

# Serving Ads from localhost for Performance, Privacy, and Profit

**Saikat Guha**, Bin Cheng, Alexey Reznichenko,  
Hamed Haddadi, Paul Francis

Max Planck Institute for Software Systems  
Kaiserslautern-Saarbrücken, Germany

October 22, 2009

Nothing certain except...

# Death Taxes Advertising

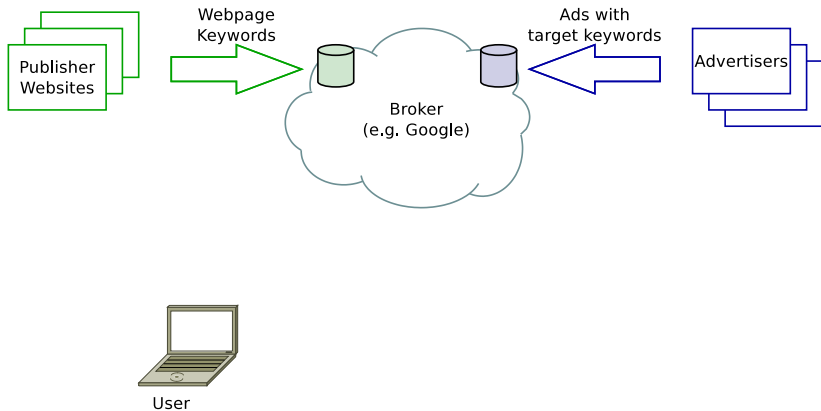
# Ads Today Suck

- ▶ Annoying
  - ▶ **Quality sucks**
  - ▶ So they push quantity, obtrusiveness
- ▶ Slow
  - ▶ **Multiple round-trips** to distant ad server
  - ▶ Stalls webpage rendering
- ▶ Invade Privacy
  - ▶ Google/DoubleClick **sees every website we visit**
  - ▶ Disgruntled employee in league with insurance company... game over.

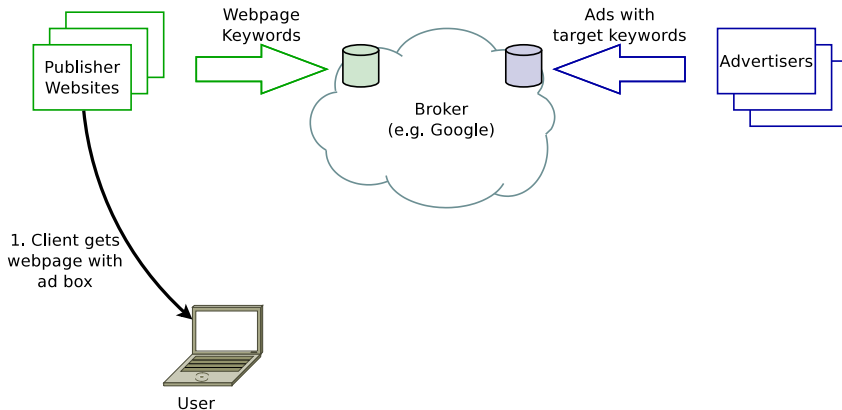
# Ads Today Suck

- ▶ Annoying
  - ▶ **Quality sucks**
  - ▶ So they push quantity, obtrusiveness
- ▶ Slow
  - ▶ **Multiple round-trips** to distant ad server
  - ▶ Stalls webpage rendering
- ▶ Invade Privacy
  - ▶ Google/DoubleClick **sees every website we visit**
  - ▶ Disgruntled employee in league with insurance company... game over.

