

Великий КРИ с капиталистическим лицом

Меня часто спрашивают, почему это я так старательно перемешиваю и дозирую информацию о себе, идущую в интернет? Почему не позволяю тэгать свои фотографии на Фейсбуке, почему отделяю друг от друга сервисы в нашу эпоху повальной конгломерации, зачем настраиваю ручки и кнопки на повышенную приватность, да ещё и время от времени путаю следы откровенной липой?

Разумеется, я далёк от мысли, что это позволит мне сохранить анонимность в Сети. Её нет уже давно ([раз](#), [два](#), [три](#)). Да что там, ФБРы и КБГы могли собирать подробнейшие досье о людях ещё лет 50 назад без всяких ынтернетов и суперкомпьютеров, и, в общем, я не питаю каких-то иллюзий на этот счёт в своём отношении. Будет нужно -- и Кому Нужно, всё узнают.

Спам, фишинг и перепродажа информации меня, как ни странно, тоже мало волнуют. Нет, раздражают, конечно. Будь моя воля, для спамеров отвели бы специальный десятый круг ада. При жизни. За разрушение ими среды доверительного общения в сети. Но эта опасность всё-таки общепонятная, и потому меньше волнующая. Собственно, тут давно известно, как надо себя вести: не светить лишней раз адреса, не делиться с кем попало критическими данными вроде своего SSN или паролей. В этом отношении государство (любое) со своей склонностью про%ывать уже собранные данные через тупоголовых исполнителей меня потенциально напрягает куда больше, чем базы данных того же Фейсбука.

Всё не то, в общем. Напрягает же меня явление относительно новое, на глазах набирающее бешеные обороты, но которое почему-то очень многие в упор не видят. Видимо, за отсутствием очевидно-наглядных исторических аналогий в прошлом. Я говорю о безоглядном сборе и анализе рассеянных данных.

Подумаем, за счёт чего контора, обрабатывающая информацию, может предоставлять лучшие сервисы? Параметров, конечно, много, но условно их можно разделить на две группы. Первая -- это "лучше работать". Писать более умные, хитрые, быстрые алгоритмы, например. И вторая "ручка" -- использовать больше данных. Тогда даже слабенькие алгоритмы, глядишь, окажутся конкурентоспособными на фоне привнесённого богатства.

Лет пять ещё назад Гугл был очевидно "впереди планеты всей" в первой области. Его алгоритмы работы с информацией рвали всех конкурентов в мелкие треугольнички, а талантливые программисты очень недурно эти алгоритмы писали.

Но технологическое преимущество в узкой области невозможно удерживать вечно. Сам факт понимания, что "вон там мёдом намазано" привлекает тем больше мыслителей и догонятелей, чем сильнее их с этой высоты пытаются спихнуть. Хоть конкуренцией, хоть патентами, хоть скалыванием горы путём регулярного выброса устаревших алгоритмов в открытый доступ -- неважно. Всё равно повторяют, переоткроют, сделают другое похожее, обойдут сбоку, купят, украдут, наконец. В качестве исторического примера смотри попытки ограничить распространение ядерного оружия или недавние успехи в космонавтике стран, ещё лет 20 назад славных разве что анекдотами про грыжу при запусках ракет.

Гугл сегодня по-прежнему лидер, но на пятках у него сидят уже многие и локтями пихают с разных сторон. Майкрософтовский [Bing](#), пусть и послабее, но всё же поисковик вполне мирового класса. Фейсбук на практике отрабатывает data mining на живых людях в таких объёмах, о которых я те же лет 5 назад только мечтал, и профилирует социальные связи с безумной точностью. Пандора умеет подбирать музыку по вкусу слушателя (довольно бездарно, но учтите, что это вообще не человек! Лиха беда начало...) А многие алгоритмы, когда-то обретавшиеся лишь в головах узкой кучки специалистов, сегодня [висят на Википедии](#) и являются второй главой букваря программиста, работающего с серьёзными данными.

В такой ситуации что для Гугла, что для его конкурентов становится особенно важен второй рычаг: сам объём информации. Чем больше ты знаешь о ситуации вокруг задаваемого вопроса, тем точнее ты можешь на него ответить. У человека это не так: переполнение оперативной

памяти на каком-то этапе ведёт к **ухудшению** результатов. Но вычислительная революция, сделавшая доступной обработку петабайтов данных, ввела в игру принципиально новую возможность: сбор, уплотнение и использование чрезвычайно **рассеянной информации**. Работает это примерно так (из русского перевода популярной лекции в двух частях: [раз](#) и [два](#)):

Google Электроэнергия и Измерение. Да, ребята, Google будет поставлять электроэнергию. Все большему и большему числу людей, они запустили публичный сервис, и будут продавать электроэнергию. Они будут знать, сколько вы используете энергии, и что вы делаете, и кстати это обязательное условие при покупке электроэнергии у Google, вы обязаны использовать так называемое «интеллектуальное замеривание» которое означает что многие ваши устройства будут подключены к Скайнету. Они смогут определять по скачкам потребления мощности холодильника, что вы его открыли, а теперь закрыли, а теперь вернулись к телевизору, включили канал, вы пошли в спальню – вы включили там свет.

