

The Lane's Gifts v. Google Report

By **Alexander Tuzhilin**

Professor of Information Systems at the Stern School
of Business at New York University,
Report published July 2006

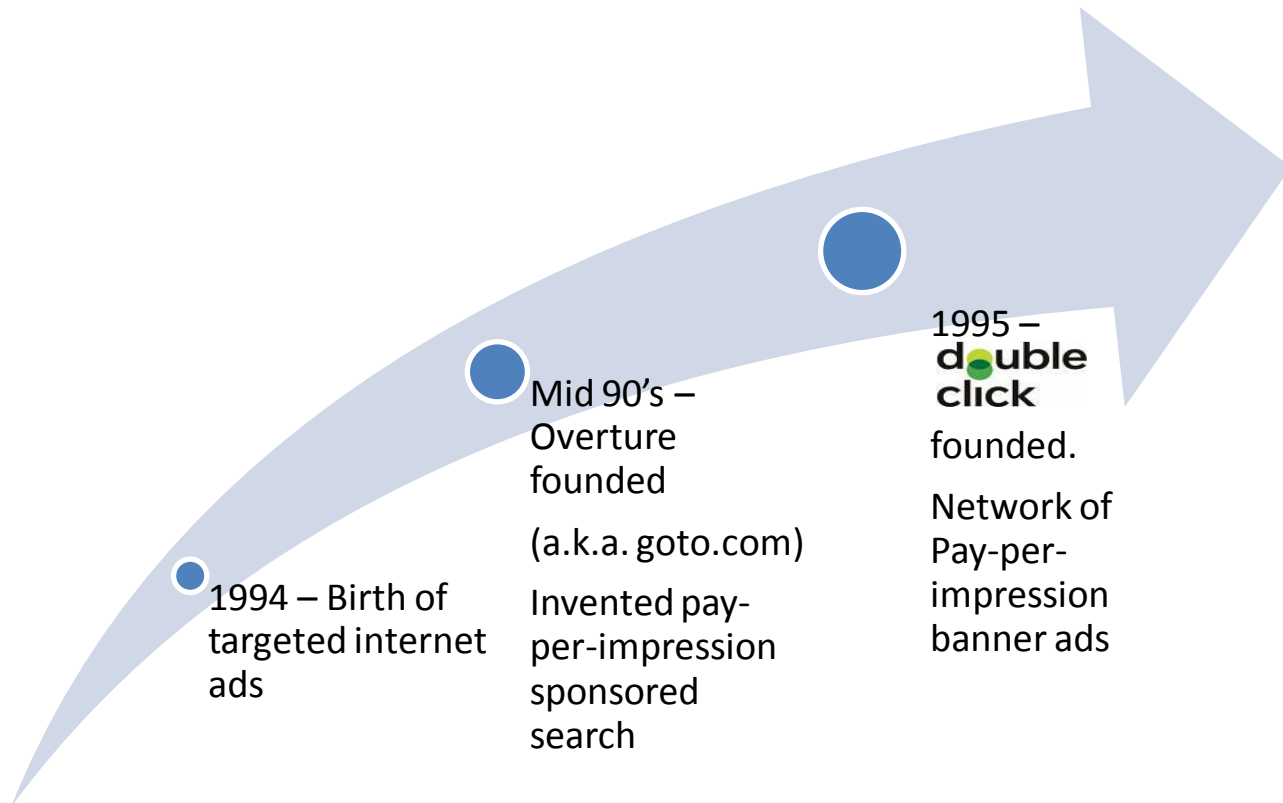
The Lane's Gifts case

- 2005 – “Lane’s Gift and Collectibles” filled a law suit against Google on behalf of all Google advertisers.
 - tired of paying for invalid clicks.
- Mid. 2006:
 - Case settled: Google agrees to refund \$90 million
 - Opened for advertisers to apply for reimbursement for clicks they believe are invalid
- Mid. 2006 - Alexander Tuzhilin was asked to evaluate Google invalid click detection efforts

