

Efficient and Scalable Software Detection in Online Social Networks

Md Sazzadur Rahman, Ting-Kai Huang,
Harsha Madhyastha, Michalis Faloutsos

University of California, Riverside

Problem Statement

- Social malware is rampant on Facebook

PC **Cyber Attacks Jump 81 Percent, Target Mobile, Social Networks**
By  [Chloe Albanesius](#) | April 30, 2012 03:02pm EST |  [4 Comments](#) |  [Email](#)  [Print](#)

CNN **83 million Facebook accounts are fakes and dupes**
By [Heather Kelly](#), CNN
updated 8:32 PM EDT, Thu August 2, 2012 | Filed under: [Social Media](#)

[Charlie Sheen death hoax spreads malware through Facebook](#)

[content.usatoday.com/communities/.../03/charlie-sheen...hoax.../1](#)

Mar 11, 2011 – If you've been clicking on links and videos about **Charlie Sheen's** alleged death, you've been had by the latest social media malware **scam**.

Facebook Scam: "Southwest Free Flights" Will Comment Spam Your Friends



February 22, 2011 by [Jolie O'Dell](#)

 166

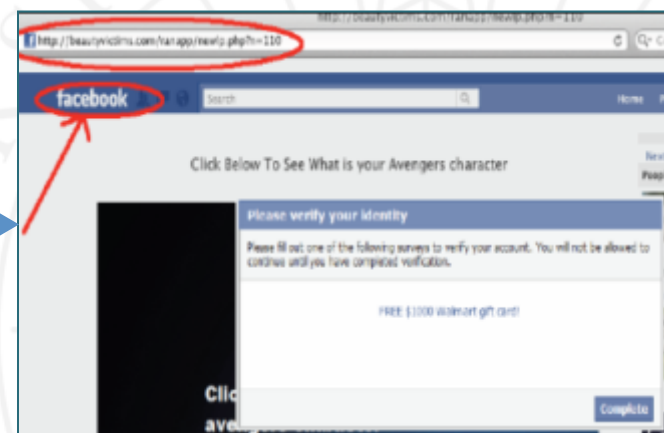
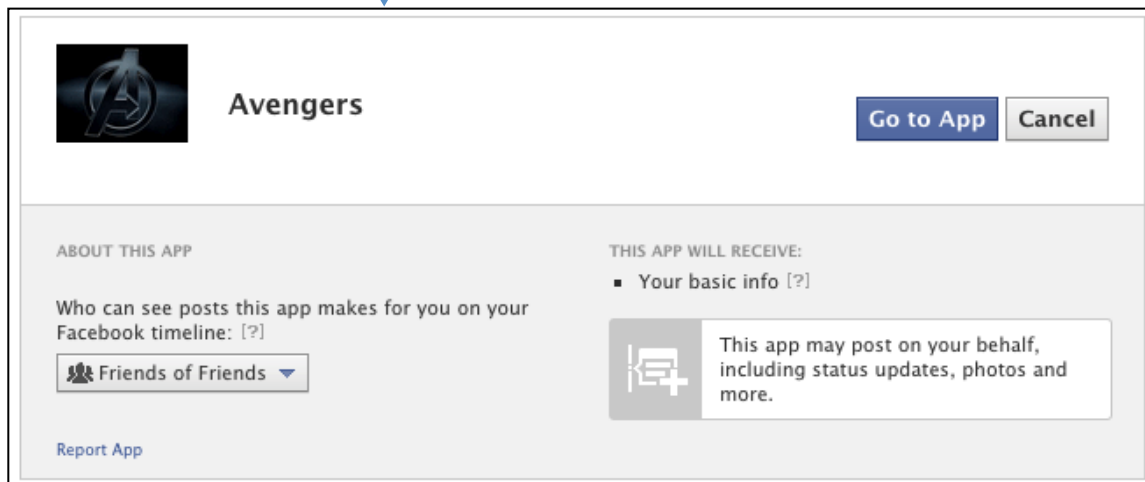
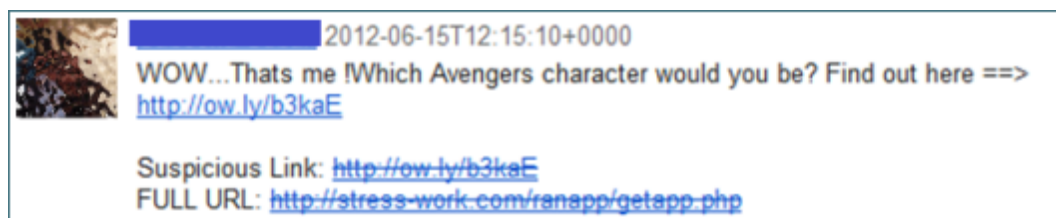
Socware – a new kind of malware

