Soundsquatting

Uncovering the use of homophones in domain squatting

Nick Nikiforakis, Marco Balduzzi, Lieven Desmet, Frank Piessens, and Wouter Joosen

(ICS 2014, 12th October, Hong Kong)

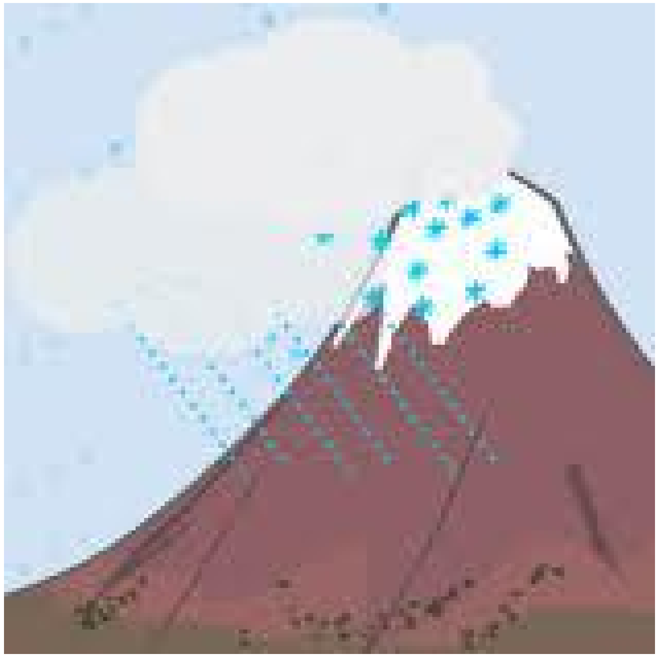
Outline

- Intro on Soundsquatting
- Generating soundquatting domains (`AutoSS`)
- Large-scale experiment
  - Findings
- User characterization
- Sound-dependent users
- Lessons learned

Soundsquatting

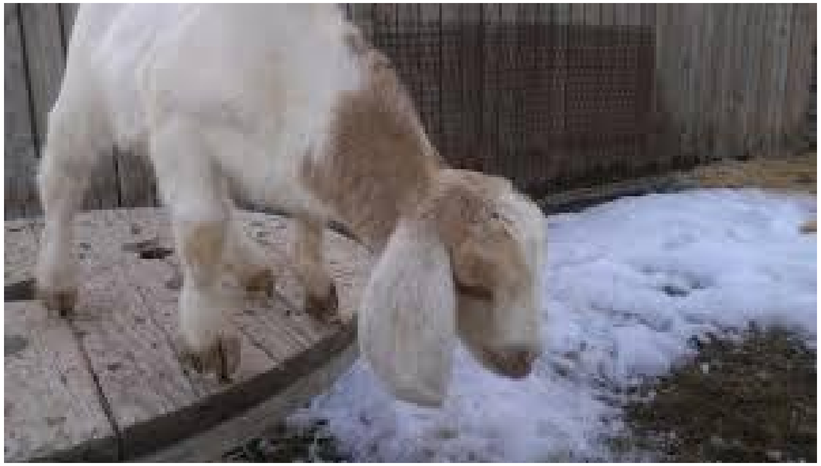
- Homophone-based squatting
- Homophones: words that have the same pronunciation, but are spelled differently
- Same meaning:
  - guarantee = guaranty
- Different meaning:
  - weather (clime)
  - whether (conj.)
  - wether (male sheep)