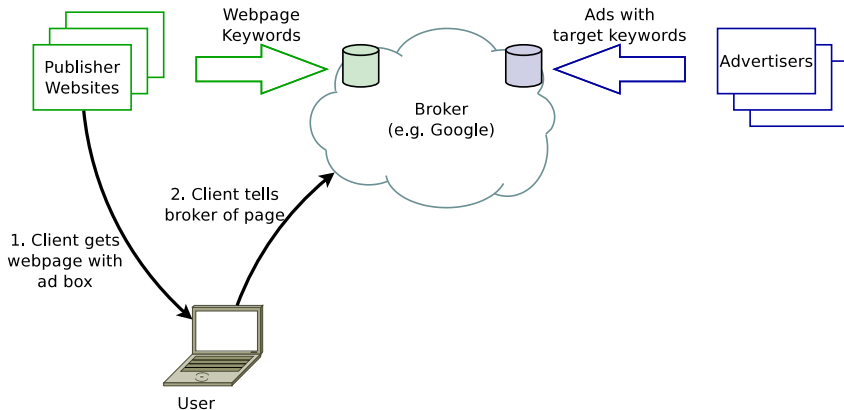
# Ads Today Suck



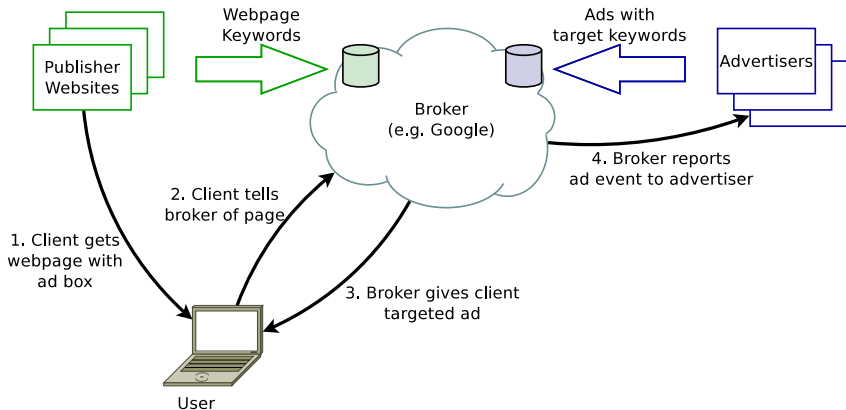
# Ads Today Suck



# Ads Today Suck

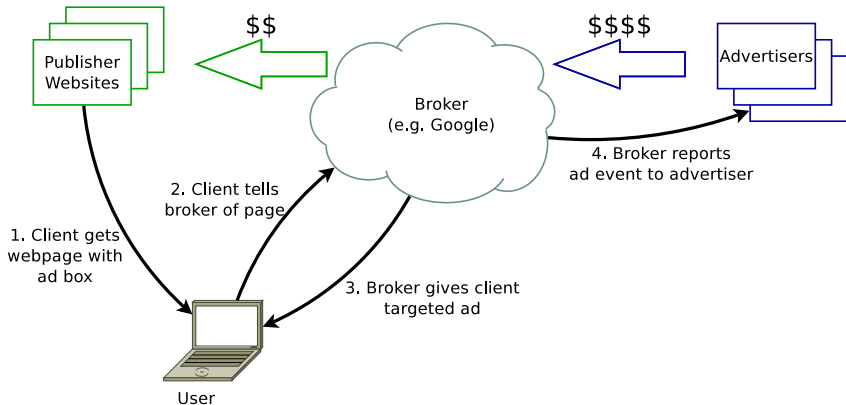


# Ads Today Suck

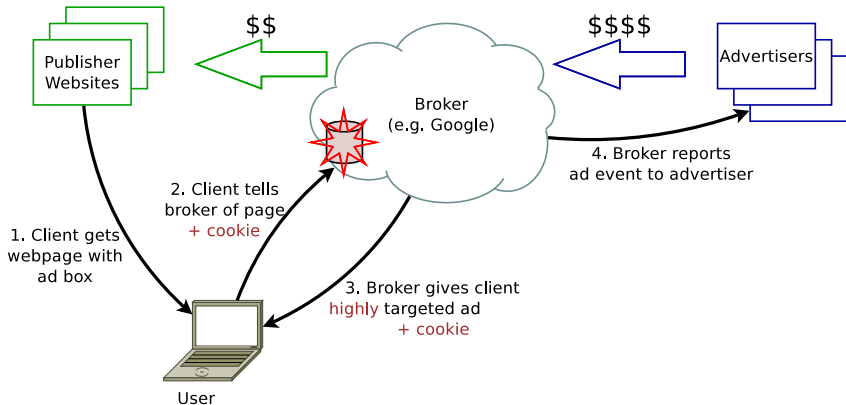




# Ads Today Suck



# Ads Today Suck



# Ads Today Suck

- ▶ Annoying
  - ▶ Quality sucks
  - ▶ So they push quantity, obtrusiveness
- ▶ Slow
  - ▶ Multiple round-trips to distant ad server
  - ▶ Stalls webpage rendering
- ▶ Invade Privacy
  - ▶ Google/DoubleClick sees every website we visit
  - ▶ Disgruntled employee in league with insurance company... game over.

# Can we design?

## Practical Private Advertising

1. ~~Clean~~ Dirty slate
  - ▶ Supports today's advertising business model
2. Private enough
  - ▶ To convince privacy-advocates and governments
3. Good at targeting
  - ▶ Increased privacy begets better personalization
4. Scalable
  - ▶ yada yada yada

# Can we design?

## Practical Private Advertising

1. ~~Clean~~ Dirty slate
  - ▶ Supports today's advertising business model
2. **Private enough**
  - ▶ To convince privacy-advocates and governments
3. Good at targeting
  - ▶ Increased privacy begets better personalization
4. Scalable
  - ▶ yada yada yada

# Can we design?

## Practical Private Advertising

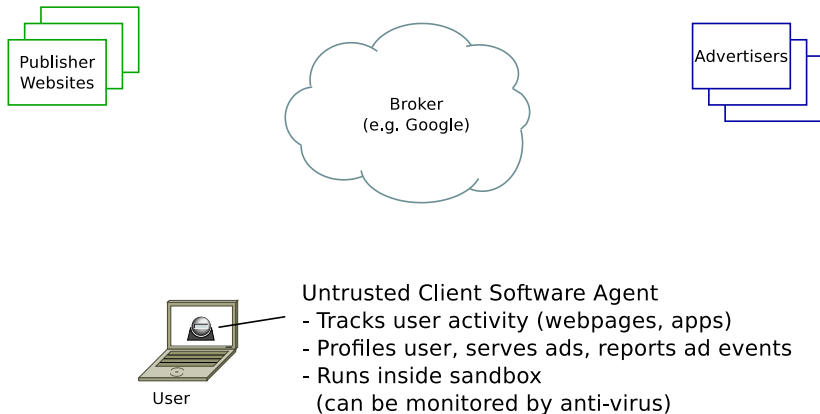
1. ~~Clean~~ Dirty slate
  - Supports today's advertising business model
2. Private enough
  - To convince privacy-advocates and governments
3. **Good at targeting**
  - Increased privacy begets better personalization
4. Scalable
  - yada yada yada