Дальше Гугл сравнит эти данные с просмотренным Вами (по Google TV!) кино, его название обнаружит в переписке по gmail с вашей девушкой, узнает, когда Вы её приглашаете на кинопросмотр и подскажет, что прямо по пути от неё уже есть удобный магазинчик, где можно купить Ваше любимое пиво, упомянутое в разговоре с другом по СМС год назад.

Всем привет от Великого КРИ, уж он нам похихикает...

Информация важна **вся**: где Вы ходите, в какие часы сидите за компьютером, с кем переписываетесь, когда заправляетесь, какие шутки любите, сколько говорите по телефону и штаны какого цвета предпочитаете. И она хранится вечно. Обработав эту хренову прорву данных, умная система **действительно** может заметить в ней нетривиальные связи и выдать Вам [что-то такое](#), что удивит Вас самих неимоверно. Причём полезное. Говоря о тех же кинофильмах, вы, может, ещё даже подумать не успели о просмотре, а система Ваши мысли уже предвосхитила и рекламку кажет, неназойливо так, в тему, и прямо в десятку, ну грех не купить.

Впрочем, "неотразимая" реклама ещё не самая большая проблема. Реклама вообще не может иметь более чем ~90% эффективность. Иначе общество само себя ею загипнотизирует. Всегда должно быть сколько-то процентов людей, так или иначе ей успешно противостоящих, ибо кому-то же надо **объективно** сравнивать товары и услуги в целях дальнейшей успешной конкуренции...

Итак, сбор и хранение информации. Даже это само по себе ещё не совсем проблема. "Приватность" (privacy) как область для законодательного регулирования была выделена людьми уже давно. Все понимают, что существуют данные "особого значения", вроде номера паспорта, записей о болезнях, электронного адреса или кредитной истории. И что если какая-то контора берёт у Вас подобную информацию для предоставления запрошенной услуги, то она, скорее всего, будет её более-менее беречь в рамках существующих законов, предписаний, и общего здравого смысла.

Другое дело информация рассеянная. Тут не то что регулирования нет, фиг бы с ним! Тут у 99% населения не выработалось пока даже туманнейшего понимания о её ценности. "А чё такого, ну кого волнует, что я им номер свой дал? Да не будут они тебе звонить, расслабься!" Заметим, речь всё-таки о телефоне. Развивать на этом фоне идею о сокрытии цвета штанов значит нарываться на намёки о паранойе. Я проверял :)

Так в чём же всё-таки опасность? Она **в сочетании** трёх факторов:

1. Оба все яйца сложены в одну корзину.
2. Яиц почему-то гораздо больше, чем два.
3. Корзины -- в руках у игроков, сломя голову бегущих стометровку.

Полное отсутствие каких-то норм в этой новой области ведёт к экспоненциальному забегу. Под давлением конкуренции игроки -- Гугл, Фейсбук, Майкрософт, Яху, и сотни поменьше --

наперегонки собирают сейчас как можно больше рассеянной информации о своих клиентах, и используют её абсолютно на своё усмотрение. В судорожном заглатывании информации обо всём и обо всех создаются и складываются невероятные массивы данных, необходимость которых для проведения заявленного бизнеса зачастую выглядит весьма сомнительно.

Вот, например, ютуб. Во время регистрации он просит ввести номер моего телефона. Зачем? Ну, подведить, что я живой человек и не спамер. Не вижу логики, но допустим, ладно. А в конце приписочка: "и ещё этот номерок Вам очень поможет восстановить доступ к аккаунту, если Вы забудете пароль". Ага! Это что же, мой номерок, оказывается, будет храниться у них там все 12 лет, пока я не забуду пароль? А не только один день для регистрации? И какой-нибудь хитрый индексирующий сервис будет об него прохаживаться и сопоставлять его с другими информационными полями, в которых он может засветиться, устанавливая всё новые связи о моей жизни?

Не было ещё в истории человечества, никогда и нигде, чтобы нарушение принципа минимальных привилегий не завершилось злоупотреблениями.

Когда правительству дано больше полномочий, чем ему реально необходимо для выполнения заявленных функций -- оно их загребает на самоукрепление и сохранение status quo. Когда возникают "лишние" деньги -- это ведёт к разворовыванию и распуханию бюджетов. Избыток власти у организаций, призванных охранять права граждан, даёт хорошо известные из истории последствия. А излишняя информация, большая, чем просто необходимая для бизнеса -- к злоупотреблению ею.