Example #1

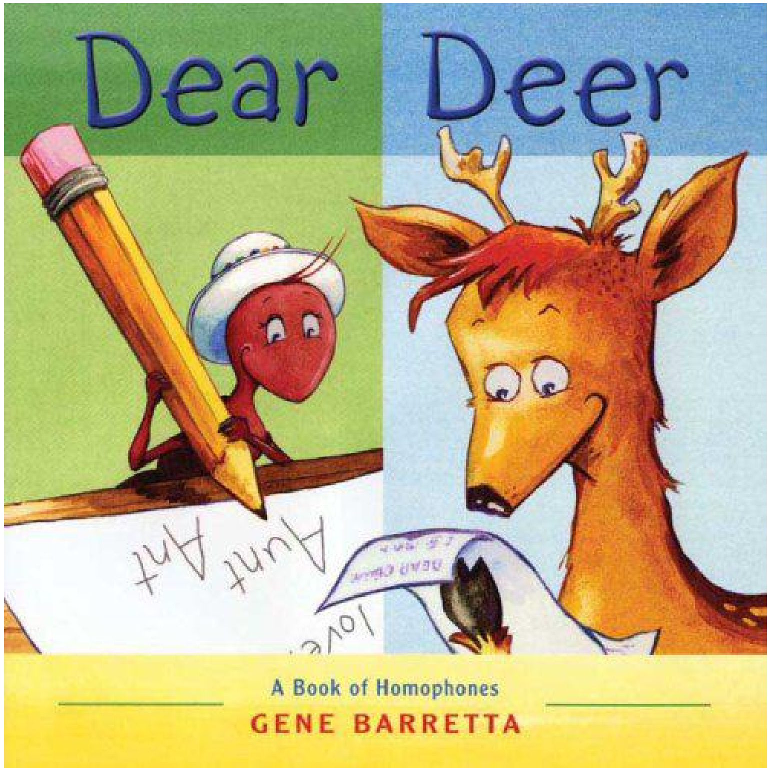


- weather

- wether



Example #2



Attack Scenario

- Attacker registers a soundquatting version of a targeted domain *(authoritative domain*),
  - e.g. youtube → yewtube.com (type of wood)
- Leverage the homophone-confusion of users
- Monetizes the hits in different forms:
  - Advertisements
  - Affiliate programs
  - Scams and information leakages
  - Phishing
  - Malware
  - Espionage (email)