# Can we design?

## Practical Private Advertising

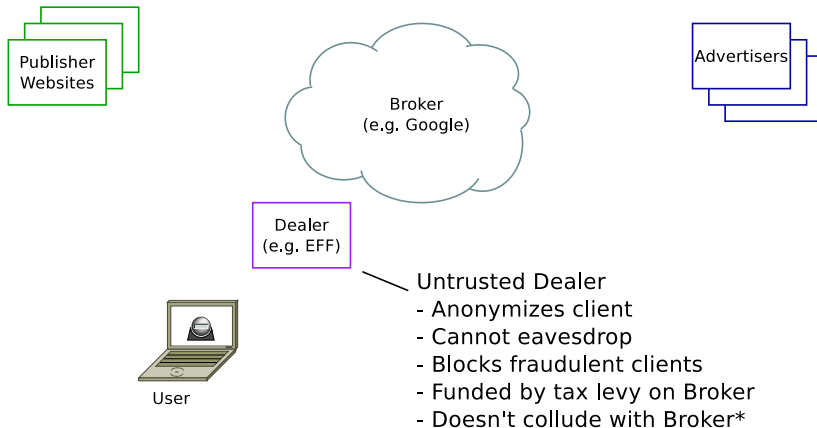
1. ~~Clean~~ Dirty slate
  - Supports today's advertising business model
2. Private enough
  - To convince privacy-advocates and governments
3. Good at targeting
  - Increased privacy begets better personalization
4. Scalable
  - yada yada yada