- Social Malware = Socware
- New propagation platform
 - Leverage users for viral propagation unlike email spam



Socware – needs new defenses

- Often hosted inside Facebook domain
 - No blacklist designed to identify malicious apps



Contributions

- Develop MyPageKeeper, a security app in Facebook
 - Obtained 12K users , effectively monitors 2.4M wall
- Design efficient, scalable socware detection method
 - 0.005% false positive, 0.3% false negative rates
 - 12 EC2 micro instances monitors 12K users every 2 hours
- Socware is a new kind of malware
 - 26% of flagged posts point back to Facebook

Roadmap

- MyPageKeeper
- Evaluation
- Interesting results
- Conclusion



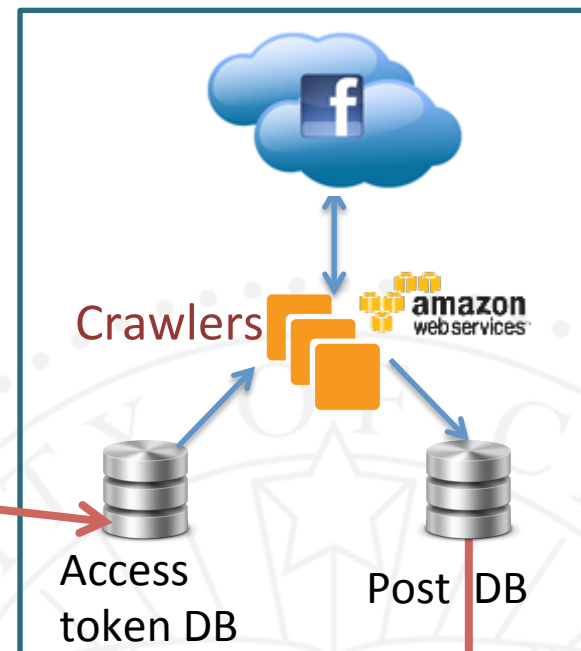
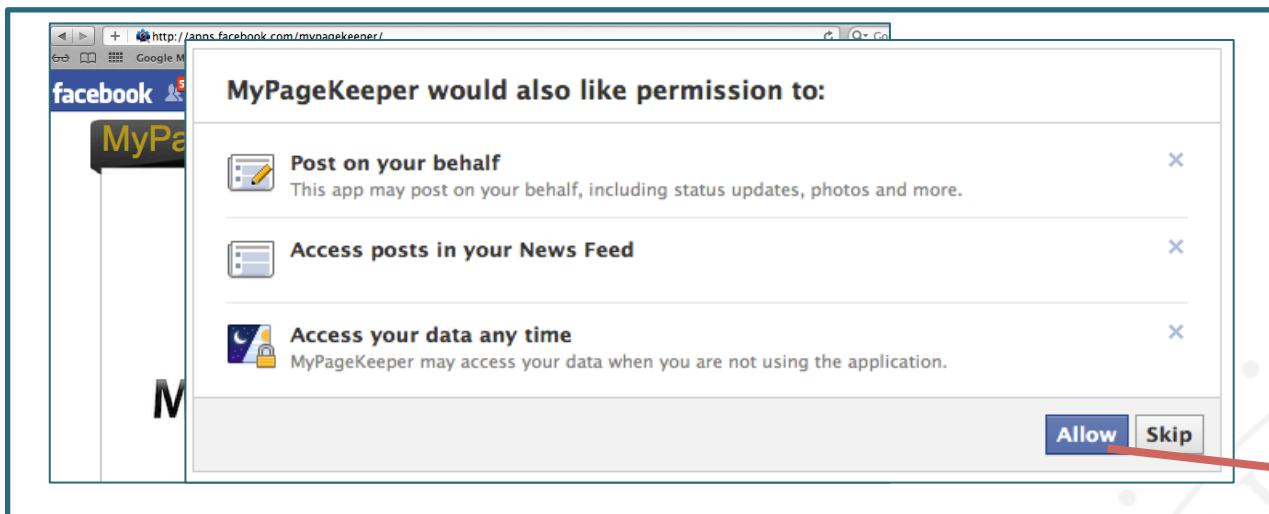
MyPageKeeper: Facebook security app

