

Конфиденциальность пользователя как вызов современному интернету

29.09.2020



ДИСКУС-КЛУБ

Выборнов Артём

- 7 лет в Big Data
- Отвечаю за разработку Big Data продуктов в компании Rambler Group
- Соавтор курса Big Data for Data Engineers
- Читаю лекции о Big Data бизнесу

RAMBLER
GROUP



Слайд о компании

МЕДИАХОЛДИНГ

Виды идентификации

User

passport
email
phone
auth.id

Device

IDFA
GAID
MAC

Browser

third-party
cookie

Domain

first-party
cookie

Часть из них перестает работать

User

passport
email
phone
auth.id

Device

IDFA
GAID
MAC

Browser

third-party
cookie

Domain

first-party
cookie

Завязка

- Текущие идентификаторы браузеров и устройств позволяют неявно собирать информацию о пользователе третьим лицам
- Весь современный интернет монетизируется основываясь на возможности неявного и свободного сбора информации о пользователе
- Если подобная возможность пропадёт - большая часть текущих бизнес моделей не будет работать

О чём мы сегодня поговорим

- Конфиденциальность пользователей это неизбежное направление развития интернета
- Действия ключевых игроков рынка
- Влияние в долгосрочной перспективе
- Очертания мира без “мягких” идентификаторов пользователя



ДИСКУС-КЛУБ

Конфиденциальность
пользователя это
неизбежный тренд

Если у вас паранойя, то это не значит что за вами не следят

- Внешние вызовы на главной странице одного популярного развлекательного портала
- Google, Yandex, Criteo, TNS, VK
- + из под некого iframe: BidSwitch, Taboola, Rambler, Pubmatic, Yahoo, Mail.ru

Name
☐ pixel.gif?ch=2
☐ pixel.gif?ch=1
▶ ☐ www.google.com
▶ ☐ cas.criteo.com
▶ ☐ widget.nl.eu.criteo.com
▶ ☐ static.eu.criteo.net
☐  recaptcha__ru.js
▶ ☐ www.google.com
▼ ☐ <iframe>
☐ pixel?google_nid=cjp&google_sc&google_ula=9130...
☐ k-AGMEStZquXXe1HjFSRyp70ldlftPC0P2mb_L9A
☐ seg?add=130915&redir=https%3A%2F%2Fsecure....
☐ cm.gif?p=84&id=k-DZijfdZquXXe1HjFSRyp70ldlftO...
☐ sync?dsp_id=46&user_id=k-EnO_-NZquXXe1HjFS...
☐ rtb
☐ sync2.204?pid=186&anket_id=k-ef5RltZquXXe1HjF...
☐ Pug?vcode=bz0yJnR5cGU9MSZjb2RlPTE5Mjgmd...
☐ tap.php?v=6434&nid=2149&put=k-_vbhiNZquXXe1...
☐ sd?id=537072953&val=k-Ku35sNZquXXe1HjFSRyp...

Пользователи не читают соглашения

Эксперимент Gamestation

By placing an order via this web site on the first day of the fourth month of the year 2010 Anno Domini, you agree to grant Us a non transferable option to claim, for now and for ever more, your immortal soul. Should We wish to exercise this option, you agree to surrender your immortal soul, and any claim you may have on it, within 5 (five) working days of receiving written notification from gamesation.co.uk or one of its duly authorised minions. We reserve the right to serve such notice in 6 (six) foot high letters of fire, however we can accept no liability for any loss or damage caused by such an act. If you a) do not believe you have an immortal soul, b) have already given it to another party, or c) do not wish to grant Us such a license, please click the link below to nullify this sub-clause and proceed with your transaction.

