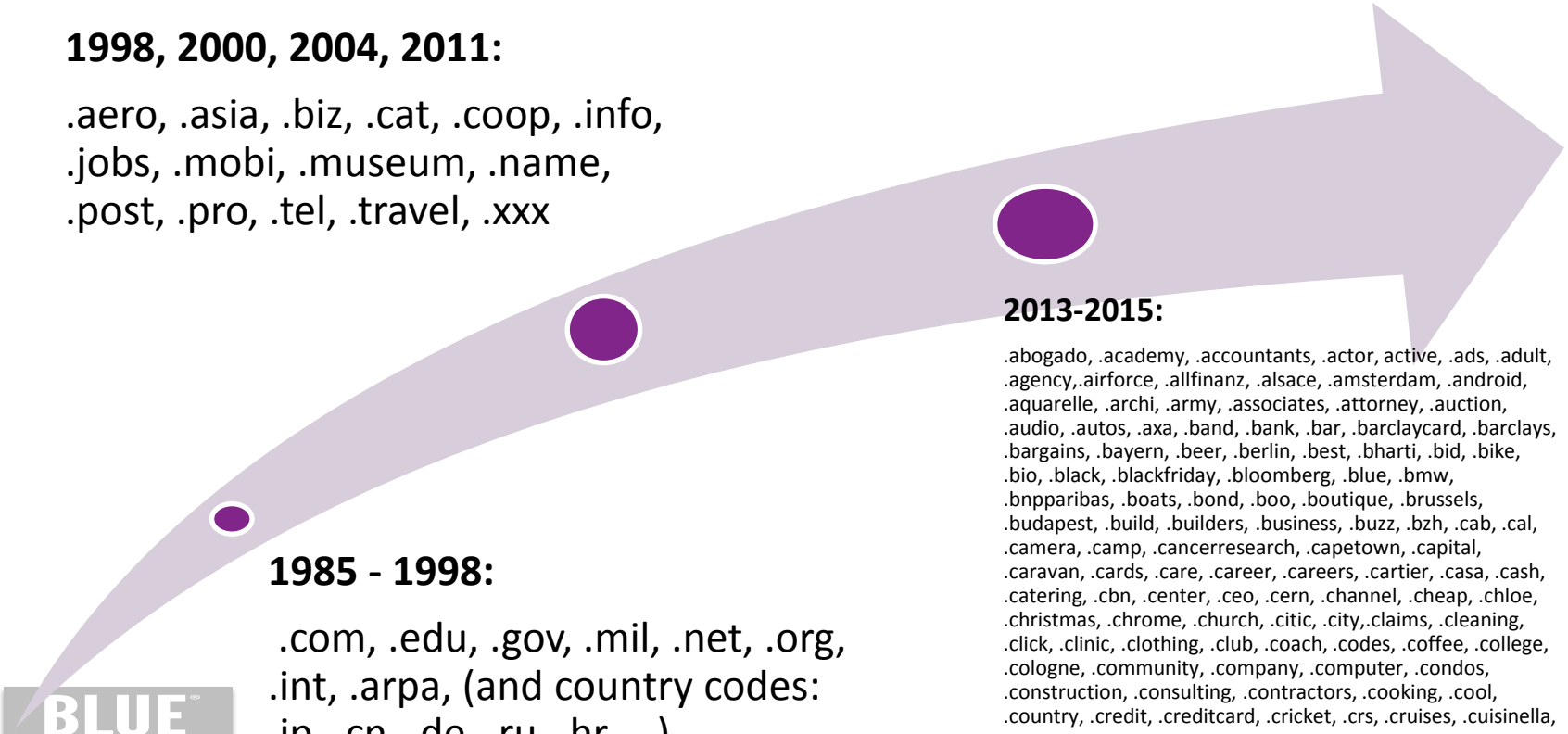
Agenda

- The TLD Explosion
- Current TLD Abuse
- Advanced TLD Abuse
- Action/Apply Slides

Problem: “All the good domains are taken”

- ICANN and IANA wanted to foster **choice** and **competition**
  - (and make some money…) ☺
- New TLDs include
  - gTLDs (“generic” or sometimes “global”), like *.xyz* or *.accountant*
    - (also in other languages, like *.maison* and *.futbol*)
  - Internationalized TLDs, like *.xn--p1ai* (.pф) and *.xn--3e0b707e* (.한국)
  - Geographic TLDs, like *.tokyo* and *.london*
  - Brand TLDs, like *.barclays* and *.hsbc*

The TLD Explosion



1998, 2000, 2004, 2011:

.aero, .asia, .biz, .cat, .coop, .info, .jobs, .mobi, .museum, .name, .post, .pro, .tel, .travel, .xxx

1985 - 1998:

.com, .edu, .gov, .mil, .net, .org, .int, .arpa, (and country codes: .jp, .cn, .de, .ru, .hr, …)

2013-2015:

.abogado, .academy, .accountants, .actor, active, .ads, .adult, .agency,.airforce, .allfinanz, .alsace, .amsterdam, .android, .aquarelle, .archi, .army, .associates, .attorney, .auction, .audio, .autos, .axa, .band, .bank, .bar, .barclaycard, .barclays, .bargains, .bayern, .beer, .berlin, .best, .bharti, .bid, .bike, .bio, .black, .blackfriday, .bloomberg, .blue, .bmw, .bnpparibas, .boats, .bond, .boo, .boutique, .brussels, .budapest, .build, .builders, .business, .buzz, .bzh, .cab, .cal, .camera, .camp, .cancerresearch, .capetown, .capital, .caravan, .cards, .care, .career, .careers, .cartier, .casa, .cash, .catering, .cbn, .center, .ceo, .cern, .channel, .cheap, .chloe, .christmas, .chrome, .church, .citic, .city,.claims, .cleaning, .click, .clinic, .clothing, .club, .coach, .codes, .coffee, .college, .cologne, .community, .company, .computer, .condos, .construction, .consulting, .contractors, .cooking, .cool, .country, .credit, .creditcard, .cricket, .crs, .cruises, .cuisinella, .cymru, …

BLUE COAT

The Top Twenty Shady TLDs

- Data as of 12/15
- Percentage of ratings in our DB with a negative security category
- Our data does skew to the negative, since that's what our systems are focused on finding
- (but still…)