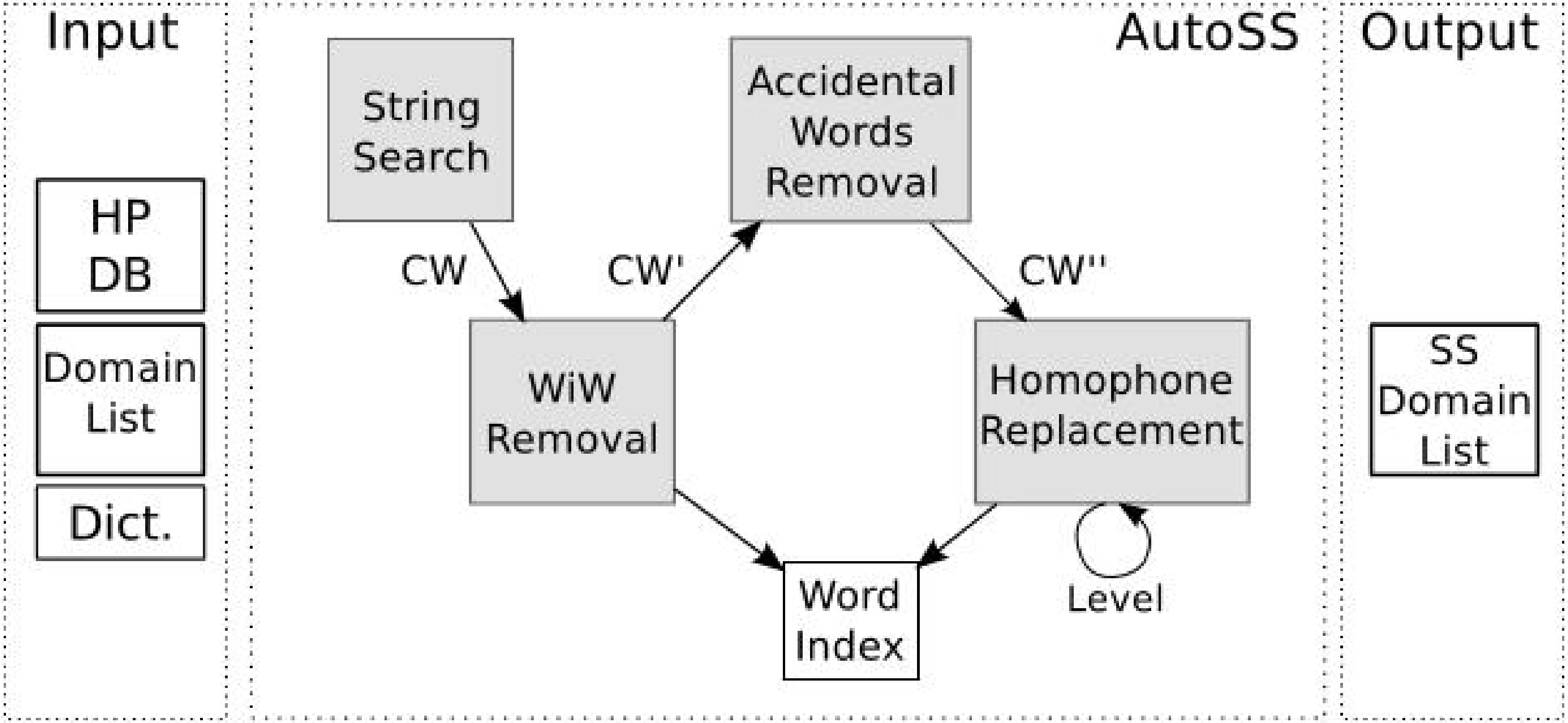
Differences with Typosquatting

- Both being domain squatting attacks, but
- Soundsquatting leverages homophone-confusion
- Typosquatting leverage “typos” (misspelling), i.e.:
  - missing dot: wwwexample.com
  - character omission: www.exmple.com
  - character insertion: www.exaample.com
  - character permutation: www.examlpe.com
  - character replacement: www.ezample.com

[27] Y.-M. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels. Strider typo-patrol: discovery and analysis of systematic typo-squatting. SRUTI’06, 2006.

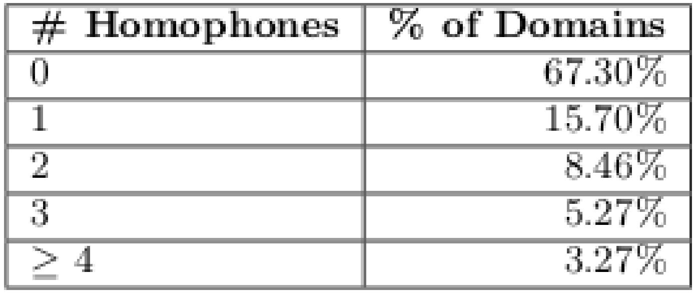
Generating soundsquatting domains

- `AutoSS` (AutoSoundSquatter)
  - `WiW:` linkedin (in, ~~ink, inked, ked, link,~~ linked)
  - `AWR:` leaseweb (lease, ~~sew~~, web)