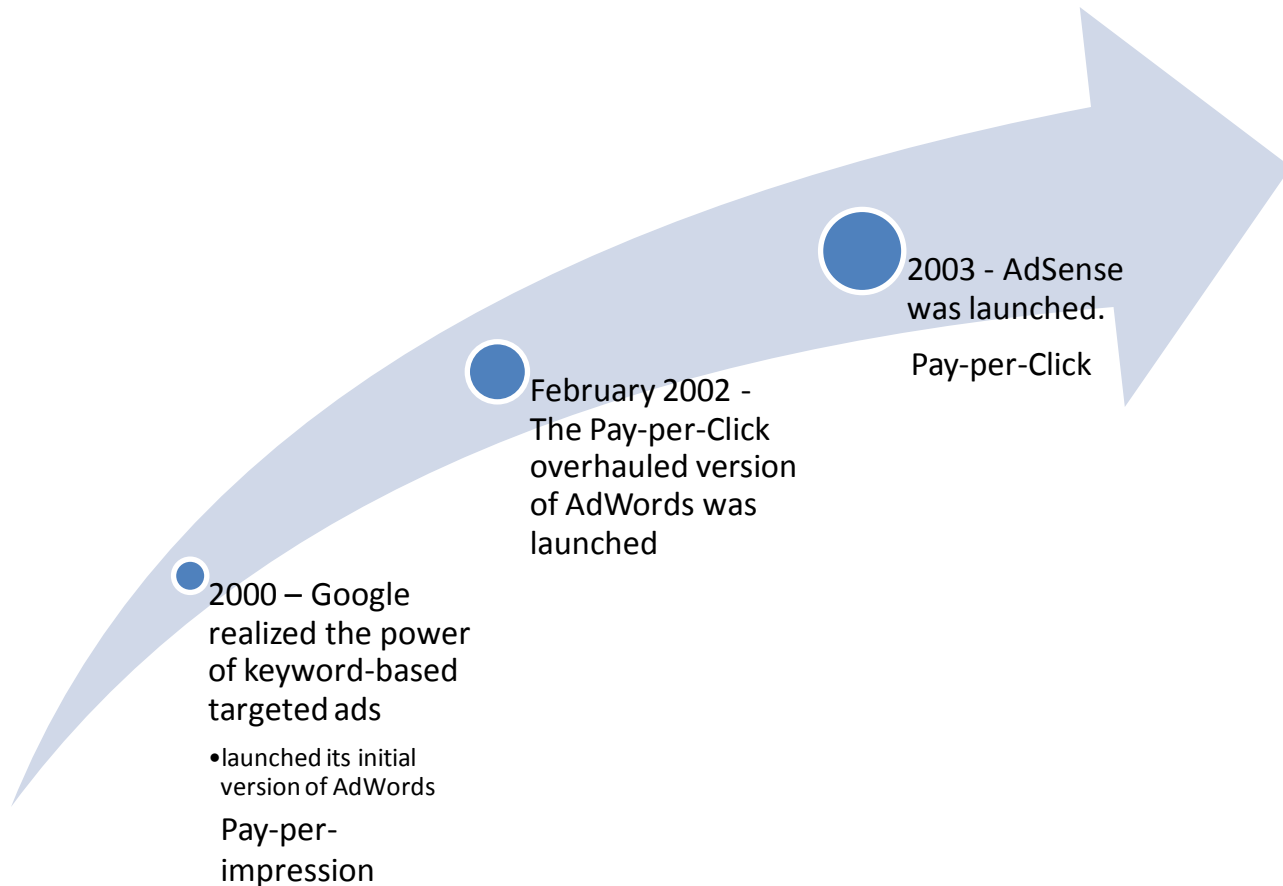
Outline

- **Background information**
- Invalid Click – Hard to define
- Google's Approach
- Conclusion

Background



Background – Google’s initiative



AdWord vs. AdSense

	AdWord	AdSense
Where	www.google.com	www.publishersSite.com
What	Query based	Content based
Who makes money	Google	Google + publisher
Who gains due to click fraud (short-term)	Google + targeted advertiser's competitors	Google + publisher + advertiser's competitors
Who loses due to click fraud (short-term)	Targeted Advertiser	Targeted Advertiser
Who loses due to click fraud (long-term)	Targeted Advertiser + Google	Targeted Advertiser + Google