# Privad Big Picture

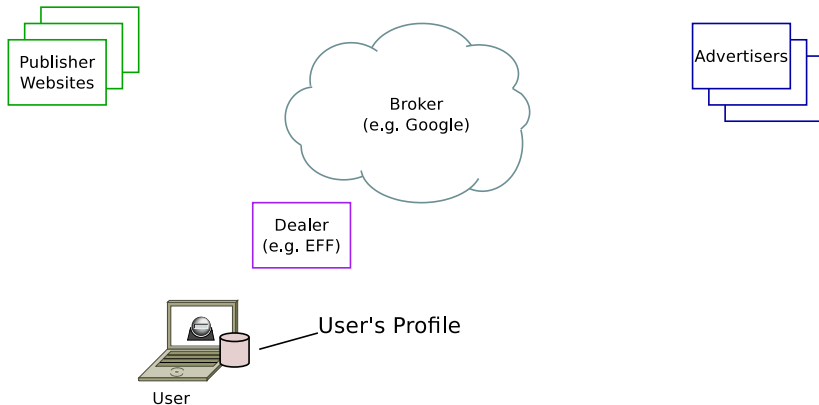




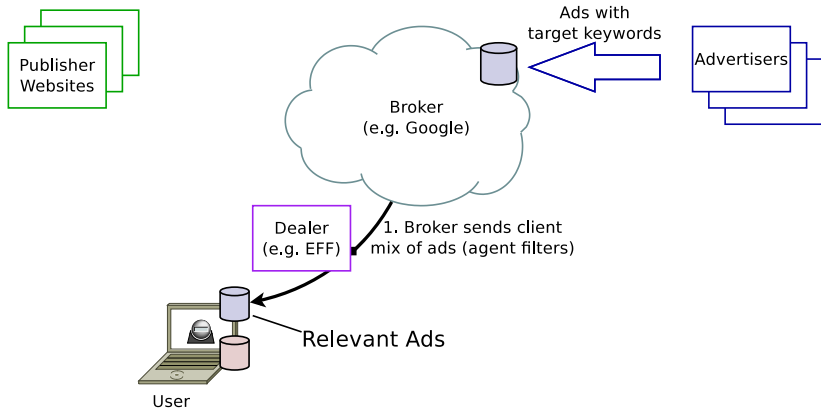
# Privad Big Picture



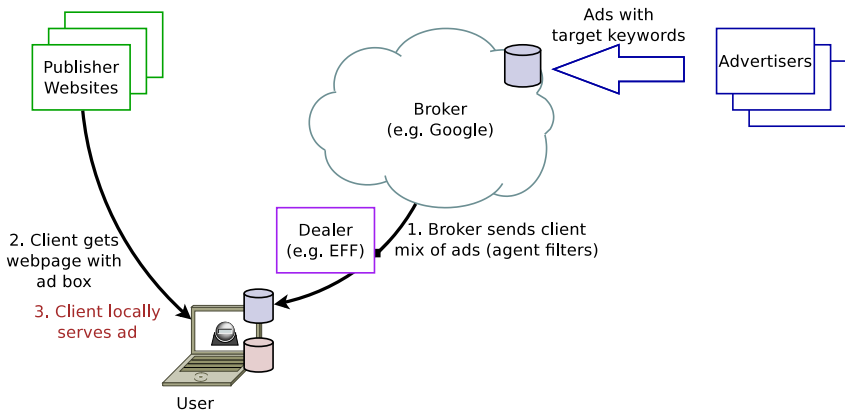
# Privad Big Picture



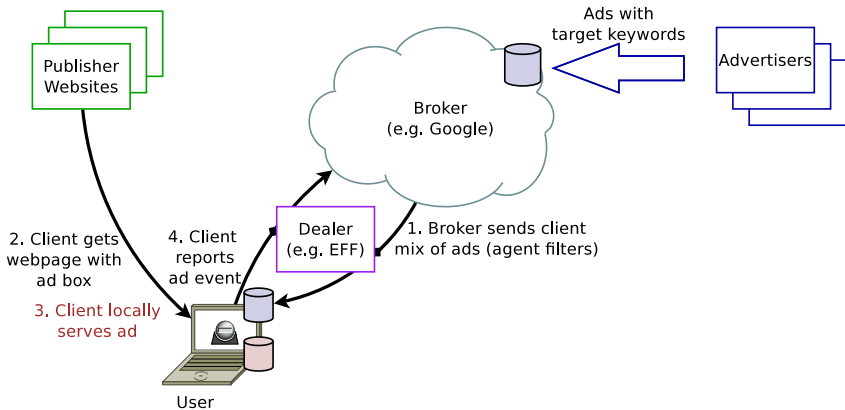
# Privad Big Picture



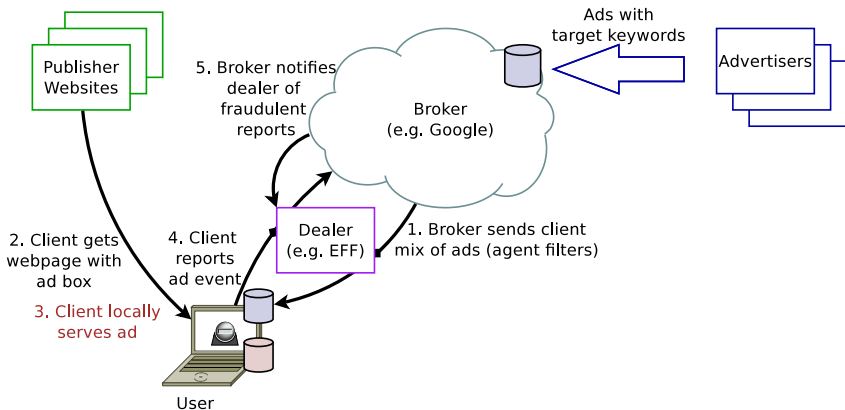
# Privad Big Picture



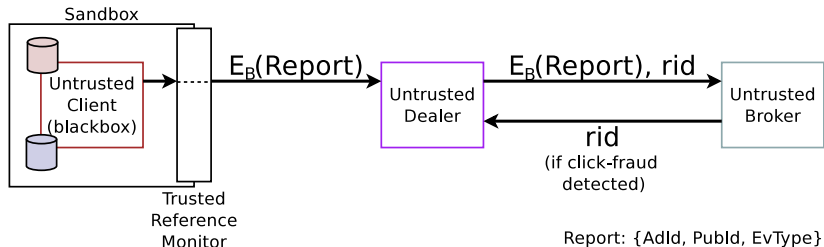
# Privad Big Picture



# Privad Big Picture














# Privad Big Picture



- ▶ Dealer learns client  $X$  clicked on some ad
- ▶ Broker learns someone clicked on ad  $Y$

# How Deep the Rabbit Hole Goes...

-  Google Ads Study
-  CoDeeN User Study
-  User profiling
-  Dissemination
-  Auctions
-  Click-Fraud
-  Anonymizing the Click
-  Crypto w/ optimizations
-  Reference Monitor
-  Privacy Analysis
-  Implementation and Microbenchmarks



# Deployability

Step 1: Convince privacy-advocates and antivirus-firms

- ▶ Not only “not bad”, but in fact “good alternative”

**NOT** for those:

- ▶ who don't see/click ads today
- ▶ use AdBlockers

For people who make Google \$20B every year.

\$\$\$\$ Installed by default with privacy-conscious browsers

Step 3: Convince or compel Google, or compete

- ▶ Better value, lower risk
- ▶ Or apply pressure through regulatory agencies

# Deployability

## Step 1: Convince privacy-advocates and antivirus-firms

- ▶ Not only “not bad”, but in fact “**good alternative**” to privacy-compromising cloud-based advertising
- ▶ Ensure user experience not degraded in any way

## Step 2: **Multiple** deployment vehicles

\$ Standalone, or bundle third-party software

- ▶ Surprisingly tenable. Based on CoDeeN study [▶ go](#).

\$\$ Or bundled with third-party software

\$\$\$\$ Installed by default with privacy-conscious browsers

## Step 3: Convince or compel Google, or **compete**

- ▶ Better value, lower risk
- ▶ Or apply pressure through regulatory agencies

# Deployability

Step 1: Convince privacy-advocates and antivirus-firms

- ▶ Not only “not bad”, but in fact “good alternative” to privacy-compromising cloud-based advertising
- ▶ Ensure user experience not degraded in any way

Step 2: Multiple deployment vehicles

- \$ Standalone, or bundle third-party software
  - ▶ Surprisingly tenable. Based on CoDeeN study [▶ go](#).

\$\$ Or bundled with third-party software

\$\$\$\$ Installed by default with privacy-conscious browsers

Step 3: Convince or compel Google, or compete

- ▶ Better value, lower risk
- ▶ Or apply pressure through regulatory agencies

# Deployability

Step 1: Convince privacy-advocates and antivirus-firms

- ▶ Not only “not bad”, but in fact “good alternative” to privacy-compromising cloud-based advertising
- ▶ Ensure user experience not degraded in any way

Step 2: Multiple deployment vehicles

- \$ Standalone, or bundle third-party software
  - ▶ Surprisingly tenable. Based on CoDeeN study [▶ go](#).