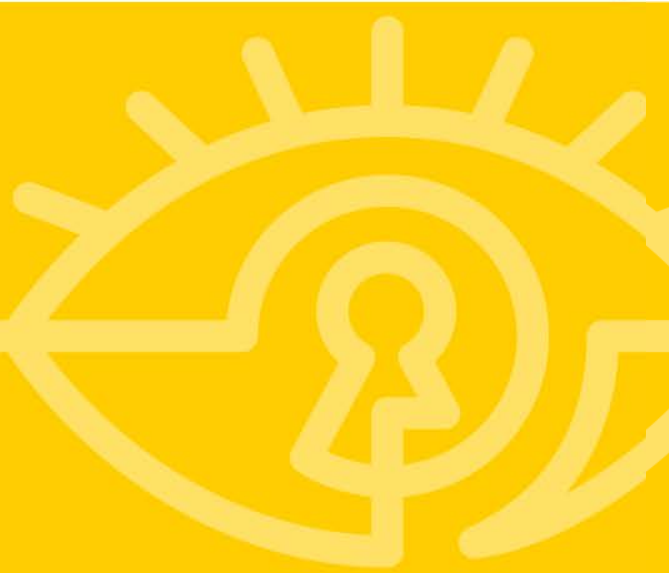
TLD	% of URLs in DB with "shady" category
.country	99.96%
.kim	99.54%
.download	99.53%
.racing	99.39%
.accountant	99.12%
.science	99.11%
.review	98.95%
.party	98.78%
.loan	98.62%
.win	98.54%

BLUE COAT

The Top Twenty Shady TLDs

- Data as of 12/15 (cont.)

TLD	% of URLs in DB with “shady” category
.bid	98.23%
.top	97.14%
.gq	95.41%
.nf	95.35%
.pw	95.27%
.link	95.19%
.ml	95.09%
.pro	94.89%
.cf	94.73%
.trade	94.38%



The TLD Registry Point of View (Case Study: *.xyz*)

The TLD Registry Point of View

- Common to operate multiple registries
  - *.xyz, .auto, .car, .cars, .college, .rent, .theatre, .security, .protection*
  - (different business models for generic vs. premium TLDs)
- *.xyz* timeline:
  - ICANN application submitted June 2012
    - 70+ pages, to document technical ability, background, etc.
    - $185,000 non-refundable application fee
    - (note that some TLDs required an auction…)

The TLD Registry Point of View

- ICANN approval in Dec 2013
  - All registries must pay ICANN $0.25 per domain sold
  - (minimum $25,000 owed per year)
- March 2014 “trademark exclusive sunrise”
  - Recommended premium fee of $250 (standard renewal fees)
- June 2014 “general availability”
  - Recommended standard retail/renewal ($10 per year)

The TLD Registry Point of View

- How's business?
- As of mid-February*:
  - Over 2M domains
  - 224 different countries
  - 128 different registrars
- Next question: How to keep the Bad Guys out?

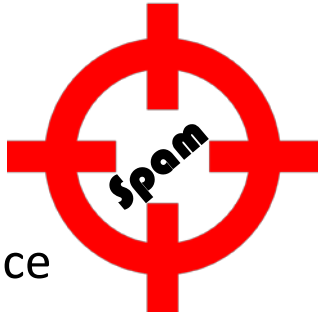
Top 10 Biggest Selling gTLDs [?]

by total | by increase

	.TLD	Domains	+ Today	
1	.xyz	2,369,039	77,145	>
2	.top	1,513,928	170,307	>
3	.club	697,747	1,149	>
4	.win	643,998	26,135	>
5	.wang (net)	629,619	6,627	>
6	.网址 (web address)	340,952	12	>
7	.science	338,555	102	>

*namestat.org

The TLD Registry Point of View



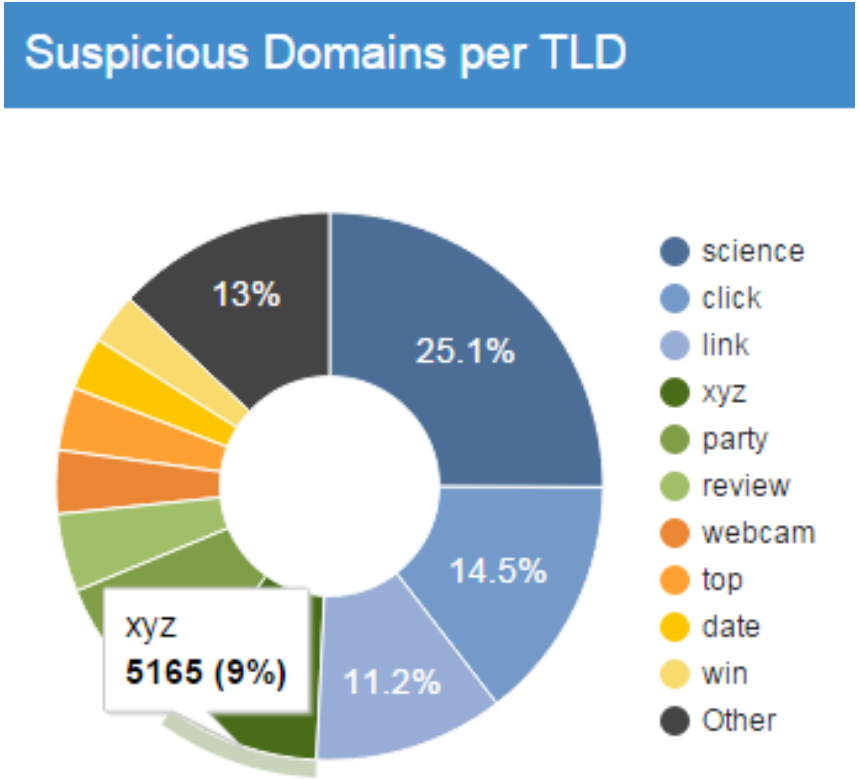
- XYZ decided to build an abuse tracking system:
  - 100+ data feeds (phishing, malware, farming, spam, …)
  - Domains assigned risk score based on severity and confidence
  - Any with a risk score are monitored more closely, but not suspended
  - Registrant's domains (all TLDs) are checked for abusive patterns
  - Suspension requires threshold (multiple feeds, high confidence)
    - Try to avoid false positives