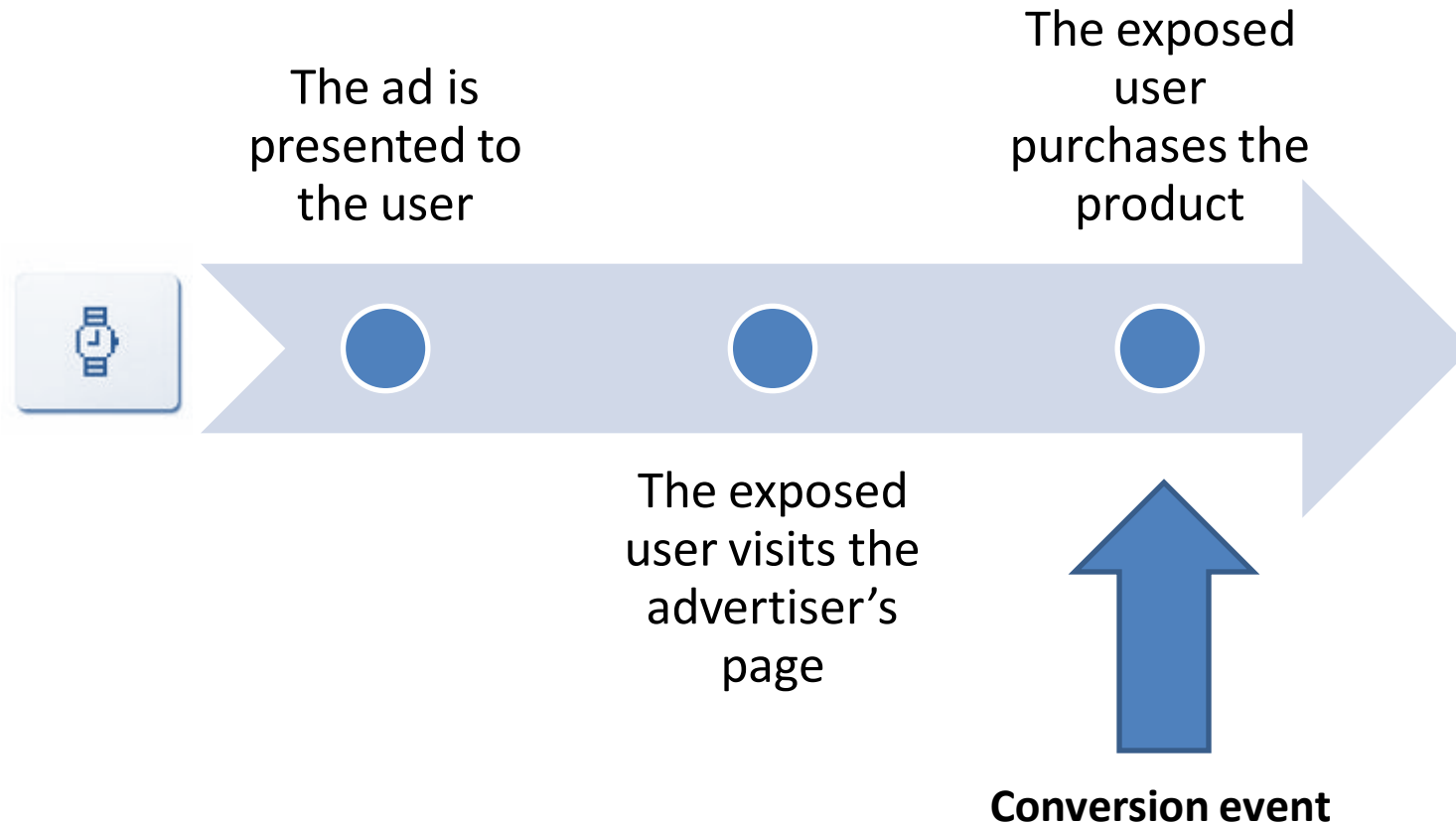
Time

When charge the advertiser?