Launched July, 2011

<https://apps.facebook.com/mypagekeeper/>

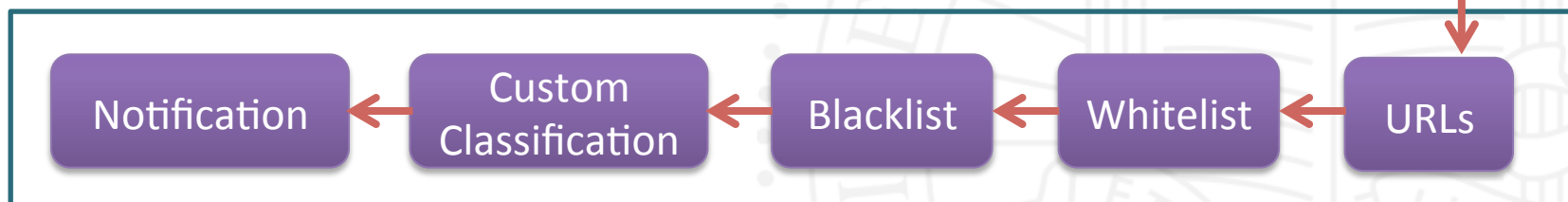


MyPageKeeper operations



1. Obtaining user's authorization

2. Crawling users wall and news feed



3. Identifying socware and notification

Classifying socware


- Previous approaches
 - URL crawling based approach [Thomas, IEEE S&P'11]
 - Expensive (large network latency)
 - Lexical, host properties of landing URL [Ma, SIGKDD'09]
 - Socware often hosted inside Facebook

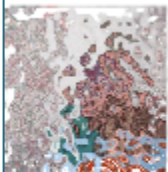
Classifying socware


- Our approach for fast and effective classification:
 - Features readily available in a post
 - Aggregate features across users
- Train Support Vector Machine classifier:
 - Spam keyword score
 - Message similarity
 - No of news feed and wall post
 - No of Like and comments
 - URL obfuscation



Spam keyword score

 [REDACTED] 2012-03-11T00:44:06+0000
Awesome! got a \$100.00 Starbucks GiftCard for **FREE!!** Only 49 cards left!!
Use this promo code when asked: 3iHm78 Get your H&M card here: ->>
<http://giftcaroz.s3-website-us-east-1.amazonaws.com>


 [REDACTED] 2012-08-03T16:43:50+0000
WOW I cant believe that you can see who is viewing your Profile! I just saw My
Top 10 Profile and Photo Peekers and I am SHOCKED!! You can also see WHO
VIEWED YOUR PROFILE here -> <http://bit.ly/M9cf8y>


 [REDACTED] 2012-07-25T05:58:31+0000
OMG I am liking it!!!..... Simple Hilarious !!!.... Wana know your
secret Death <http://bit.ly/LXv0sM>


Awesome, Free, wow, OMG...