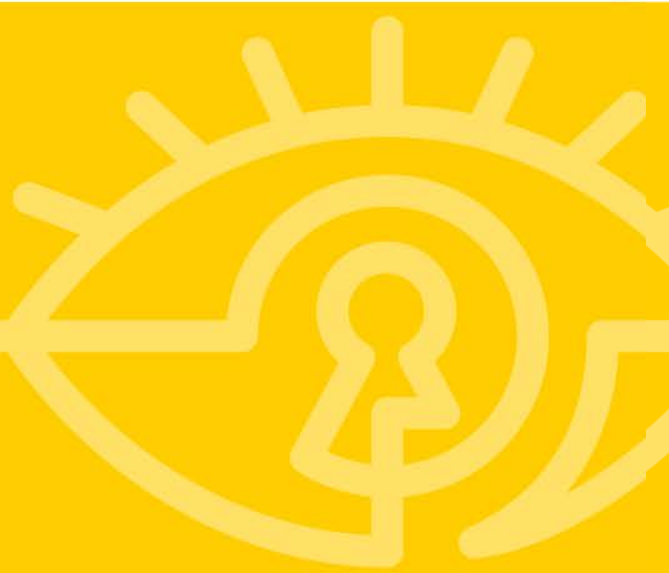
The TLD Registry Point of View

- *.xyz* abuse tracker, continued:
  - Registrars notified of abuse/suspension
    - Investigate payment fraud, and contact registrant
    - Registrant may not know site is compromised
  - Domain is temporarily suspended to avoid infecting visitors
  - Contact address provided to registrant; upon contact:
    - Clean-up instructions provided
    - Site is un-suspended (but still on “closely monitor” list)

The TLD Registry Point of View

- How are they doing?
- Well, out of over 2M domains registered, 5165 bad apples isn't too bad (~0.22%)
- (our numbers are calculated differently, but we have seen definite improvement)





TLDs and Ambiguity

The .Zip Controversy

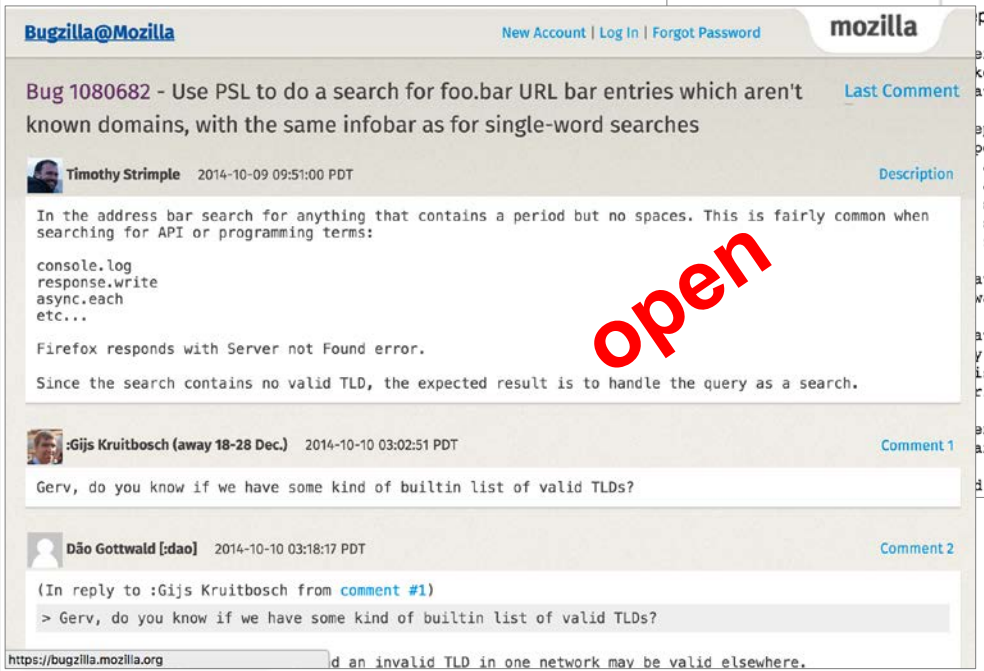
- In our first big “Shady TLD” report, we had *.zip* as #1
- Skeptics pointed out that there was only one real *.zip* domain
  - (and it was simply Google’s page talking about their new TLDs)
- Q: How can a TLD with no domains be “100% Shady”
- A: Because there is now ambiguity in the world
  - *(.zip* used to be a File Extension, but now it ***might*** be a domain…)

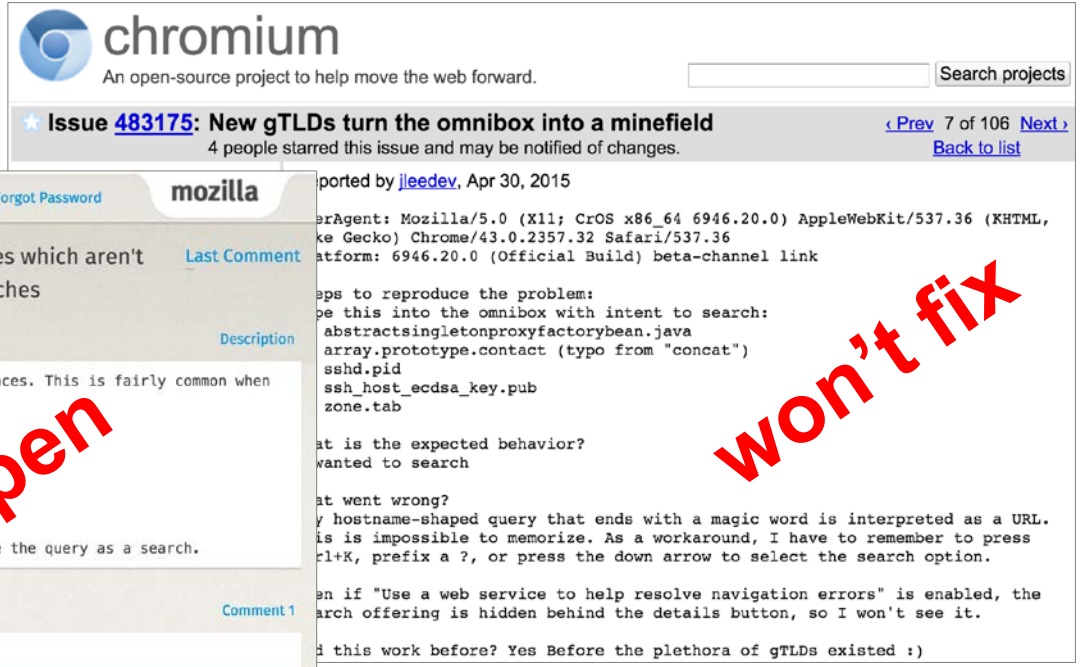
Pity the Poor Browser Coders…

- The “omnibox” (address bar) is causing some heartburn…
- Browsers let you type URLs and search terms in same place
- Traditional strategy: use “context” to tell apart
- But *.zip* (etc.) no longer has a single context…