- When the ad is being shown to the user
 - *CPM – Cost per Mille*
- When the ad is being clicked by the user
 - *CPC – Cost per Click*
- When the ad has “influenced” the user (conversion event)
 - *CPA – Cost per Action*



Cost-per-Action



Two effectiveness measures

- Click-Through Rate (CTR)

- $CTR = \frac{\#ads_clicked}{\#ads_presented}$

- Conversion Rate

- The % of visitors who took the conversion action

Cost-per-click Advertising Model

- Ad Rank – How high the ad is placed on www.google.com (example on next slide)
- Cost-per-Click (CPC)
- Quality Score – quality of the keyword/ad pair
 - Depends on the Click-through-rate (CTR)

$$\text{Ad Rank} = f(\text{CPC}, \text{QualityScore})$$

Pay-per-Click AdWord model

The screenshot shows a Google search for "buy stocks" with the following elements:

- Search Bar:** "buy stocks" entered, search button, and "Advanced Search" link.
- Results Summary:** "Results 1 - 10 of about 2,450,000 for buy stocks. (0.11 seconds)"
- Organic Results (Left):**
 - Buy Stocks** - www.ETRADE.com - 100 Commission-Free Stock Trades. Apply Now at E*TRADE Securities.
 - Buy Stocks for \$4** - www.ShareBuilder.com - No Account or Investment Minimums. No Inactivity Fees. Start Today!
 - Take Control of Trading** - www.Scottrade.com - Powerful trading tools, \$7 online trades, fast executions. Apply now.
 - 10 stocks to buy now** - money.cnn.com/popups/2006/fortune/invguide_stocks/index.html - 36k -
 - How to buy stocks** - CNNMoney guide to investing in mutual funds. Everything you need to know about investing.
 - How To Buy Stocks | eHow.com** - How to Buy Stocks. Buying stock in a company is relatively easy once you've researched the stocks you're interested in and have a broker or brokerage ...
 - How to Buy Stocks Online » Financial Investing - Financial Rebel** - Jul 17, 2007 ... Buying stocks online is a fairly straightforward process. Gone are the days of calling up your broker, unless you want to.
 - ING DIRECT | ShareBuilder: Buy Stocks Online and invest your money ...** - Buy stocks online and find new ways to invest your money at ShareBuilder.
 - Online Trading - TD AMERITRADE - Online Stock Trading and Investing** - Review mutual fund families - Buy CDs now - Learn about Apex rewards - Understand Asset Protection - Plan with Amerinvest ...
 - How to Buy Stocks - wikiHow** - wikiHow article about How to Buy Stocks. ... Before buying stocks, make sure you have a decent idea of how to choose which stocks to buy ...
 - How to buy ... stocks - MarketWatch** - SAN FRANCISCO (MarketWatch) - Stock tips spill from everywhere: on television, at parties, in the gym. E-mail boxes are full of pitches for can't-miss ...
 - How to Buy Stocks - Mahalo** - Buying stock is easier and quicker than it's ever been before, but certainly no less risky. If you're a first time investor, you'll want to prepare yourself ...
 - Learn how to Invest, Buy Stocks, Sell Stocks, Investing and ...**
- Sponsored Links (Right):**
 - Buy Stocks** - Straightforward Pricing. Powerful Tools. Sign Up Now at TD AMERITRADE TDAMERITRADE.com
 - \$0 Stock Trades at Zecco** - Not An Introductory Offer. No Gimmicks. No Tricks. www.Zecco.com
 - Stock Trading** - \$8.95 Online Equity Trades. Powerful Tools and No Service Fees. www.Schwab.com
 - Stock Trading Online** - \$4.95 per Trade, Market or Limit Tools, charts, newsreader and blogs www.TradeKing.com
 - Buy Stocks** - Trade stocks with optionsXpress. Free tools to help with your trade optionsXpress.com
 - Top Stocks Of 2008** - Exciting Oil Find, Read Now to Invest. Find Out Top Stocks Here! TopStockAlert.com
 - Online Stock Trading** - Trade Our Capital. Utilize Our Resources & Experience. www.keystonetradinggroup.com San Diego, CA
 - Online Broker** - Discover a better way to invest. TWS Financial, Member FINRA & SIPC www.TradeWallStreet.com

AdWord – Ranked after the Ad Rank

Problems with CPC

- Good click-through rates (CTRs) are not indicative of good conversion rates
- No “built-in” fundamental protection (endogenous) mechanisms against click fraud

Invalid click

From Wikipedia:

“Click fraud occurs in pay per click online advertising when a person, automated script or computer program imitates a legitimate user of a web browser clicking on an ad, for the purpose of generating an improper charge per click.”