[Click here to nulify your soul transfer.](#)

Данные о пользователе ему не принадлежат

- Сегодня пользователь не способен контролировать использование информации о себе
 - Cookies/IDFA/GAID позволяют легко и неявно отслеживать каждый шаг пользователя любому участнику рынка вызов пикселя которого размещён на странице / трекер которого размещён в приложении
 - Эти же идентификаторы позволяют разным компаниям легко обмениваться данными о пользователях
 - Пользовательские соглашения составлены так, что невозможно пользоваться сервисом и при этом не разрешить передачу информации о себе третьим лицам

И это проблема

- Это беспокоит самих пользователей
- Об этом беспокоятся государства (ФЗ о “Защите персональных данных”, GDPR, etc)

И это проблема

- Это беспокоит самих пользователей
- Об этом беспокоятся государства (ФЗ о “Защите персональных данных”, GDPR, etc)
- Об этом говорят владельцы крупнейших экосистем (Apple, Google)

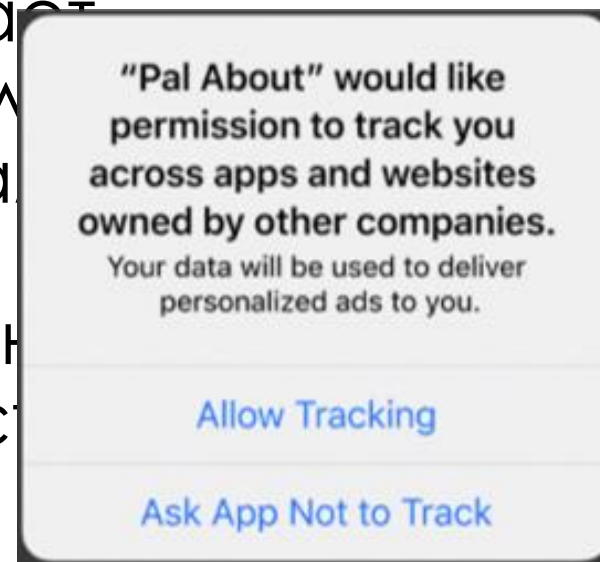


ДИСКУС-КЛУБ

Действия ключевых игроков рынка

Идентификация в мобильных приложениях

- Базируется на системных идентификаторах: IDFA (iOS) / GAID (Android)
- С iOS 14 (осень 2020 года) Apple начинает явно спрашивать пользовательское согласие на сбор его данных внешними трекерами
- Apple предлагает своё решение (SKAdNetwork) для атрибуции трафика с точностью до пользователя, а с точностью до определённых групп пользователей



MAC адрес

- MAC адрес - уникальный идентификатор сетевого устройства подключенного к интернет
- В контексте сбора аудитории обычно идёт речь про MAC адрес модуля WiFi
- MAC адрес который устройство сообщает при поиске WiFi рандомизируется очень давно (препятствие для WiFi sniffing)
 - IOS 8 - 2014 год
 - Android 8 - 2017 год
- С этой осени Apple начинает периодически менять MAC адрес для всех устройств даже после успешного подключения к известной сети