Message similarity

Similar text in socware posts

 2012-08-03T16:43:50+0000
WOW I cant believe that you can see who is viewing your Profile! I just saw My Top 10 Profile and Photo Peekers and I am SHOCKED!! You can also see WHO VIEWED YOUR PROFILE here -> <http://bit.ly/M9cf8y>


 2012-08-04T02:26:13+0000
WOW I cant believe that you can see who is viewing your Profile! I just saw My Top 10 Profile and Photo Peekers and I am SHOCKED!! You can also see WHO VIEWED YOUR PROFILE here -> <http://bit.ly/M9cf8y>


 2012-08-03T22:48:39+0000
WOW I cant believe that you can see who is viewing your Profile! I just saw My Top 10 Profile and Photo Peekers and I am SHOCKED!! You can also see WHO VIEWED YOUR PROFILE here -> <http://bit.ly/M9cf8y>

 2012-08-04T01:00:04+0000
WOW I cant believe that you can see who is viewing your Profile! I just saw My Top 10 Profile and Photo Peekers and I am SHOCKED!! You can also see WHO VIEWED YOUR PROFILE here -> <http://bit.ly/M9cf8y>

Different text in non-socware posts

 2012-05-13T22:38:50+0000
they sound very good i like them [http:// www.youtube.com/watch?v=d9NF2edxy-M&feature=share](http://www.youtube.com/watch?v=d9NF2edxy-M&feature=share)

 2012-05-03T21:31:30+0000
Augie Phillips - just your style. [http:// www.youtube.com/watch?v=d9NF2edxy-M&feature=share](http://www.youtube.com/watch?v=d9NF2edxy-M&feature=share)

 2012-05-18T00:07:37+0000
I know the economy is bad, but canada has government funded musical programs, im sure these guys could'a bought more instruments. They do kill this performance though [http:// www.youtube.com/watch?v=d9NF2edxy-M&feature=share](http://www.youtube.com/watch?v=d9NF2edxy-M&feature=share)

 2012-05-22T12:57:01+0000
Check this version out. [http:// www.youtube.com/watch?v=d9NF2edxy-M&feature=share](http://www.youtube.com/watch?v=d9NF2edxy-M&feature=share)

Other features

- News feed post and wall post count
- Like and comment count
- URL Obfuscation



Implementation

- Non-trivial engineering
 - Apache, Django, Postgres
- Designed for scale
 - Crawler instances
 - 12 micro instances in Amazon EC2 to run crawlers
 - Checker instances
 - Work load partitioned across checkers