Example of Click Frauds

- Firm A has an ad budget of 100\$/day
- Firm B depletes this budget with fake clicking.
 - > No more ads for Firm A that day
- Firm A publishes an ad at www.firmB.biz
- Firm B clicks on the ad several time without any plans of buying anything
 - Firm A has to pay for fruitless clicks and Firm B gets paid for invalid clicks.

Different kind of problems with the Cost-per-Click Model

- Unethical advertisers of AdWords will try to use up budgets of other advertisers
- Unethical publishers of AdSense will try to enrich themselves
- Google launched a beta CPA model March 2007 to handle these problems.

Outline

- Background information
- **Invalid Click – Hard to define**
- Google's Approach
- Conclusion

Invalid click – Hard to define

- Consider the case of a double-click, i.e., two clicks on the same ad impression by the same browser, where the second click follows the first one within time period p
 - *What is the threshold p which splits the clicks into valid and invalid? 10 sec ? 1 sec?*
- Consider clicks on different ads by same viewer leading to the same page.

Recognizing Invalid Clicks (1)

Anomaly-based

- *i.e.* a normal average clicking frequency on an ad is <1 clicks/week per user. If someone clicks on it 100 times/week \Rightarrow abnormally large clicking activity

Challenges:

- Identify groups of clicks from “same user”, “same ad”, etc.
- identify what the “normal” clicking activities
- Define what “deviation from the norm” is

Recognizing Invalid Clicks (2)

- *Rule-based*
 - set of rules identifying invalid or invalid clicking activities
 - *i.e.* “IF Double-click occurred THEN the second click is Invalid”
- *Challenges:*
 - *Are the conditions reasonable?*
 - *i.e. duplicate click was in the start treated by Google as a valid click => the customers had to pay for it.*
 - *Are the conditions consistent (to the definition of invalid click)?*

Recognizing Invalid Clicks (3)

- *Classifier-based*
 - Build a statistical model based on the past data that can classify new clicks into valid or invalid
 - Assign probability to the classification
- Challenges:
 - Need to manually label a large training set, which might be an issue in itself.
 - Does the classifier manage to capture the conceptual description of an invalid click?
 - Concept drift and adversarial classification

Operational Definitions of Invalid Clicks

- Google uses:

- Mainly rule-based and anomaly-based approaches.

No machine learning

Unsupervised learning

- For some minor cases the classifier approach

Supervised learning

Fundamental problem of the Cost-per-click Model

Publish the rules?

Yes – unethical users will take advantage of the information (adversarial problem).

No – no overview over what the advertisers exactly is charged for.

Outline

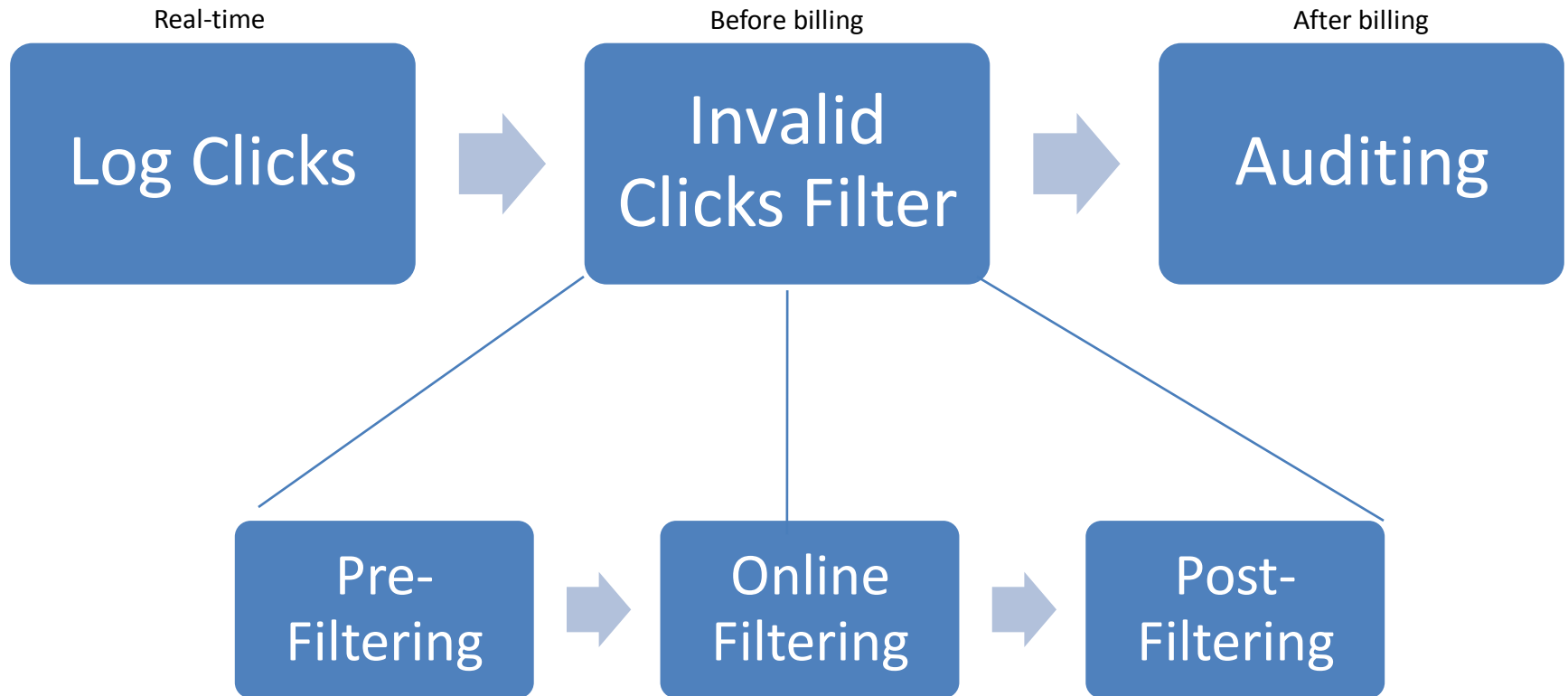
- Background information
- Invalid Click – Hard to define
- **Google's Approach**
- Conclusion