Cookies

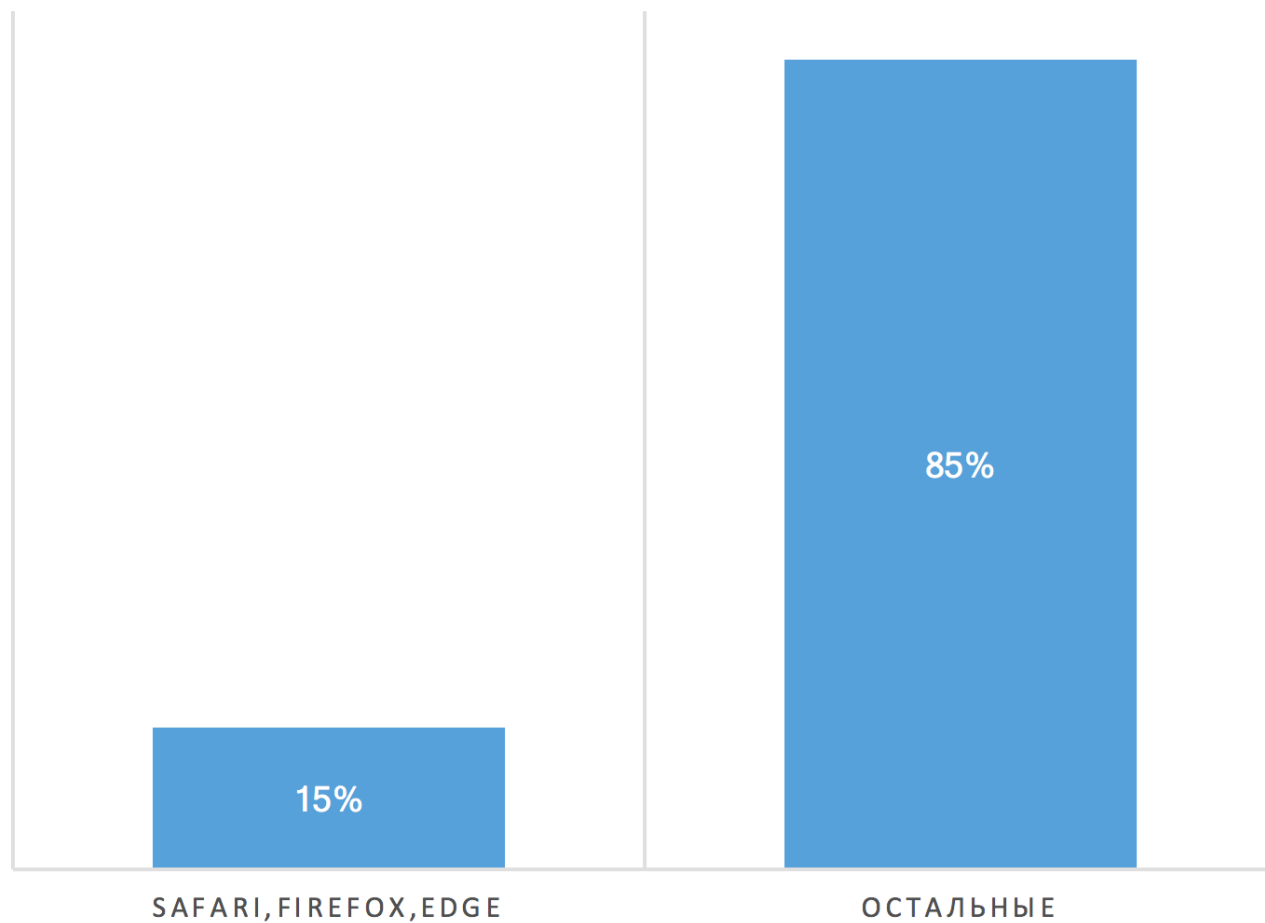
- Cookies - классический способ идентификации пользователей интернета
- First-party cookies
 - Куки которые web-приложение ставит в домен той площадки на который сейчас находится пользователь
- Third-party cookies
 - Куки которые web-приложение ставит в домены отличные от того на котором сейчас находится пользователь