В такой ситуации не за горами (то есть, логарифмически близко) время, когда кто-нибудь из крупных игроков по-крупному, крепко в спешке лопухнётся и навредит огромному количеству людей. Как? Простейший воображаемый способ -- это утечка данных, но не сырых, а уже обработанных. Положим, обнаружится связь между паттернами покупок в магазинах и сильной предрасположенностью к курению. Даже у тех, кто ни разу не пробовал. И это попадет в плохие руки.

Или пусть без утечек. Допустим, обработав чьи-то рассеянные данные, можно установить у человека наличие obsessive-compulsive disorder, выражающегося в неконтролируемом желании покупать определённые товары и услуги. Угадайте с трёх раз, как эти результаты будут использованы? Подсказка: вполне возможно, что никто из людей в Гугле этого и не заметит. Знать будет только робот, распределённая цифровая система. А люди отметят лишь рост продаж и порадуются.

Или совсем простое. Фиг с ним, пусть утечёт только рассеянная информация. Но мне всё равно как-то не нравится, что для получения доступа к номеру телефона, адресу, марке машины, переписке, СМСкам, истории поисков в интернете за прошедшие 10 лет, списку установленных на телефоне приложений, к именам всех друзей и любовниц, предпочтениям в кино, книгах, музыке, фотомоделях, хобби, фотографиям, местам пребывания за прошедшие годы по GPS-у и т.д. хакеру теперь надо взломать только **один** аккаунт, а не двенадцать разных.

А теперь умножаем масштаб события на полмиллиарда человек.

Ха-ха три раза.

Во только когда народ в таком масштабе треснется мордой об стену, тогда, наконец, люди начнут хотя бы местами призадумываться, что тому же Гуглу, например, **совсем необязательно** иметь доступ к моему календарю и переписке, чтобы продавать программы для телефона, что Пандоре, по уму, в 99% случаев нафиг не нужны мой ЖЖ и Фейсбук, и что попытки хотмэйла организовать один "паспорт" для всех сервисов -- несусветная глупость.

Но это будет потом. Сначала должна случиться плюха планетарного масштаба, какой мы ещё не видели. Недавние утечки wikileaks на этом фоне покажутся репетицией детсадовского утренника.

Я не тешу себя надеждой не быть зацепленным в этом бардаке. Но я надеюсь, что на фоне миллиардов, сделавших все свои данные **сверхдоступными** для обработки те, кто попробует на этом нажиться, будут ориентироваться на лёгкие, доступные цели "массовых"

пользователей. Что им, в отличие от ФБР и ФСБ, будет некогда и лень копаться в более запутанных и фрагментированных случаях вроде моего при неограниченной доступности лёгких.

Разумеется, в этом нет с моей стороны злорадства или попытки выйти сухим из воды. Скорее, очень уж не хочется в неё лезть. Не хочется играть в игру, окончание которой уже столь явно вырисовывается в недалёком будущем.

Я -- за выбор. За то, чтобы каждый мог выбирать, какой уровень сервиса ему получать в зависимости от готовности делиться информацией. Сегодня такого выбора, по сути, нет.

07.03.2011

===

Text Author(s): Eugene Bobukh === Web is volatile. Files are permanent. **Get a copy:** [[PDF](#)] [[Zipped HTML](#)] === **Full list of texts:** <http://tung-sten.no-ip.com/Shelf/All.htm>] === **All texts as a Zip archive:** <http://tung-sten.no-ip.com/Shelf/All.zip>] [mirror: <https://1drv.ms/u/s!AhyC4Qz62r5BhO9Xopn1yxWMSxtaOQ?e=b1KSiI>] === **Contact the author:** h o t m a i l (switch name and domain) e u g e n e b o (dot) c o m === **Support the author:** 1. **PayPal** to the address above; 2. **BTC:** 1DAptzi8J5qCaM45DueYXmAuiyGPG3pLbT; 3. **ETH:** 0xbDf6F8969674D05cb46ec75397a4F3B8581d8491; 4. **LTC:** LKtdnrau7Eb8wbRERasvJst6qGvTDPbHcN; 5. **XRP:** ranvPv13zqmUsQPgazwKkWCeEaYecjYxN7z === **Visit other outlets:** Telegram channel <http://t.me/eugeneboList>, my site www.bobukh.com, Habr <https://habr.com/ru/users/eugenebo/posts/>, Medium <https://eugenebo.medium.com/>, Wordpress <http://eugenebo.wordpress.com/>, LinkedIn <https://www.linkedin.com/in/eugenebo>, ЖЖ <https://eugenebo.livejournal.com>, Facebook <https://www.facebook.com/EugeneBo>, SteemIt <https://steemit.com/@eugenebo>, MSDN Blog https://docs.microsoft.com/en-us/archive/blogs/eugene_bobukh/ === **License:** Creative Commons BY-NC (no commercial use, retain this footer and attribute the author; otherwise, use as you want);
=== **RSA Public Key Token:** 33eda1770f509534. === **Contact info** relevant as of 7/15/2022.

===