Google's Approach

The Click Quality team's mission statement:

- Protect Google's advertising network (long-term profit) and provide excellent customer service to advertisers. We do that by:
 - monitoring invalid clicks/impressions and removing its source
 - Reviewing all client requests and responding in a timely manner
 - Developing and improving systems that remove invalid clicks/impressions and properly credit clients for invalid traffic
 - Educating advertisers and employees on invalid clicks/impressions.

Google's Process



- Automated monitoring
- Manual Reviews
 - Proactively
 - Reactively

Overview: Google's Approach

- Prevention

- Discouraging invalid clicking

- Hard to make duplicate accounts

- Hard to make fake accounts

- Don't pay for fraudulent activities

} -Building walls

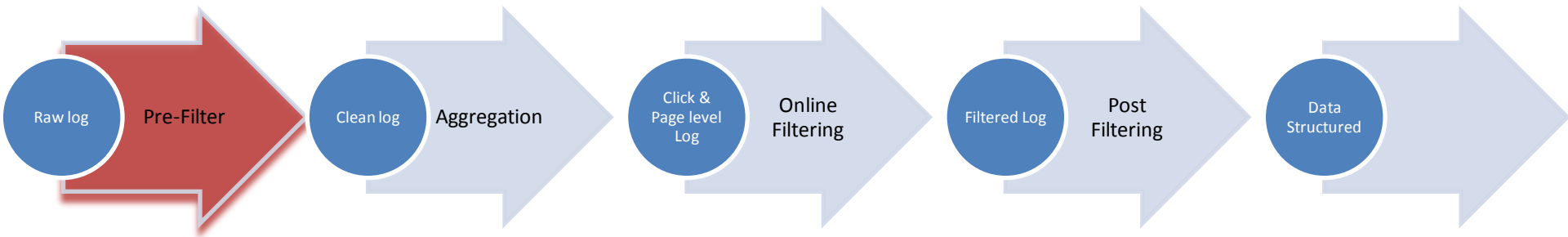
} -Very limited punishment

- Detection

- Detecting and removing invalid click

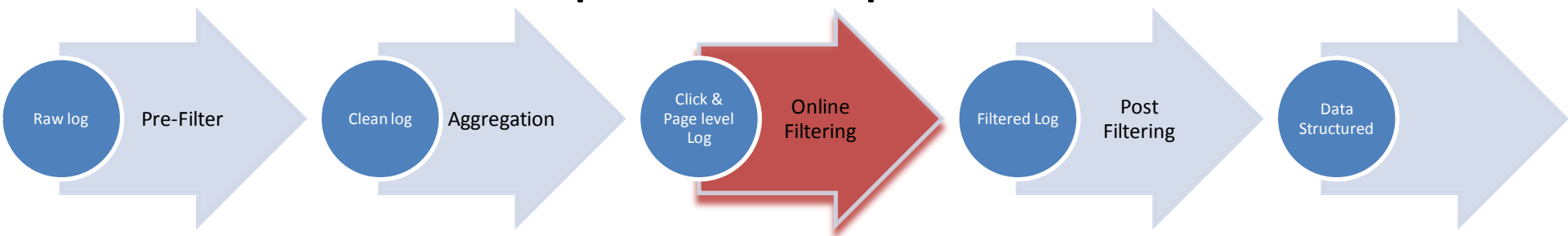
Pre-Filtering

- Clicks removed from log in order to keep the performance statistics clean
 - Google test clicks removed
 - From Google's IPs
 - Meaningless clicks removed
 - Improperly recorded clicks



Online Filtering

- Rule-based filters and anomaly-based filters
- Detection within a short time window
- Clicks are identified and marked as invalid and advertisers are not charged for them
- The invalid clicks are removed at the end of the filtering process => the filter sees all the clicks; can compare multiple related clicks



Performance of the Online Filters

- The typical way of presenting performance of a classifier is with a Confusion Matrix
- Unfortunately, Google does not know which clicks are actually valid
 - > Have to measure performance through indirect evidence

Performance of the Online Filters contd.

- The indirect evidence:
 - If newly added filters only suggest a few additional invalid clicks
 - The offline filters suggest only a few additional invalid clicks

Performance of the Online Filters

- From indirect evidence, Online Filters seem to be effective.
- This surprised the author; are the filters too simple?
- Answer:
 - Reasonable performance due to:
 - Combination of filters
 - Simplicity of most attacks
 - Some complex filters (although most of them are simple)

The long tail of invalid clicks

- Massive amount of invalid clicks from only a few types of inappropriate activities
- Long tail: many infrequent idiosyncratic activities
- Google's filters easily catches the left part
- Question 1: Why do criminals continue?
- Question 2: Which activities are shrinking/expanding?