Roadmap

- MyPageKeeper
- **Evaluation**
- Interesting results
- Conclusion



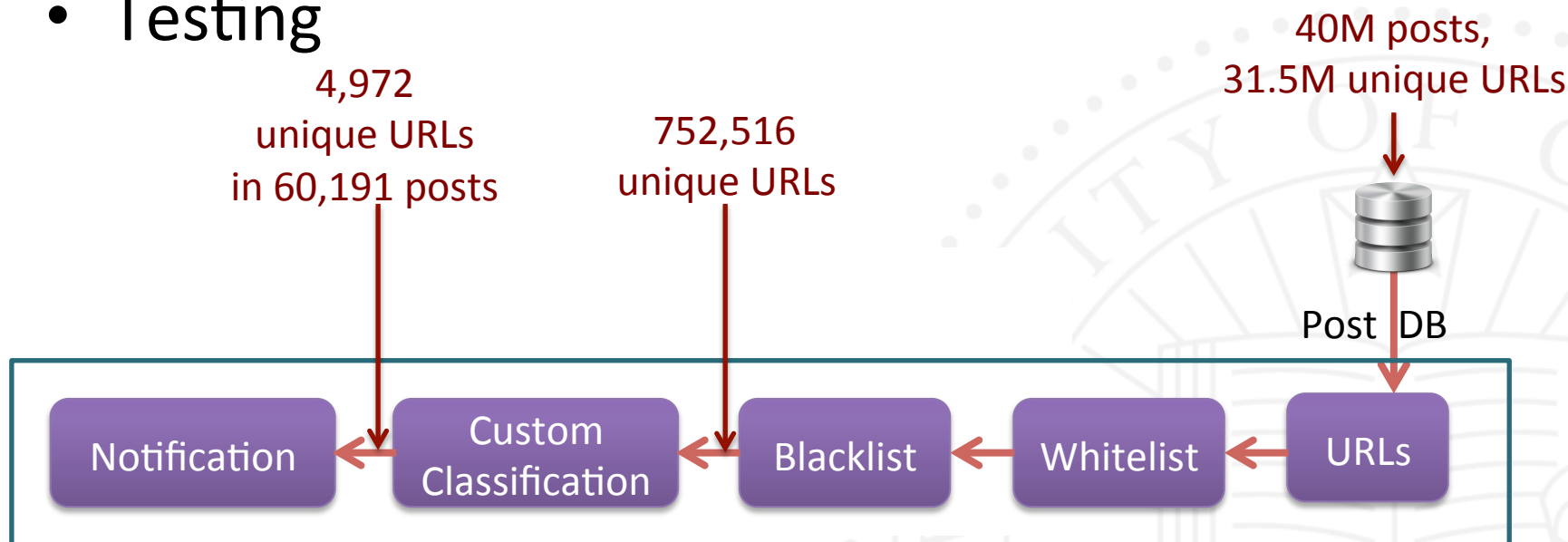
Evaluation: Dataset

Collected from June to October 2011

Data	Total count	# of distinct URLs
MyPageKeeper users	12.5K	-
Friends of MPK users	2.4M	-
News feed posts	38.8M	29.5M
Wall posts	1.8M	1.5M
User reports	679	333