Cookies

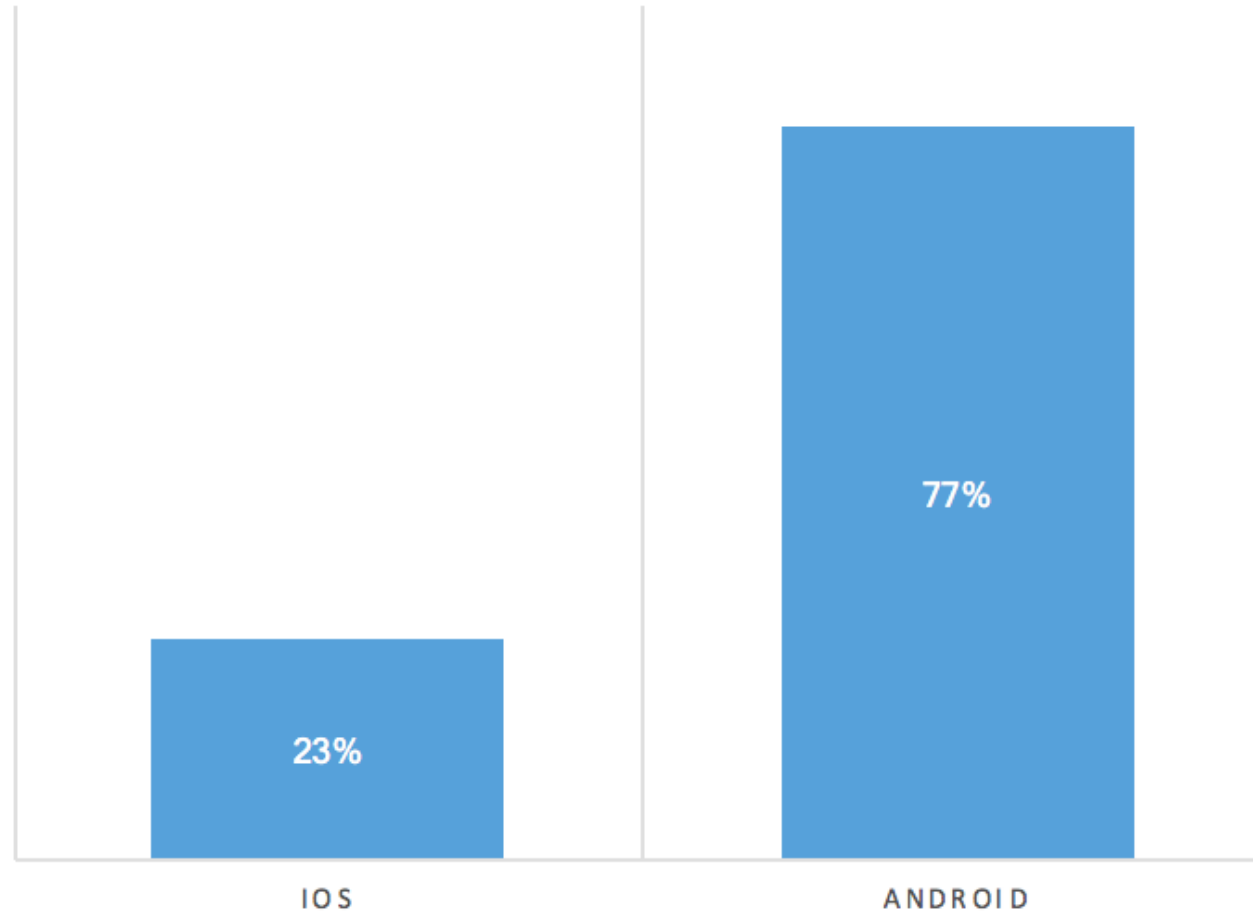
- Safari, Firefox, Edge в течение последнего года
 - Запретили установку third-party кук
 - Сильно ограничили время жизни кук first-party выставленных client-side (то есть из JS). То есть выставить по настоящему долго живущую куку может только сама площадка server-side
- Chrome
 - Объявил, что поддерживает тренд по блокировке third-party кук, но на данный момент всё ещё разрешает их использовать
 - Обязал всех явно разделять first-party и third-party с помощью заголовка samesite
 - Призвал всех обсудить что делать дальше (Chromium privacy sandbox)

**DON'T
PANIC**

Рынок браузеров в России



Рынок мобильных приложений в России



Chromium Privacy Sandbox

- Рано или поздно будет отключена возможность cross-site идентификации через third-party куки, а также сильно ограничена возможность прочей cross-site идентификации (в том числе и fingerprint)
- Google призывает всех вместе подумать как жить в НОВЫХ УСЛОВИЯХ
- <https://www.chromium.org/Home/chromium-privacy/privacy-sandbox>