It’s not a bug, it’s a “feature”…

http://j.mp/1QWbVd3



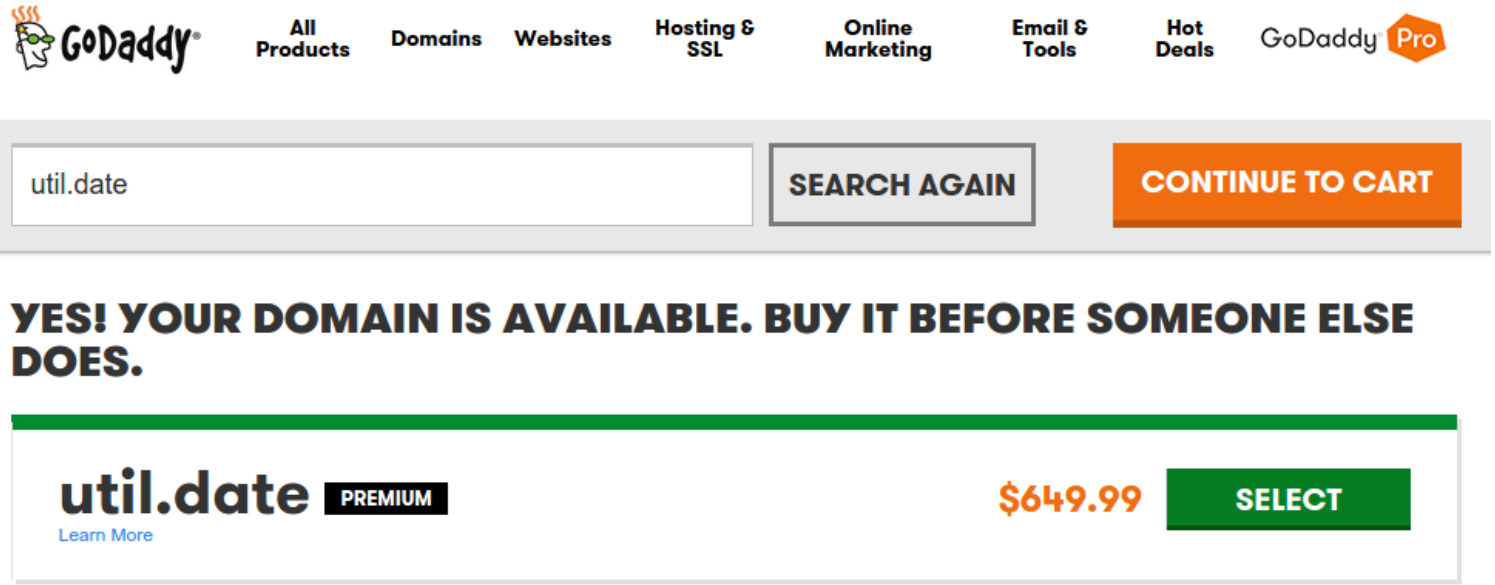


http://j.mp/1OJiU4p

.Zip Got Us Thinking… .Date Got Us Worried

- .Date -- designed to be a “premiere” TLD for dating sites
  - (when we profiled it, none of the spam/scam used this angle)
- This started the wheels turning in our heads…
- http://java.util.date, anybody?

No Fancy Hacks Needed (Just $)



Collisions – Namespaces and gTLDs

gTLD	Java 8 collision(s)
.active	org.omg.PortableInterceptor.ACTIVE
.channel	java.nio.channels.Channel
.date	java.util.Date, java.sql.Date
.engineering	javax.print.attribute.standard.MediaSize.Engineering
.global	javax.xml.bind.annotation.XmlElementDecl.GLOBAL
.graphics	java.awt.Graphics
.group	java.security.acl.Group
.info	javax.sound: .sampled.DataLine.Info, .sampled.Line.Info, .sampled.Port.Info, .midi.MidiDevice.Info, .sampled.Mixer.Info

https://github.com/dhh1128/ns-gtld-collide

Collisions – Namespaces and gTLDs (cont.)

gTLD	Java 8 collision(s)
.lease	java.rmi.dgc.Lease
.media	javax.print.attribute.standard.Media
.menu	java.awt.Menu
.na	javax.print.attribute.standard.MediaSize.NA
.name	java.util.jar.Attributes.Name, javax.lang.model.element.Name, javax.naming.Name, javax.xml.soap.Name
.properties	java.util.Properties
.style	javax.jws.soap.SOAPBinding.Style, javax.swing.text.Style

What about:

.NET?

System.Drawing.Graphics

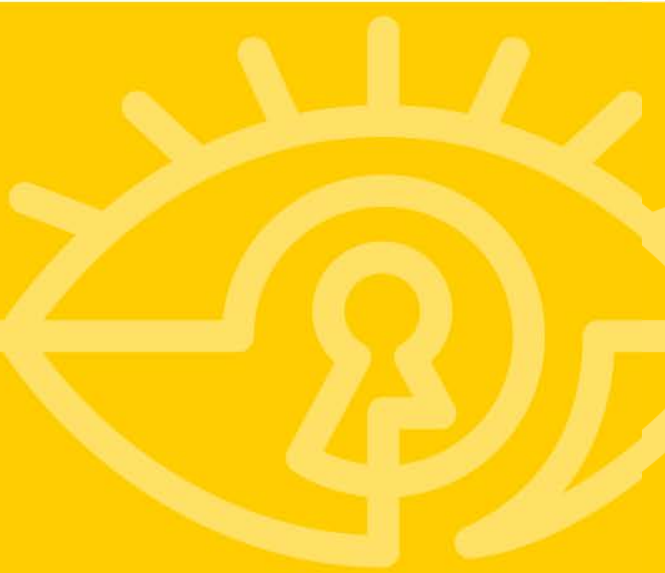
Python?

datetime.date

Ruby?

RSS::Rss::Channel

https://github.com/dhh1128/ns-gtld-collide



Intermission: A Quick Look At IDNs*

*See our RSA 2014 Paper:

Where in the World is xn--80atbrbl6f.xn--p1ai?