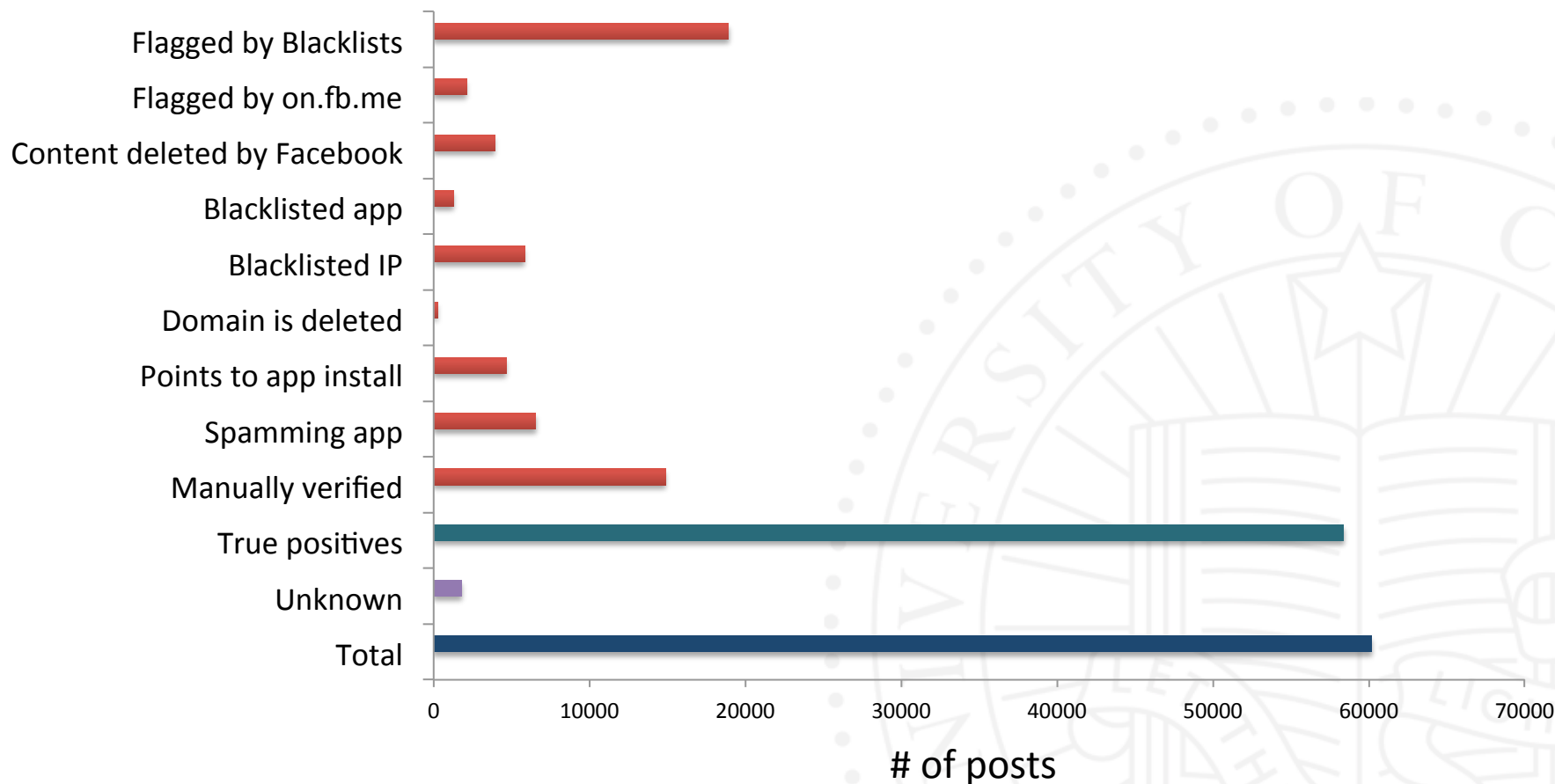
Is MyPageKeeper accurate?