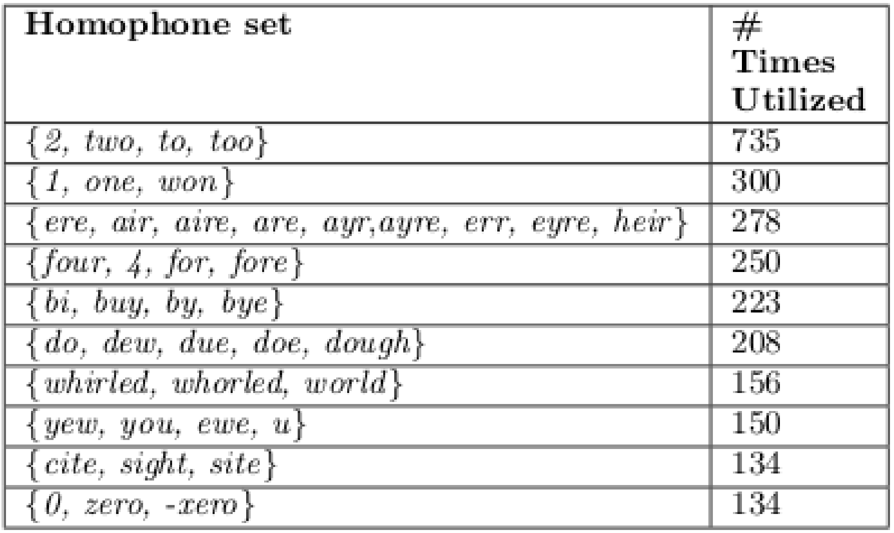


Uncover Soundsquatting

- Large-scale experiment: Alexa Top 10K
- Homophone databases (1,337 sets)
- 67.3% domains contained no homophones
- *8,476 soundsquatting domains*



# Homophones	% of Domains
0	67.30%
1	15.70%
2	8.46%
3	5.27%
≥ 4	3.27%



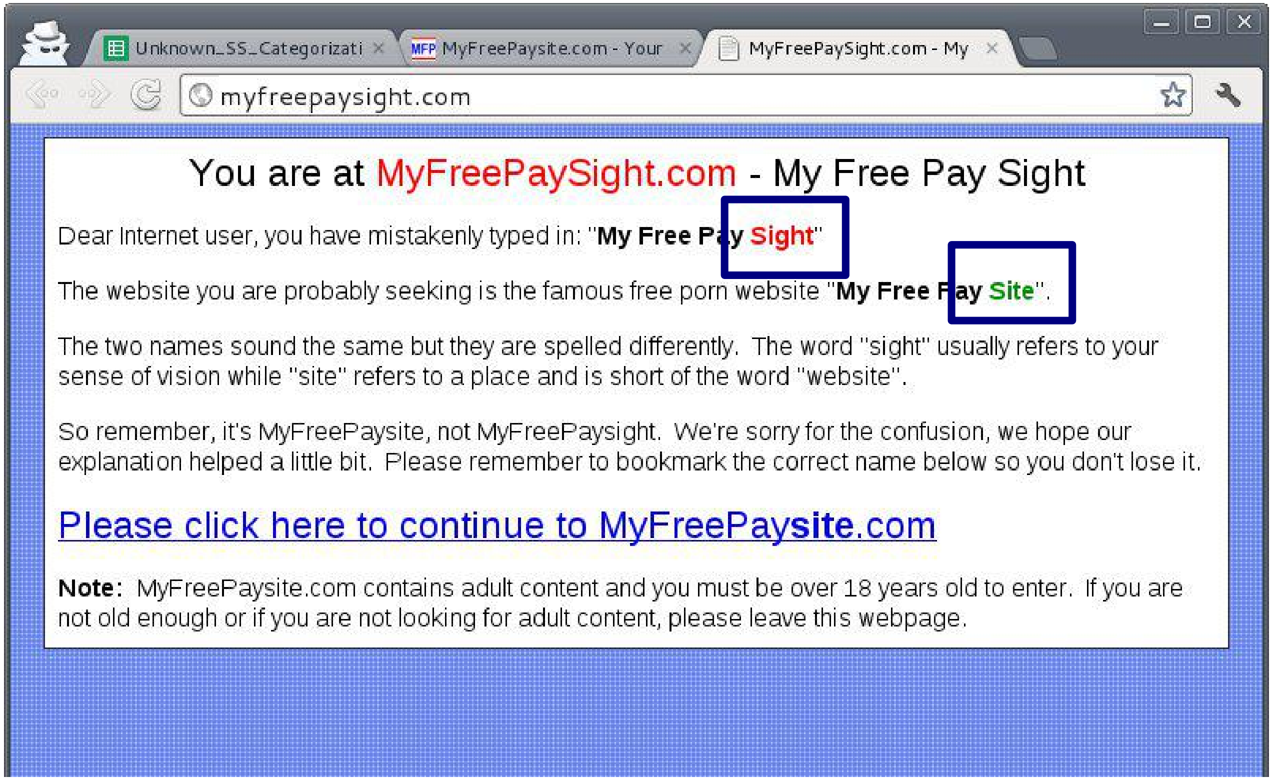
Homophone set	# Times Utilized
{2, two, to, too}	735
{1, one, won}	300
{ere, air, aire, are, ayr,ayre, err, eyre, heir}	278
{four, 4, for, fore}	250
{bi, buy, by, bye}	223
{do, dew, due, doe, dough}	208
{whirled, whorled, world}	156
{yew, you, ewe, u}	150
{cite, sight, site}	134
{0, zero, -xero}	134