International Criminals Hiding Out in Internationalized Domain Names

IDNs and Homograph Attacks

- Substitute Unicode chars for common letters:
  - Presto! Lookalike domains! (wikipedia.org vs. wikipedia.org)
- IETF and ICANN set rules https://www.icann.org/resources/pages/idn-guidelines-2011-09-02-en
  - (basic idea: don't allow mixed character sets)
  - Some non-conformance grandfathered in
  - HTTP clients & bind servers may ignore https://tools.ietf.org/html/rfc5891
  - (most importantly) **knowledge isn't universal**

IDNs and Homograph Attacks

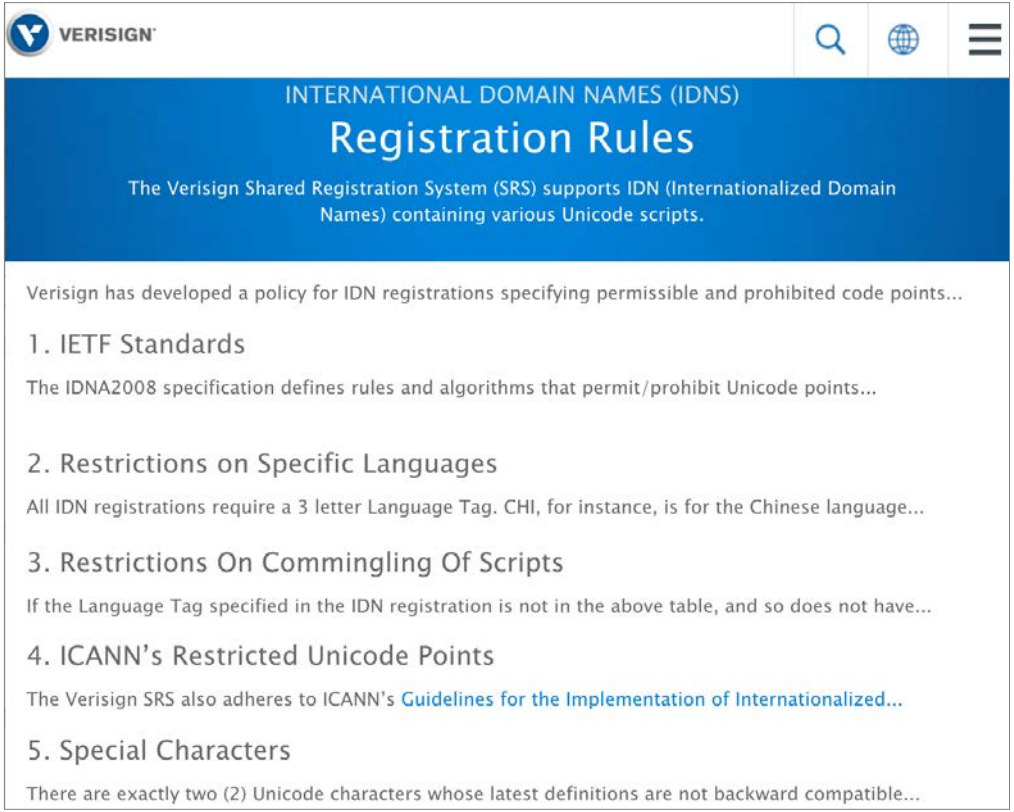
Verisign, for example,

has rules...

But does ***everyone***

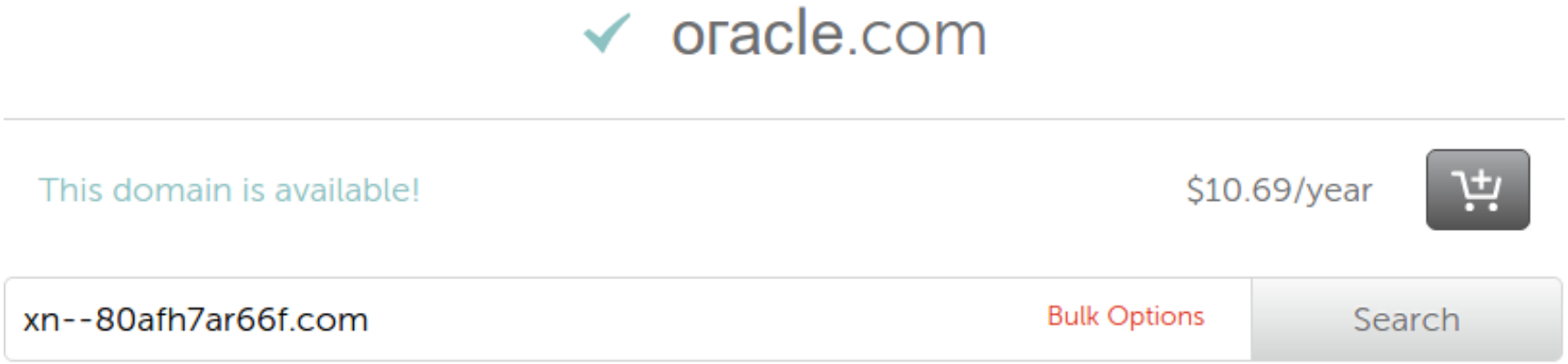
follow them?

All the time?



http://j.mp/1OJitXO

Case Study: Registering oracle.com



u+043e u+0433 u+0430 u+0441 u+04cf u+0435

Case Study: Registering oracle.com

1:35 **You: I need to get prompted for IDN language code so I can buy a nonEnglish domain, but it skips that step and then says that the IDN language code is missing.**

1:35 Dmitry M*: Hello Daniel… may I know the domain name in question?

1:37 Daniel Hardman: The domain is "xn80afh7ar66f.com". It's Russian. I just want checkout to prompt me for the IDN language code so I can answer that question.

1:42 Dmitry M: As I understand, you are attempting to register oracle.com, am I right?

1:42 Daniel Hardman: Yes, that is the punycode reversal.

1:43 Dmitry M: Thank you, please hold on.

1:50 Dmitry M: Daniel, please hold on, I am registering domain manually for you.

2:15 Dmitry M: Daniel, as I see, we have some issues with IDN registration at the moment. I deeply apologize for these inconveniences. We need to contact our upstream provider in order for them to investigate this issue.

11 Days Later…

Good catch!

Alexander N. <billing@[redacted].com>

To: Hardman, Daniel;

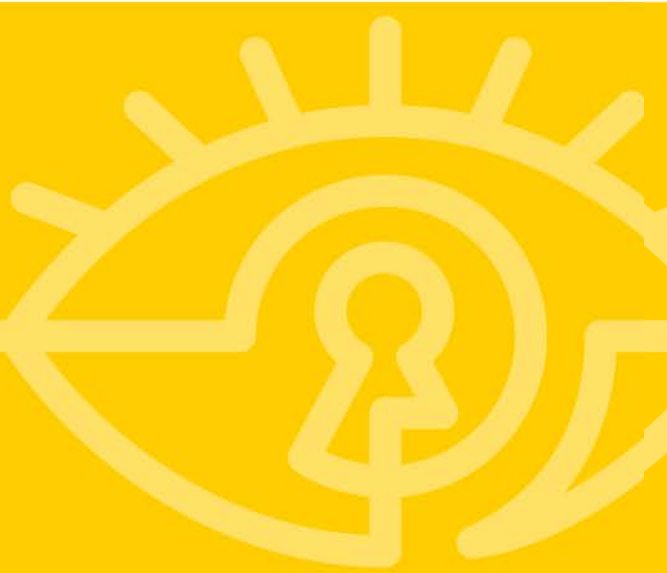
Thank you for your patience.
We have received an update from our upstream provider.
Kindly note that the encoding type is Russian, but some of the characters are not Russian.
Unfortunately, the Registry does not allow mixed encoding because the resulting domain names can be used in phishing attacks.