\$\$ Or bundled with third-party software

\$\$\$\$ Installed by default with privacy-conscious browsers

Step 3: Convince or compel Google, or compete

- ▶ Better value, lower risk
- ▶ Or apply pressure through regulatory agencies

# Deployability

Step 1: Convince privacy-advocates and antivirus-firms

- ▶ Not only “not bad”, but in fact “good alternative”

@Microsoft: Want to “frakkin’ kill” Google?

Step 2: M

- ▶ Already own endhost OS and browser. Can deploy agent easily.

- ▶ Already an ad broker.

\$

- ▶ Already got FTC, anti-virus, and privacy watchdogs on your side (against Google).

Step 3: C

- ▶ Better value, lower risk
- ▶ Or apply pressure through regulatory agencies

# Status

- ▶ Protocols defined
  - ▶ Stable: Dissemination, Reporting, Reference Monitor, Crypto w/ optimizations
  - ▶ May evolve: Auctions, Click-Fraud
- ▶ Implemented, pilot deployment
  - ▶ Firefox plugin
- ▶ Next steps
  - ▶ Talk to privacy-advocates, brokers
  - ▶ Real deployment and measurements...












# Summary

- ▶ Practical privacy-preserving online advertising
  - ▶ Better targeting, significantly better privacy, no changes to business models, scalable
- ▶ Full system\*
  - ▶ Profiling, Dissemination, Auctions, Reporting, Click-Fraud, Scalability, Auditing, Deployment incentives
- ▶ Call to action
  - ▶ If you hate online ads, help fix it!
  - ▶ Lots of interesting research directions  
(and low-hanging fruit!)

---

\*See <http://mpi-sws.org/tr/2009-004.pdf>

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)



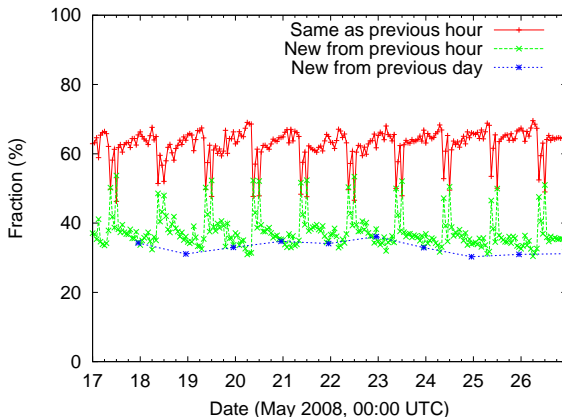
# Understanding Google Search Ads

- ▶ Sampled Google search ads for 1 month
- ▶ Every 30 minutes
- ▶ 1.3K random keywords  
(from 100K keyword dictionary)
- ▶ Geo-diverse vantage points

# Understanding Google Search Ads

**Ad Skew:** 10% (generic) ads shown 80% of the time.

**Ad Churn:** 30%–40% ads change hour-hour/day-day.  
5%–10% replaced permanently.



# Understanding Google Search Ads












**Ad Skew:** 10% (generic) ads shown 80% of the time.

**Ad Churn:** 30%–40% ads change hour-hour/day-day.  
5%–10% replaced permanently.

Design implications:

- ▶ Generic ads: may disseminate widely and cache.
- ▶ Rest **cannot flood**. Update traffic too high.

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

# Understanding CoDeeN users

- ▶ CoDeeN click stream for 1 month
- ▶ Filtered bots using CoDeeN's bot detector
- ▶ 31K users; some bots still

# Understanding CoDeeN users

**Ad Block:** Only 10–20%; tad low?

**Third-party Crap:** 21%; surprisingly high?

	Users	Ad Views	CTR	3rd-Party Toolbars	Ad Blockers
China	7308	39K	0.5 %	22 %	12 %
Saudi Arabia	6710	56K	2.7 %	40 %	9 %
United States	1420	19K	0.9 %	13 %	17 %
U.A.E	1322	8K	1.7 %	35 %	8 %
Germany	956	5K	1.5 %	7 %	19 %
<u>Worldwide</u>	30987	189K	2.5 %	21 %	12 %

# Understanding CoDeeN users












**Ad Block:** Only 10–20%; tad low?

**Third-party Crap:** 21%; surprisingly high?

Deployment implications:

- ▶ Ad-supported business models still viable
- ▶ Many users will install anything, and forget?  
(if it isn't disruptive)
- ▶ Even for somewhat tech. savvy users; likely more so for typical users

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)



# Profiling the User












Multiple complementary approaches

- ▶ **Crawling:** Broker maps website-keywords. Client queries anonymously.
  - ▶ Identical to today (but private)
  - ▶ Sophisticated classifiers
  - ▶ Not for sites with user login. Or desktop apps.
- ▶ **Scraping:** Client scrapes websites
  - ▶ Simple classifiers
  - ▶ May be combined with anonymized access to sophisticated classifiers
  - ▶ Works for sites with user login. And desktop apps.

# Profiling the User

- ▶ **Metadata:** Website embeds keywords in webpage served.
  - ▶ Incentivise by offering part of ad revenue
  - ▶ Client tracks and sends in report which websites contributed profile info that led to click. (different from website showing adbox)
- ▶ **User/Social Feedback:** Direct user feedback (+/-) on ads. Client may also affect clients of OSN friends.