- Training
 - 2500 positive and 5000 negative sample posts
- Testing



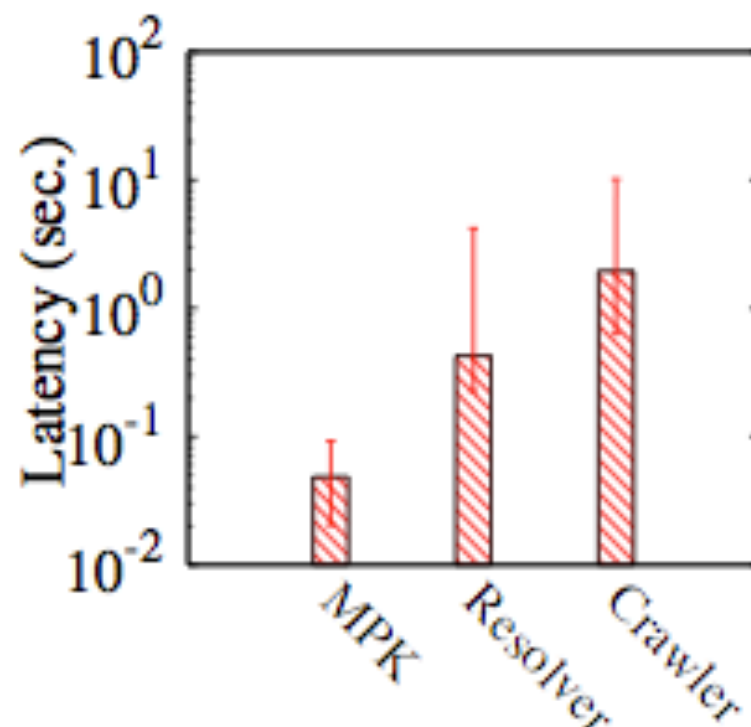
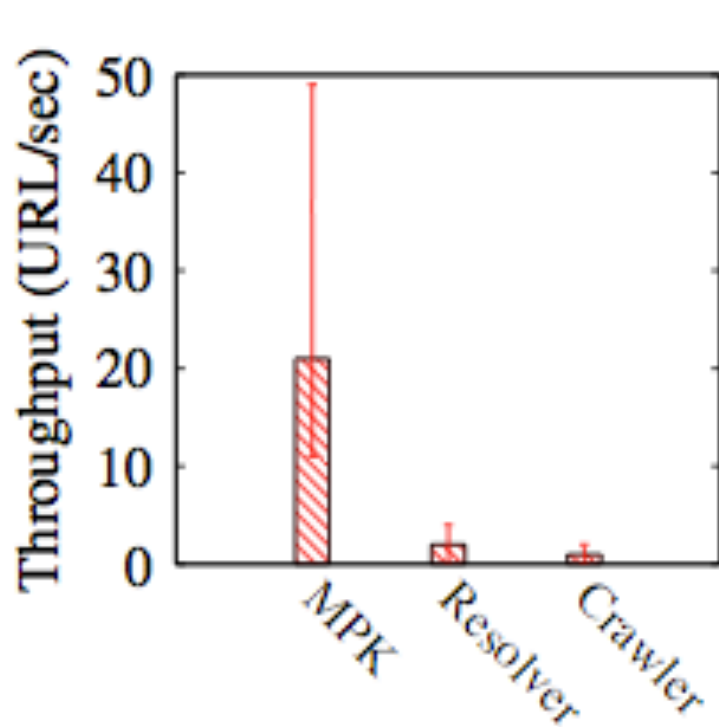
MyPageKeeper is accurate

97% socware identified by classifier true positives.



MyPageKeeper is efficient

20x better throughput than other approaches



Blacklist does not suffice

96.6% socware identified by classifier

Source	% of posts	Overlap with classifier (# of URLs)
Google SBA2	0.4%	0
Phishtank	0.5%	1
Malware Norm	0.2%	0
Joewein	0.7%	11
APWG	0.6%	0
Spamcop	1.0%	0
All blacklists	3.4%	12
MyPageKeeper Classifier	96.6%	-

Roadmap

- MyPageKeeper
- Evaluation
- **Interesting results**
- Conclusion



Socware is prevalent in Facebook

- 49% MyPageKeeper users exposed to socware
- 26% of socware post points back to Facebook



Novel Parasitic behavior: Like-as-a-Service

Artificially inflate number of Likes in Facebook page

Play and Win

2011-11-19T21:18:32+0000

Just got a better score on Raging Bid's Bouncing Balls contest and I am now in 14660th place. I am getting closer to winning a Sony Bravia 3D HDTV. Who thinks they can beat my score? Click here to try: <http://sups.us/7e58d9>

facebook

Search for people, places and things

Dave White Acura Play & Win

Like 156k

Play & Win!

How to Play Rules Leaderboard Best Friends

"LIKE" OUR PAGE TO PLAY & WIN

Try your skills playing Gem Swap II for a chance to win: a 42 inch Flat Screen TV

Dave White Acura

Like Message

Why is Play & Win asking for these permissions?

From Play & Win: We need permission to keep you informed of your current position and any new games available to you.

Allow Skip

Novel Parasitic behavior: Like-as-a-Service

- Identified 721 distinct LaaS URLs in Facebook
- Other Like-as-a-Service app: Games, FanOffer, Latest Promotion

Conclusion

- We presented MyPageKeeper, a security app
- We show that
 - Reach of socware is widespread
 - Existing defenses not sufficient to identify socware
 - MyPageKeeper is accurate and efficient
- We identified LaaS, new aggressive marketing trend

Thank you!

Questions?



<https://apps.facebook.com/mypagekeeper/>
<http://mypagekeeper.org>