Alexander N.

Ticket ID: [redacted]

Registering oracle.com -- Analysis

- The rules were followed, and worked, in this case
  - But this was for a registrar in the US
  - With a clearly illegal value (the “I”)
  - And it took 11 days to catch. ☹
- Note: the “one” digit (ascii 0x31) is legal in Russian names…
  - …and orac1e.com is available…
- Conclusion: IDN-type attacks constrained, not impossible



Putting It All Together:
A Fun New Spearphishing Attack
(no attachments to open; no links to click)

Spearphishing PoC

From: [a believable looking name]

Subject: worried about new java CVE; may affect our codebases

Some of you that follow java vulnerabilities may have heard about the problem that was recently reported with java's javax.sound.midi.MidiDevice.Info class. Reading metadata from a specially deformed MIDI file can allow an attacker to break out of a JVM's sandbox.

What's really worrisome--and where this impacts us--is that chrome and firefox both allow extensions/plugins to play MIDIs for events (e.g., to signal a successful screenshot, the end of a download, a form submission, etc)--and if client-side java is active on a website when MIDI playback occurs, java owns the playback channel.

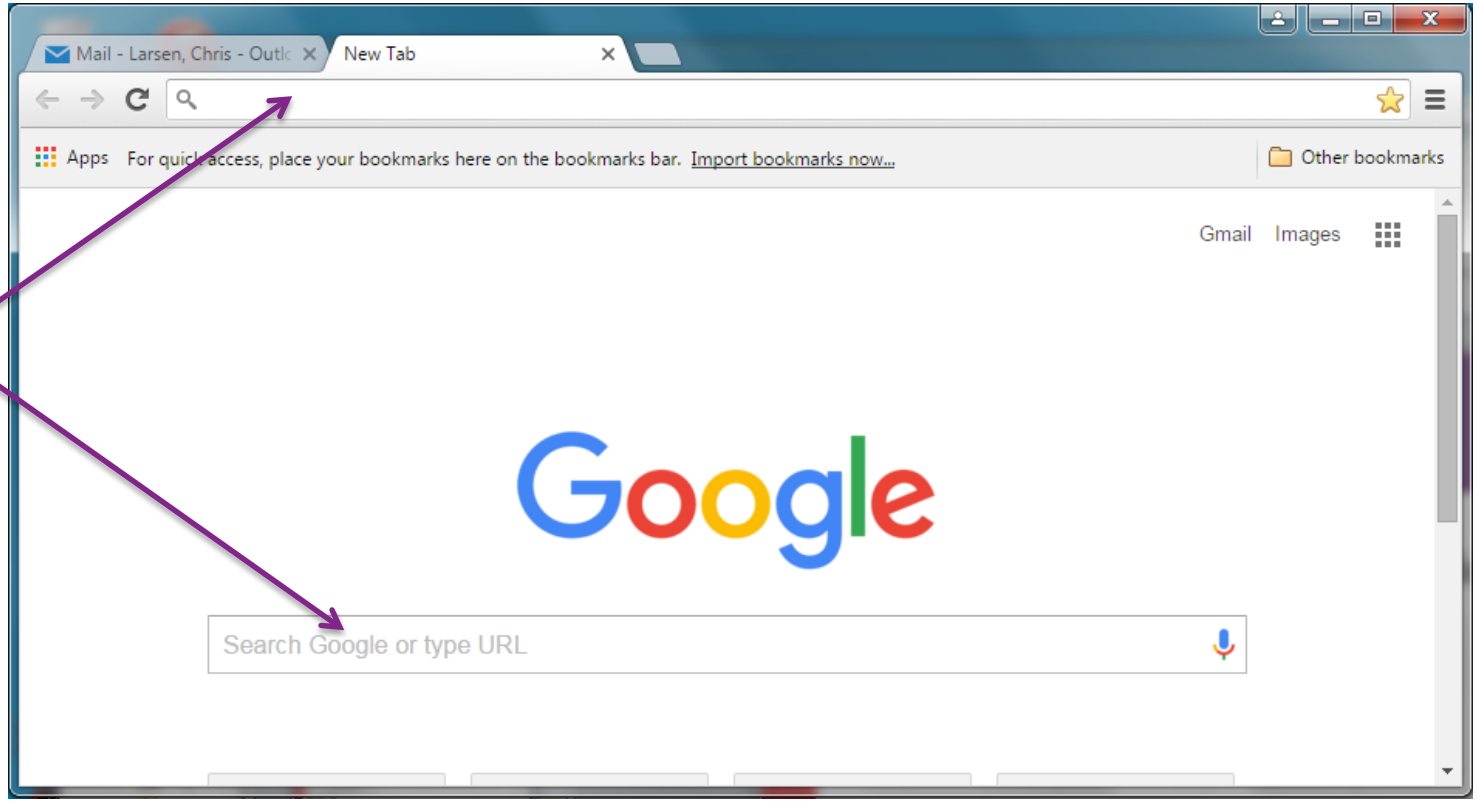
[...]

When Google researchers at Project Zero posted about this (Feb 4), they said there was a way to block the exploit by influencing the classloader with some system properties related to audio. Our team is going to be doing some research to see if we can release a quick patch. I recommend that you google javax.sound.midi.MidiDevice.Info for more info.

--[believable name]