Method of Categorization

- Identify already-registered domains
  - IP and WHOIS lookups
  - Verification against known registrants
  - *1,823 soundsquatting domains online*
- Crawler based on PhantomJS (agent-less)
  - 10 seconds visit
  - Screenshot, HTML and URL chain dumps
- Semi-automated analysis
  - Parked, offline (404), under-construction
  - Use of signatures, the rest *(417 sites)* manually

Characterization Results

- 155 Authoritative-owned domains
- 301/302 HTTP redirection



You are at MyFreePaySight.com - My Free Pay Sight

Dear Internet user, you have mistakenly typed in: "**My Free Pay Sight**"

The website you are probably seeking is the famous free porn website "**My Free Pay Site**".

The two names sound the same but they are spelled differently. The word "sight" usually refers to your sense of vision while "site" refers to a place and is short of the word "website".

So remember, it's MyFreePaysite, not MyFreePaysight. We're sorry for the confusion, we hope our explanation helped a little bit. Please remember to bookmark the correct name below so you don't lose it.

Please click here to continue to MyFreePay**site**.com

Note: MyFreePaysite.com contains adult content and you must be over 18 years old to enter. If you are not old enough or if you are not looking for adult content, please leave this webpage.

Best forms of monetizing

- Parked/Ads/For Sale domains
  - 954 cases, 52.3%
  - Ads constructed on demand
  - Use of domain-parking agencies
- Affiliate-abusing domains
  - 32 cases
  - Use of affiliate programs
  - Commission every time the use visit the soundsquatted domain of an authoritative site, e.g.
    - `mybrowsercache.com` → `http://www.mybrowsercash.com/index.php?`<u>`refid=312044`</u>

Hit Stealing

- 22 Cases
- Redirect the traffic to a competitor site
- Most targeted business categories: adult, online shopping and travel
- Example:
  - online gaming site game5.com: soundsquatted as gamefive.com (parked → gaming site)
  - transvestite-oriented porn site ashemaletube.com: soundsquatted as ashemailtube.com which redirects to trannydates.com

Scams

- 16 domains
- Lure visitors into subscribing to fake lotteries and surveys
- `vhone.com, soundquatting version of vh1.com`
  - Electronic business
  - “Survey-scam” promising techie prizes in change of private information
  - Names, email addresses, mobile phone numbers

Promoting-related domains

- 7 cases of domains promoting something or someone related to the authority domains
- `teambeechbody.com ss for teambeachbody.com`
- beech (wood) VS beach (coastline)