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

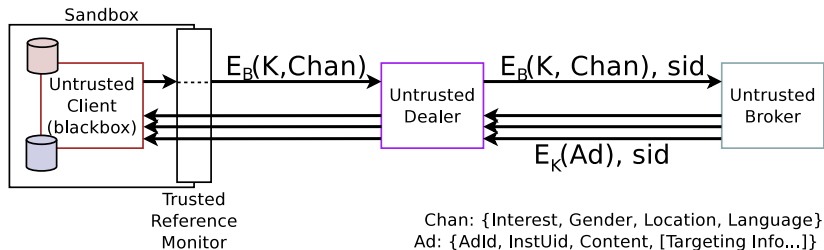
# Ad Dissemination

- ▶ Broker doesn't learn anything about client
- ▶ Simplest: Flood all ads to all clients
  - ▶ Won't work. Easily 2+ GB per month, probably much more. Based on Google Ads study [▶ go](#).
- ▶ We propose **privacy-preserving Pub-Sub**

# Ad Dissem: Privacy-preserving Pub-Sub












- ▶ Define categories of ads
  - ▶ Amazon defines over 100K of these, e.g.  
`electronics.camera+photo.panasonic.camcorders.-`  
`accessories.memory+media.media.minidv`
  - ▶ Actual number is scalability-privacy tradeoff
- ▶ Client subscribes to channels (through Dealer)
  - ▶ Channel is ad category plus broad demographics  
e.g. gender, location, language
- ▶ Broker publishes ads (through Dealer)
  - ▶ Ads nearing daily budget not published
  - ▶ Not all ads published match client because of sensitive demographics e.g. marital-status
  - ▶ Published ads expire after some time

# Ad Dissem: Privacy-preserving Pub-Sub



- ▶  $K$  unique to this subscription
- ▶ Dealer learns client  $X$  subscribed to some chan
- ▶ Broker learns someone subscribed to channel  $Y$
- ▶ Broker cannot link multiple subscriptions from same client. (Otherwise can build up profile over time)

# Questions?

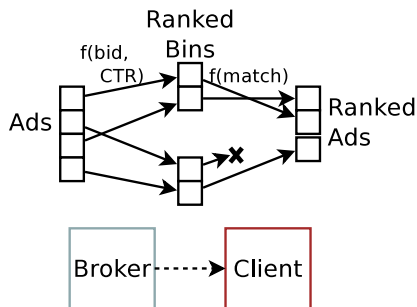
-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

# Auctions

- ▶ Fair marketplace where advertisers influence frequency and position of ads through bids
- ▶ Preserve user privacy, and advertiser bid privacy
- ▶ **Design-I:** Simple Auction
- ▶ **Design-II:** Combined Auction
  - ▶ Identical to **Google's GSP Auction** today
- ▶ Will evolve as new approaches are added

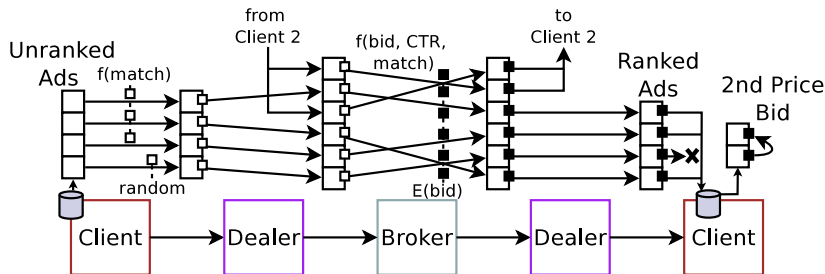


# Auctions: Simple Auction














- ▶ Coarse-grained but very simple
- ▶ Channel granularity. Bins ranked by global metrics. Ads in bins ranked by user metrics.
- ▶ No changes to protocols; no impact on privacy

# Auctions: Combined Auction



- ▶ Identical to Google model. Incl. 2<sup>nd</sup> price.
- ▶ Fine-grained. Per user ads ranked by global and user metrics.
- ▶ Private for both user, and advertiser

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

# Detecting Click-Fraud

- ▶ Client is untrusted. Protocol is public.
  - ▶ Much like today (browser, HTTP)
- ▶ **No silver bullet.** Constant arms race.
- ▶ Basic approach: **Defense in depth**
  - ▶ **Lots** of overlapping detection mechanisms
  - ▶ Each requires time and effort to circumvent
  - ▶ Together raise the bar considerably
- ▶ Will evolve as new approaches are added












# Detecting Click-Fraud

- ▶ **Thresholds:** Dealer flags clients with abnormally high number of subscriptions, views, clicks, or click-through ratio.
  - ▶ Forces attacker to use botnet
  - ▶ Cannot use same botnet for multiple attacks
- ▶ **Blacklists:** Dealers use lists of known bots (from antivirus or network telescope). Dealers share list of banned clients.
  - ▶ Limits window of time a bot is useful.
- ▶ **Honeyfarms:** Broker operates honeyfarm susceptible to botnet infections.
  - ▶ Honeyfarm detection armsrace. Advantage Broker.