Chromium Privacy Sandbox

- Chromium privacy sandbox не просто про работу с куками, а в целом про работу с пользовательскими данными во всей экосистеме Google
- Google как владелец подавляющей части рынка платформ не хочет “сломать” интернет и собирается двигаться постепенно
- Пока это так (как минимум до 2022 года) - можно жить в привычной парадигме мира

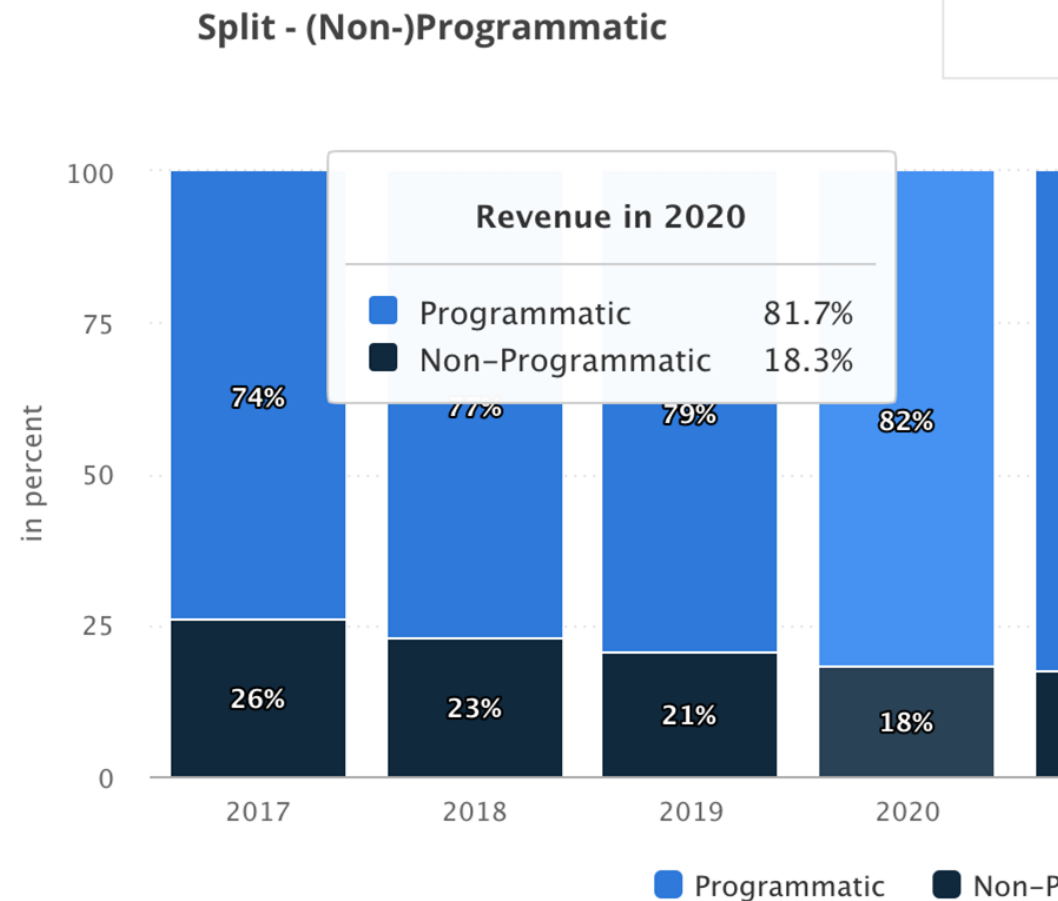


ДИСКУС-КЛУБ

Влияние в долгосрочной перспективе

Programmatic

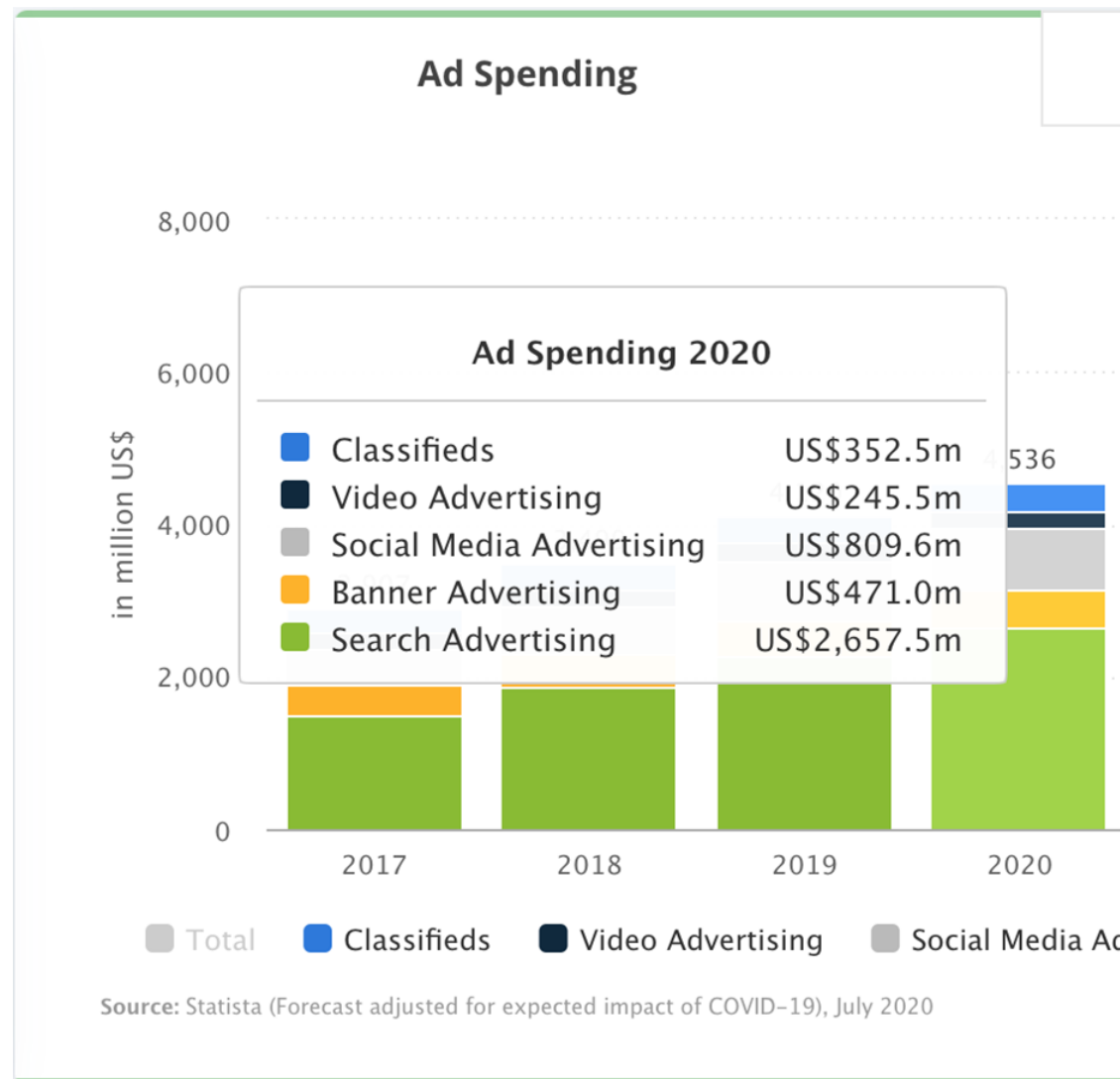
- Пропадёт возможность
 - Таргетироваться на пользователя
 - Измерения результатов рекламных кампаний с точностью до пользователя



Source: Statista (Forecast adjusted for expected impact of COVID-19), July 2020

Это затронет всех

- Инфраструктура заточена под third-party cookie и под GAID/IDFA
- Пропадёт возможность обогащения профиля пользователя на одной площадке данными с другой



Сломается не только реклама

- Кросс-доменная авторизация
- Кросс-доменная аналитика и измерения аудитории проекта
- Задачи основанные на обмене данными об интернет аудитории между различными игроками
- ...