On-line fitness club

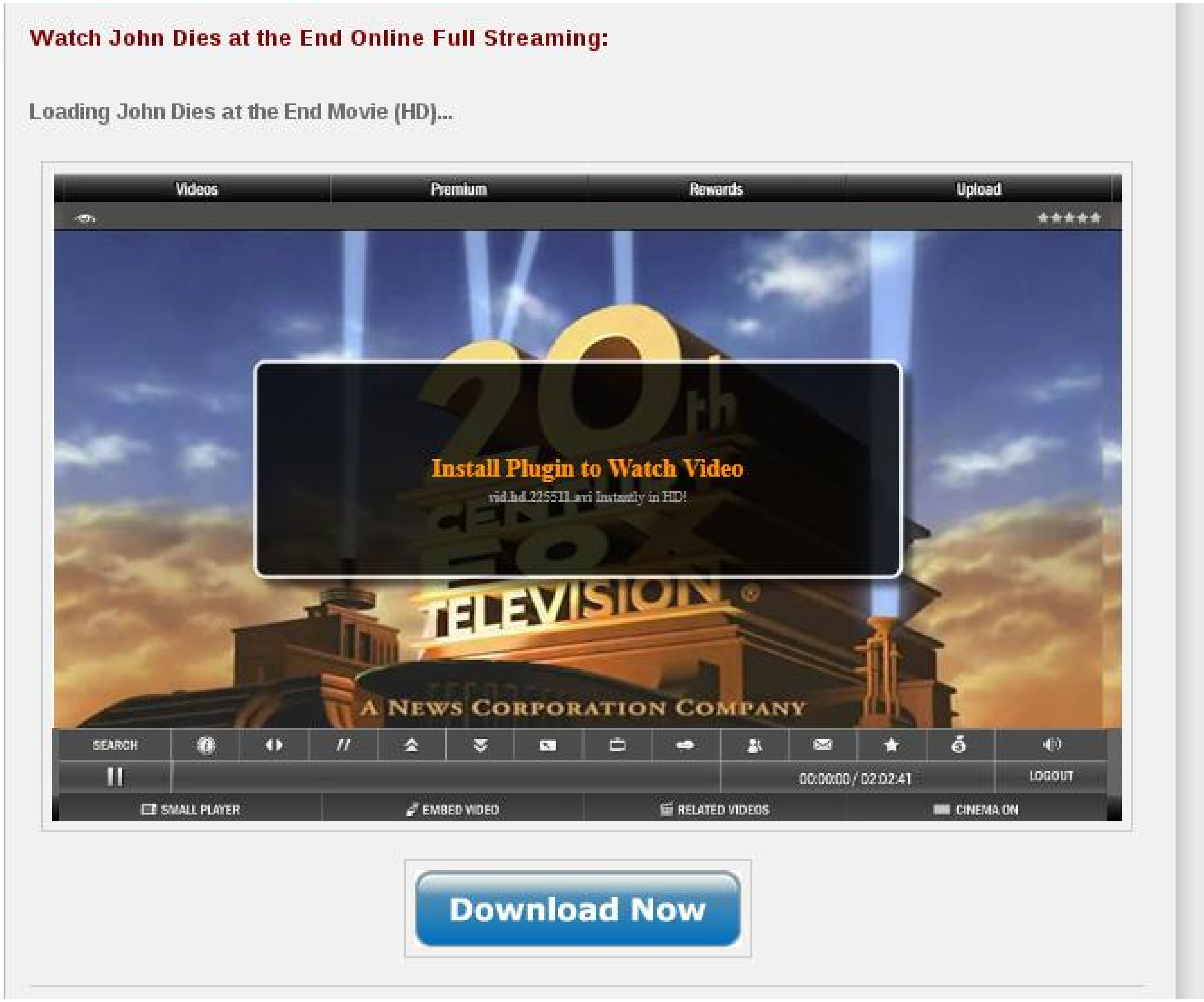
- Promotes a specific coach
  - working for the authoritative domain's organization

Other Malicious Intents

- `utube.com ss_for YouTube`
  - Videos to social-engineer the users
  - Divulging personal information
  - Installing malicious browser extensions

- `movreal.com ss_for movreel.com`
  - Free of charge video-streaming provider
  - Hosts malicious content

Social-engineering to spread malware



Watch John Dies at the End Online Full Streaming:

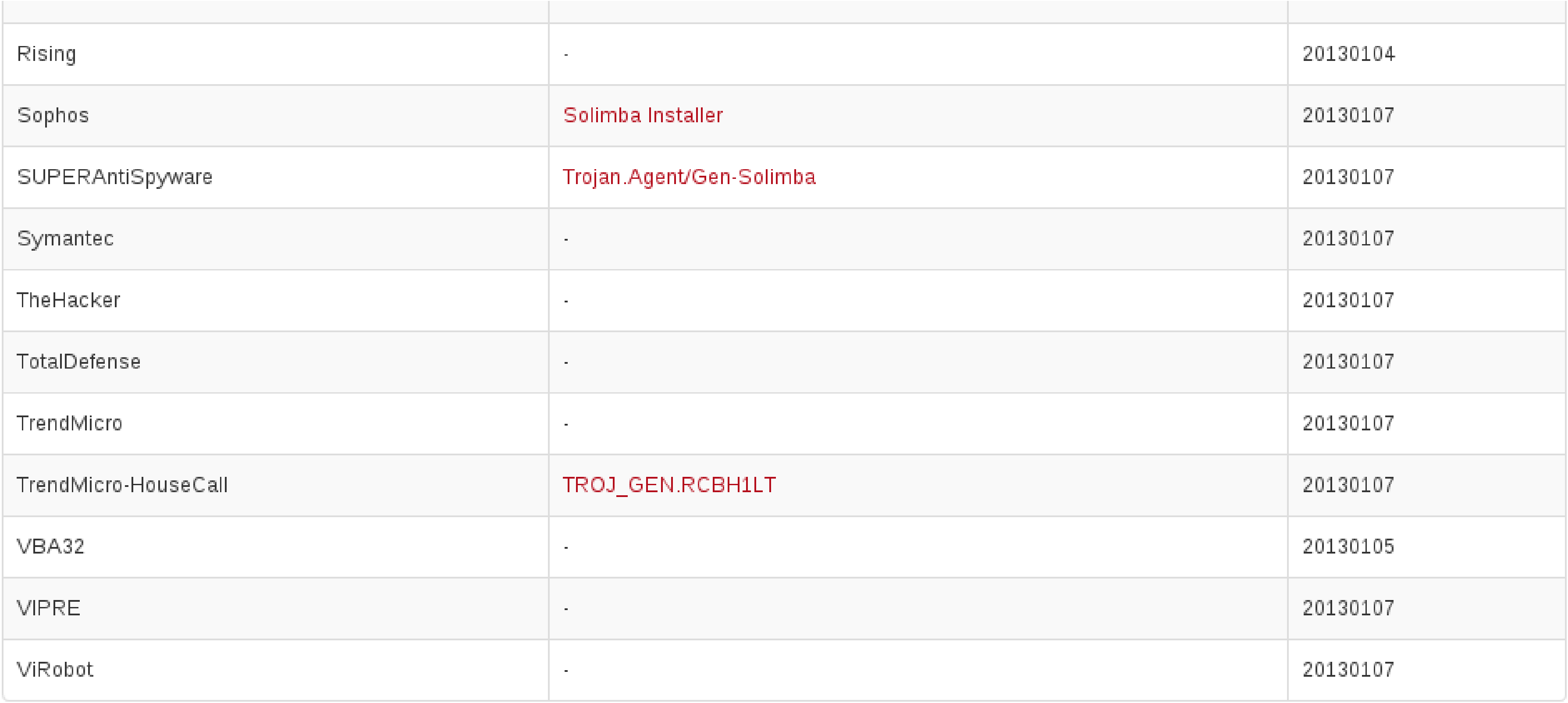
Loading John Dies at the End Movie (HD)...

Install Plugin to Watch Video

vid.hd.225511.avi Instantly in HD!