# Detecting Click-Fraud

- ▶ **Historical Statistics:** Broker tracks historical volume of views, and click-through-rates for each publisher, and each advertiser. Flags abrupt changes.
  - ▶ Forces gradual attacks
  - ▶ Buys time for other approaches
- ▶ **Bait Ads:** Synthesized ads with content from one ad, and targeting information from a completely different ad. Expect few legit clicks.
  - ▶ Think CAPTCHAs for ads.
  - ▶ Attacker could use cheap human labor
    - ▶ Potentially more time-consuming
    - ▶ Bait = semantic. CAPTCHA = syntactic.
    - ▶ Especially in non-English-native countries

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

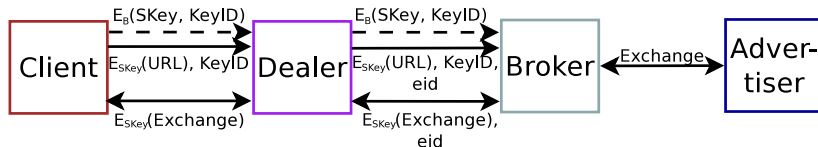
# Anonymizing the Click

## User Privacy vs. Advertiser:

- ▶ Open question: **What is “good enough”?**
- ▶ Advertiser can see IP address if user clicks; also knows targeting info of ad that matched user. May link multiple clicks.
  - ▶ But clicks are rare; but payoff could be significant
  - ▶ Anonymizing proxy? Proxy learns profile. TOR?
  - ▶ Approach: **anonymizing the click**
  - ▶ Good enough? Don't know.
- ▶ Advertiser may link to user identity through credit-card
  - ▶ Single-use credit card tokens?
- ▶ Or shipping address for physical products
  - ▶ Anonymous remailers? (i.e. TOR for post)














# Anonymizing the Click



- ▶ Client pre-establishes (single-use) SKey
- ▶ User privacy preserved
  - ▶ Broker, Advertiser don't learn which Client.
  - ▶ Dealer doesn't learn what Advertiser.
- ▶ Broker drops out at some point
  - ▶ Informs user what advertiser can learn
  - ▶ Open question: when?
    - ▶ After landing page?
    - ▶ Certainly before user inadvertently reveals PII
    - ▶ Or advertiser could encrypt exchange

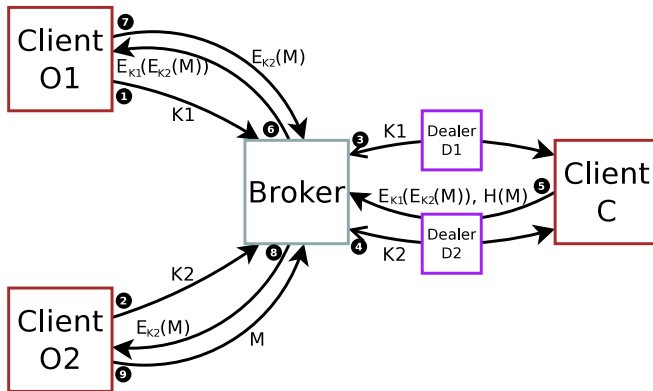
# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

# Cryptographic Overheads

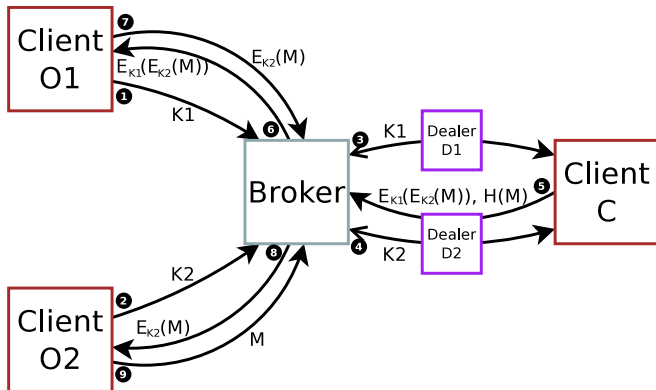
- ▶ Symmetric key operations quite fast
  - ▶ With hardware, can operate at line speeds
- ▶ Biggest concern: public-key operations
- ▶ Insight: Leverage idle clients
  - ▶ Save on datacenter costs (cores, cooling)

# Offloading Public-Key Operations



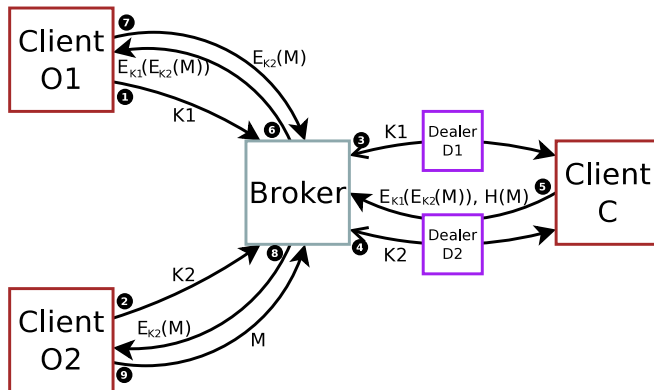
- ▶ Broker learns  $M$  **without any public-key ops.**
- ▶  $D1, D2$  do not learn  $M$ . Can't MITM.

# Offloading Public-Key Operations














- ▶ Broker, O1, O2 do not learn client identity.
- ▶ New keys for each message. Broker **cannot link**.