The long tail of invalid clicks

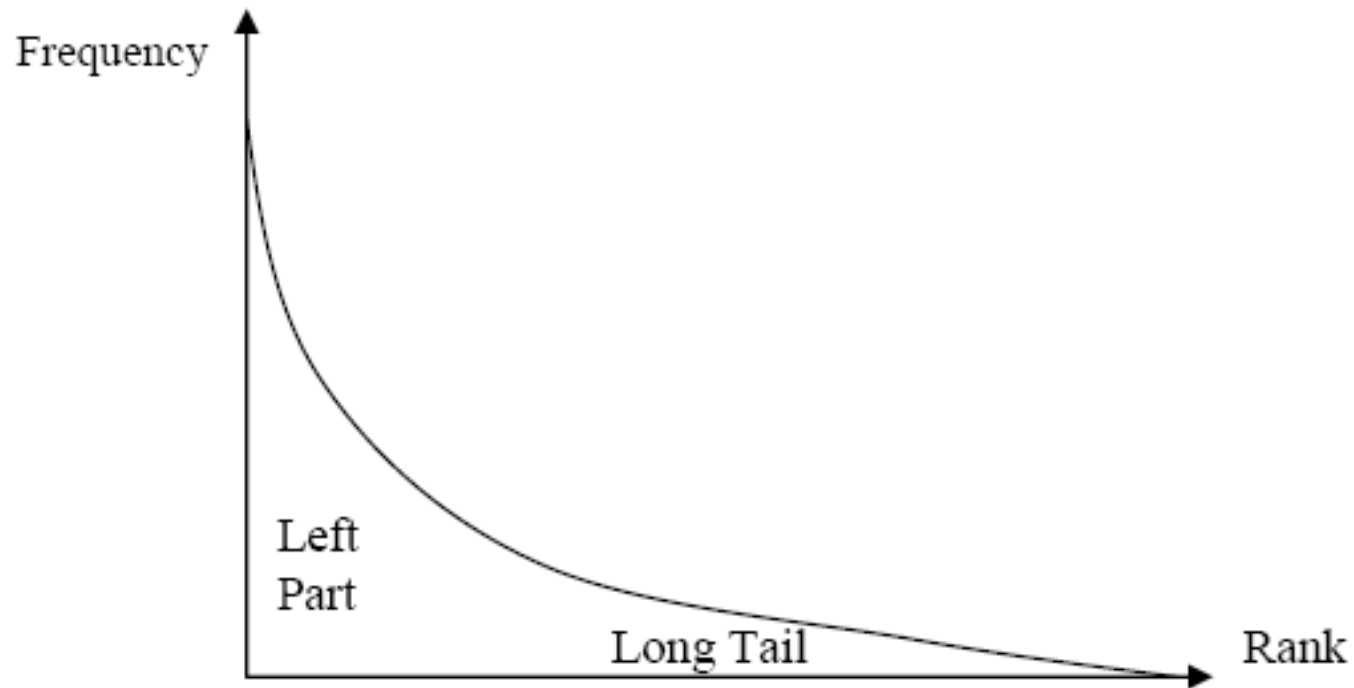


Figure 1: The Zipf's Distribution and the Long Tail of Invalid Clicks.

Online Filters: The future threats

- Today: Filters perform well and seem to be accurate
- Future: New attacks might pretend to shift towards the long tail
 - E.g. botnets

Are the filters biased?

- Is it profitable for Google to filter laxly
 - > let through some invalid clicks
- I.e. consider filter:
 - *“If signal X associated with a click is above the threshold level a then mark the click as invalid”*
- Low a => lost short-term profit for Google
- High a => gain in short-term profit for Google

Are the Online filters biased? Contd.

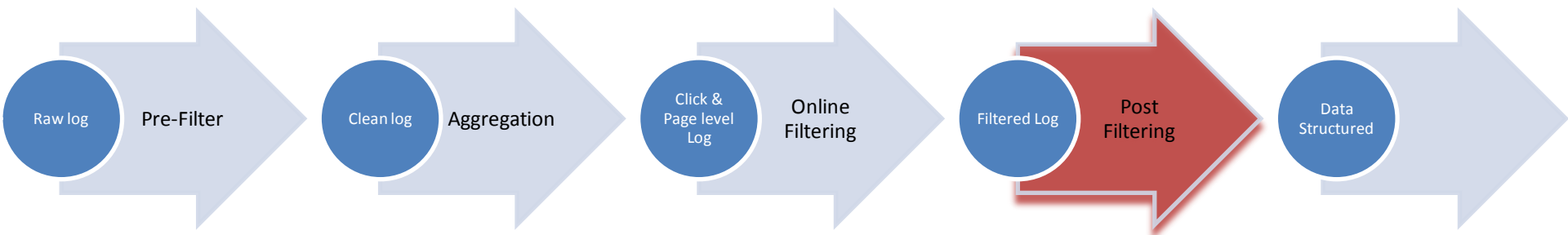
- Is it worth gaining some revenue but losing the advertisers trust?
- Author's investigation:
 - Click classification is purely a engineering decision, with no input from the finance department
 - Except the case regarding duplicate click
 - This case was a big concern for the management
 - Still, despite the short-term, they decided to change the policy.

What is missing in Google Filters?

- More supervised learning approaches
- Using the conversion data in filters
 - Hard to collect conversion data
 - Very sparse data
 - Conversion can take some time after the actual click => hard to use in online filters.

Post/offline detection methods

- No real-time constraint, no computational constraint.
- Automated
 - Alert systems
 - Automated termination system for AdSense publishers
- Manual
 - Handle complaints
 - Handle alerts



Performance of detection methods

- Indirect evidence:
 - Newly added and revised filters detects few additional invalid clicks
 - Same for the offline methods
 - Increase in # of clicks marked invalid, but not in complaints from AdSense publishers.
- No hard evidence :/
 - > but can conclude that the filters works reasonably well

Conclusion

- The conceptual definition of invalid click assume human intent:
 - No method satisfying this definition for algorithmically detecting invalid clicks
 - > Need operational definitions:
 - Anomaly based
 - Rules based
 - Classifier based
- No complete data on actual valid/invalid clicks but
 - Complaints from advertiser indicate invalid
 - Complaints from publishers indicate valid