“Provides” Solimba

- Adware campaign
- Installer for other malware



Rising	-	20130104
Sophos	Solimba Installer	20130107
SUPERAntiSpyware	Trojan.Agent/Gen-Solimba	20130107
Symantec	-	20130107
TheHacker	-	20130107
TotalDefense	-	20130107
TrendMicro	-	20130107
TrendMicro-HouseCall	TROJ_GEN.RCBH1LT	20130107
VBA32	-	20130105
VIPRE	-	20130107
ViRobot	-	20130107

Other Malicious Intents

- 2 Phishing Cases
  - Banks

- Fake email providers
- Steals email credentials
- `innbox.lv → InBox`

User Characterization

- We registered 30 soundsquatting domains
  - Show blank page and log
- Understand who and why users (victims) access them
- Bot/human detection:
  - `useragentstring.com` = 716 bot signatures
  - `stopforumspam.com` = 350,000 IPs of bots

Auth. Domain	Homophone pair	SS Domain	#Human Req. (per month)
thefreedictionary.com	{*the,thee*}	theefreedictionary.com	283 (39.86%)
fc2.com	{*2, too*}	fctoo.com	165 (44.84%)
jimdo.com	{*do, doe*}	jimdoe.com	150 (38.27%)
turbobit.net	{*bit, bitt*}	turbobitt.net	132 (36.07%)
leboncoin.fr	{*coin, quoin*}	lebonquoin.fr	110 (74.32%)
adserverplus.com	{*ad, add*}	addserverplus.com	98 (60.49%)
profitclicking.com	{*profit, prophet*}	prophetclicking.com	56 (48.28%)
hostgator.com	{*gator, gaiter*}	hostgaiter.com	45 (45.92%)
sitesell.com	{*sell, cel*}	sitecel.com	44 (40.00%)
discuz.net	{*disc, disk*}	diskuz.net	43 (40.19%)
tube8.com	{*8, ait*}	tubeait.com	42 (43.30%)
clixsense.com	{*sense, scents*}	clixscents.com	40 (44.44%)
a8.net	{*8, eight*}	aeight.net	48 (43.24%)
newegg.com	{*new, gnu*}	gnuegg.com	37 (36.63%)
redtubelive.com	{*red, read*}	readtubelive.com	44 (51.76%)
fiverr.com	{*err, air*}	fivair.com	33 (37.93%)
exoclick.com	{*click, clique*}	exoclique.com	32 (45.71%)
theglobeandmail.com	{*mail, male*}	theglobeandmale.com	35 (38.46%)
pastebin.com	{*bin, been*}	pastebeen.com	35 (39.77%)
ku6.com	{*6, sics*}	kusics.com	28 (33.33%)
...	...	...	...
Total Requests per Month (30 domains):			1,718

Findings

- `jimdo.com` = provider hosting personal pages
  - Squatting error in the SLD
  - `jimdoe.com` reached out for `awesomegrizzlybears.jimdoe.com`, `karatedojo-oppeln.jimdoe.com` and `armaniwoe.jimdoe.com`
- Global problem: 123 different countries
- Our soundsquatting domains received different emails related to social-networking invitations and shipment of products

Targeting Sound-dependent users

- Experiment: `youtube.com` and `yewtube.com` by email to a sound-dependent user
- Six popular readers:
  - `Win XP, Win 7, OS X (built-in functionality)`
  - `Thunder, Linux's ORCA, Android's Skyvi (220,000 users)`
- The sound is identical → no mean to distinguish a legitimate link from a malicious
- Proposed Solution: spelling mode

Conclusions

- Uncover soundsquatting
- New type of domain squatting based on words sound-similarity, rather than typos
- We conducted ethical experiments
- Attackers abuse soundsquatting in different forms (scams, malware, ads)
- `AutoSS` as prevention strategy
  - Detect suspicious soundsquatting domains beforehand – TrendMicro

Thanks!

Questions?

Nick Nikiforakis, Marco Balduzzi, Lieven Desmet, Frank Piessens, and Wouter Joosen
(ICS 2014, 12th October, Hong Kong)