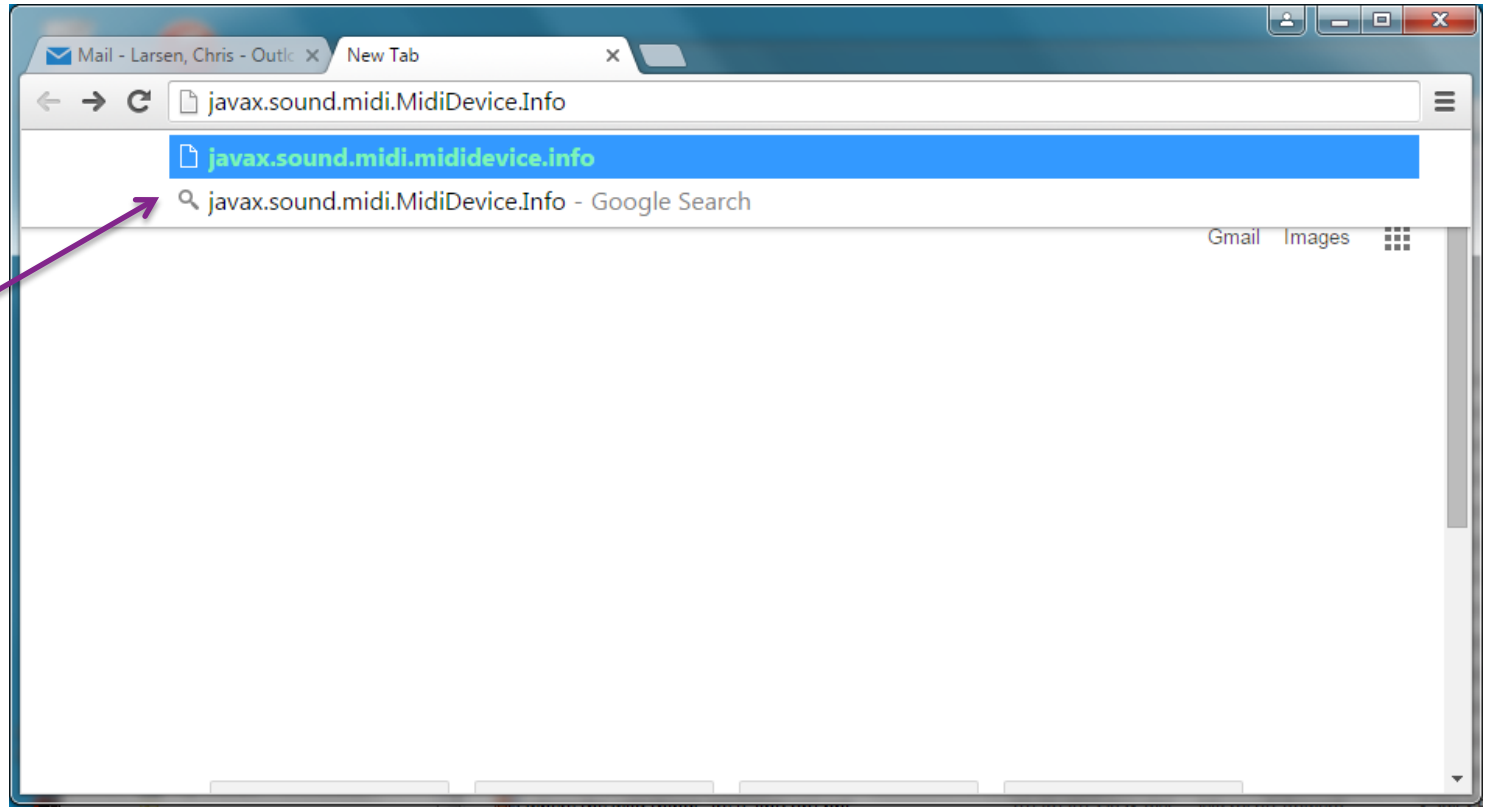
Spearphishing PoC

Pick either one



Spearphishing PoC

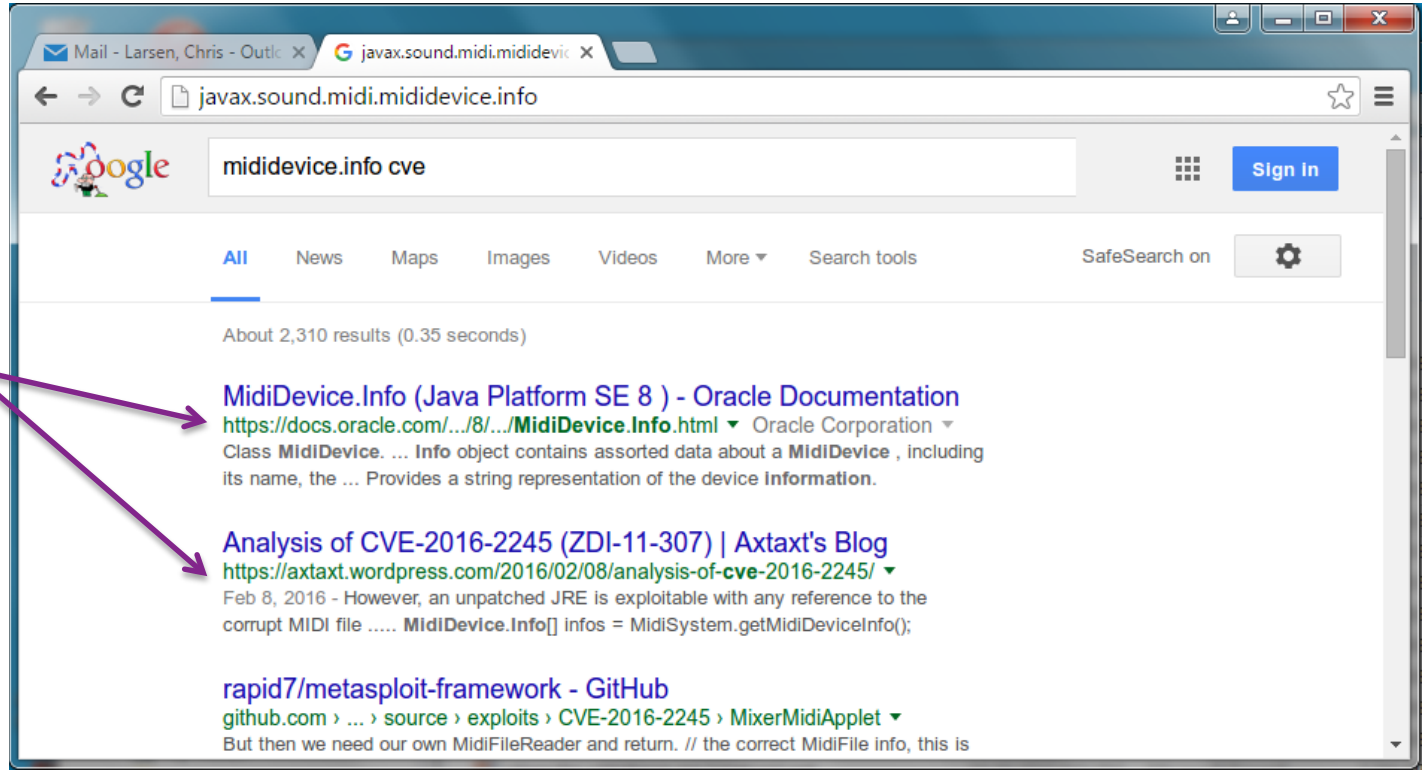
If you don't explicitly choose the 2nd option, you take the bait…