Как изменится рекламный рынок

- С точки зрения пользователя
 - Пользователь сможет активнее влиять на то, как именно используются данные о нём
 - Реклама станет менее релевантной
- С точки зрения рекламодателей
 - Реклама с пользовательским таргетингом станет дороже
 - Будет значительно сложнее оценить эффективность рекламы

Как изменится рекламный рынок

- С точки зрения площадки / приложения
 - Для эффективной монетизации аудитории потребуются самим активнее работать с собственными данными
 - Сильно деградируют возможности монетизации для маленьких площадок / приложений
 - Возрастет значимость всех прочих идентификаторов (уровня домена / приложения и уровня всего пользователя)

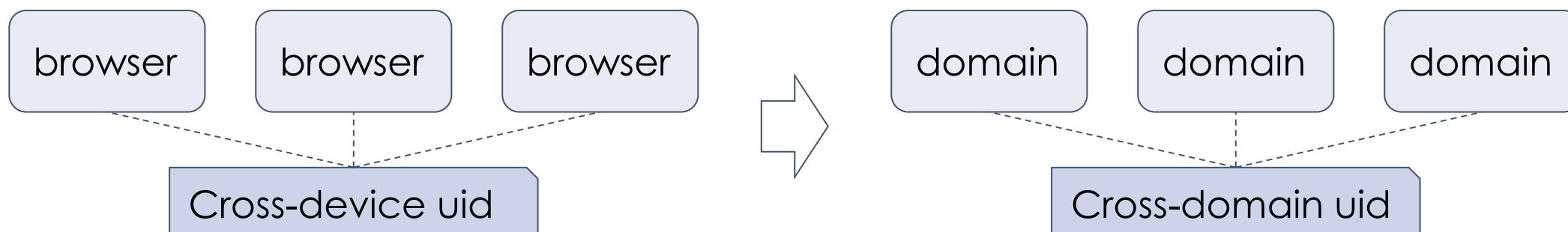


ДИСКУС-КЛУБ

Очертания мира без “мягких” идентификаторов пользователя

Гонка вооружений

- Развитие fingerprint-подходов
- Вероятностная склейка пользователей между разными доменами/приложениями



Гонка вооружений

- Развитие новых способов неявной идентификации (в том числе CV, аудио, etc)
- CV уже хорошо умеет в решение задачи классификации (кто изображен на картинке из заранее внесенных в базу) и очень близко к промышленному решению задачи кластеризации (по набору лиц сгруппировать лица всех проходящих



Пользователь => Клиент

- В случае владельцев инвентаря
 - Повышение доли пользователей за логином
 - Мотивация пользователя заходить на разные сайты под одной авторизацией (универсальные id, системы лояльности)
 - Использование “твердых” идентификаторов: email/phone
- В случае рекламодателей
 - Тщательный сбор и обработка всех пользовательских данных
 - Работа с пользователями на основе CDP-like решений

Развитие in-house решений

- Создание собственных DMP/DSP для своей платформы
- Фрагментация рынка programmatic платформ

Развитие прочих таргетингов

- Таргетирование на контекст запроса
 - Местоположение, Погода
 - Устройство, ОС, Модель
 - Сайт, Страница, Тематика
 - ...
- Таргетирование на “когорты” пользователей
 - TV-like подход - инсайды о площадке
 - Сбор аудитории в рамках крупных срезов

ИТОГ

- Конфиденциальность пользователей это неизбежное направление развития интернета
- Ключевые игроки приступают к блокировке всех идентификаторов, позволяющих неявно собирать информацию о пользователе. Но пока влияние изменений на рынок не столь существенно
- Эти изменения ломают текущие подходы к интернет рекламе (и не только). Всем игрокам придётся к ним адаптироваться
- Хорошего ответа как это сделать пока нет, но общие очертания нового будущего уже видны



ДИСКУС-КЛУБ

Ваши вопросы