# Offloading Public-Key Operations



- **20x** performance improvement in real deployment. See Microbenchmarks [► go](#)

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

# Reference Monitor Design

- ▶ **Blackbox monitoring** of client
  - ▶ Allows brokers to have proprietary code in client
  - ▶ Allows for complex clients
- ▶ Monitor itself **very simple**
  - ▶ Open source
  - ▶ Created by privacy-advocates, or anti-virus vendor, or browser vendor, and verified by another
  - ▶ Correctness verified manually














# Reference Monitor Design


What it does:

- ▶ **Validates message** contents
  - ▶ Client gives it plain text
  - ▶ Monitor validates, then encrypts
  - ▶ Thus no covert channel in salts, paddings, etc.
- ▶ Source of all randomness in messages
  - ▶ Specifically, **generates session keys** for Pub-Sub Ad Dissemination [▶ go](#)
  - ▶ Thus no covert channel in keys
- ▶ Staggers message bursts
  - ▶ May **add arbitrary delay/jitter**
  - ▶ Disrupt any covert channel in message timing
  - ▶ All protocol exchanges designed with this in mind (i.e. completely asynchronous)

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

# User Privacy

- ▶ vs. Publisher
  - ▶ Privad doesn't change anything here
  - ▶ Client free to use anonymizing proxies as today
- ▶ vs. Advertiser
  - ▶ In theory, Privad doesn't change anything
  - ▶ In practice, Privad has better targeting. Advertiser can infer more on click.
  - ▶ Approach: Anonymizing the Click 
- ▶ vs. Broker, vs. Dealer
  - ▶ **Unlinkability**: no user information can be associated with user's identity using internal or external means.












# User Privacy

1. **No Personally Identifying Information (PII)**, except IP address, explicitly **leaves client**
  - ▶ Validated by Reference Monitor [▶ go](#)
2. Dealer knows IP address, but **no other user information**
3. Broker has access to user information, but **not IP address**
  - ▶ Cannot link user information from multiple messages over time
  - ▶ Very little user information in any given message
  - ▶ Cannot de-anonymize user using external databases

# Privacy Non-Goals

- ▶ Protecting ad targeting information
  - ▶ Desirable or undesirable debatable
  - ▶ e.g. cigarette companies targeting pre-teens
  - ▶ OTOH, targeting as competitive edge
- ▶ Protecting against malware
  - ▶ Malware can see client data
    - ▶ OS could impose process based ACL (e.g. SELinux)
  - ▶ But fundamentally, malware can anyway spy on user

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)

# Implementation and Pilot Deployment

## Implementation:

- ▶ Client and Simple monitor
  - ▶ 150kB Firefox addon<sup>†</sup>; 4.2K LoC
  - ▶ Simple profiling (Facebook, Google Ad Preferences)
  - ▶ Ad dissemination, combined auctions, ad event reporting, crypto offload
- ▶ Broker, Dealer
  - ▶ Java servlet; 800 LoC and 300 LoC
- ▶ Wire protocol
  - ▶ JSON over HTTP; 2.4K LoC
  - ▶ In retrospect, mistake. Everything optimized; serialization/deserialization for text-based RPC now bottleneck.

---

<sup>†</sup>See <http://adresearch.mpi-sws.org>

# Implementation and Pilot Deployment

## Deployment:

- ▶ Client scrapes Google ads, adds synthetic targeting and bid information
- ▶ Broker publishes to other clients
- ▶ Clients inject ads into existing Google adboxes
- ▶ Handful of alpha testers ( $\sim 70$ )



# Implementation and Pilot Deployment












## Challenges:

- ▶ Webpage **scraping is laborious**
  - ▶ 20% of client code for just 2 websites
  - ▶ Not to mention keeping up-to-date
  - ▶ Could crowd-source module development/maintenance
  - ▶ Could build tools to generate scraping code
- ▶ Defining ad categories and **mapping scraped information non-trivial**
  - ▶ Currently, scraped info well structured. Categories superset of scraped info. Mapping trivial.
  - ▶ Problematic for unstructured information
  - ▶ Potentially, one-time manual effort plus small maintenance effort

# Microbenchmarks

- ▶ Client: workstation, laptop, netbook
  - ▶ Serving: < 30ms for 100K local ads; **10x faster than today**
  - ▶ Crypto: **Unnoticeable** 50–200ms; anyway async.
- ▶ Broker: 3GHz single-core
  - ▶ Subscribe/Reports without offload: bottleneck public-key ops. ( $\sim 280$  req/sec)
  - ▶ with offload: **bottleneck RPC** >6K req/sec
  - ▶ Publish: bottleneck symmetric-key ops. 750M ads/day
  - ▶ Auctions: depends on privacy 30K–80K ads/sec
- ▶ Dealer: 3GHz single-core
  - ▶ 200K clients per core. Client polls; **bottleneck sockets**

# Questions?

-  Google Ads Study ..... [▶ go](#)
-  CoDeeN User Study ..... [▶ go](#)
-  User profiling ..... [▶ go](#)
-  Dissemination ..... [▶ go](#)
-  Auctions ..... [▶ go](#)
-  Click-Fraud ..... [▶ go](#)
-  Anonymizing the Click ..... [▶ go](#)
-  Crypto w/ optimizations ..... [▶ go](#)
-  Reference Monitor ..... [▶ go](#)
-  Privacy Analysis ..... [▶ go](#)
-  Implementation and Microbenchmarks ..... [▶ go](#)