Spearphishing PoC

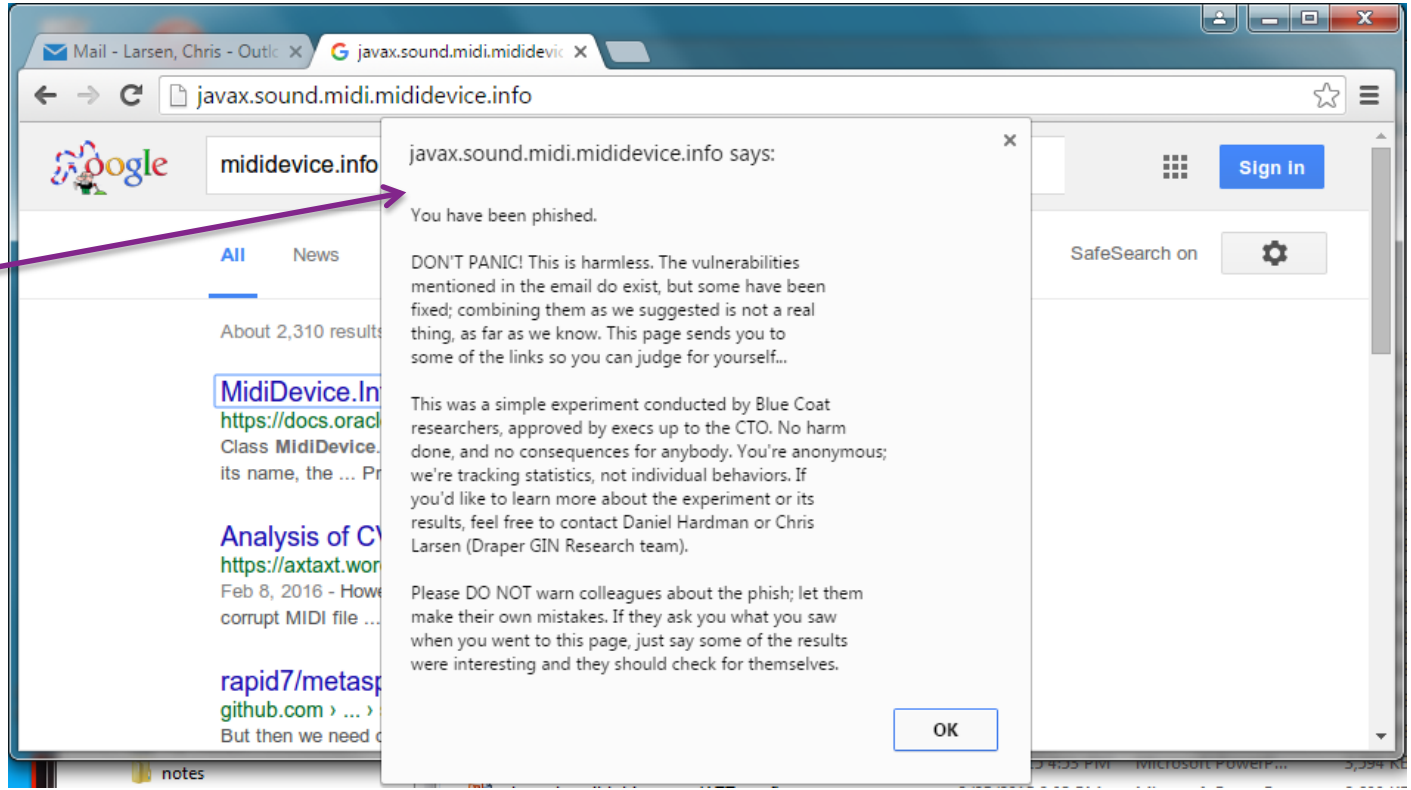
Originally, we were going to have all the links be to our fake IDN version of oracle.com…

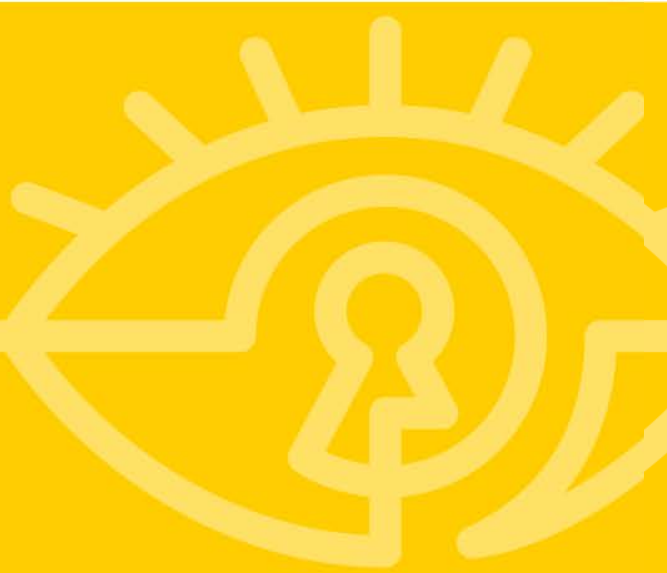
…but then the Internet handed us a live CVE!



Spearphishing PoC

All of the links bring up this pop-up. The OK button then takes them to the real article.





Wrap-up:
Action/Apply Slides

“Apply” Slide #1: The Top 20 Shady TLDs

TLD	% of “shady” URLs in DB
.country	99.96%
.kim	99.54%
.download	99.53%
.racing	99.39%
.accountant	99.12%
.science	99.11%
.review	98.95%
.party	98.78%
.loan	98.62%
.win	98.54%

Circulate an e-mail about these TLDs in your org, to raise awareness.

…and maybe add a note about possible IDN abuse.
(Use mouseover!)

TLD	% of “shady” URLs in DB
.bid	98.23%
.top	97.14%
.gq	95.41%
.nf	95.35%
.pw	95.27%
.link	95.19%
.ml	95.09%
.pro	94.89%
.cf	94.73%
.trade	94.38%

“Apply” Slide #2: Ambiguous TLDs

Collision TLDs (Java)	
.active	.lease
.channel	.media
.date	.menu
.engineering	.na
.global	.name
.graphics	.properties
.group	.style
.info	

Another e-mail to summarize the “ambiguous” TLDs and their abuse scenarios…

…but this one needs to be targeted more at your engineers.

https://github.com/dhh1128/ns-gtld-collide

Questions ???

(Btw, contact chris.larsen or daniel.hardman
@ bluecoat.com
if you want to use our spearphish site and get stats)

Useful Links:

https://namestat.org/
https://ntldstats.com/
https://github.com/dhh1128/ns-gtld-collide