دانشگاه تهران
پردیس دانشکده های فنی
دانشکده مهندسی برق و
کامپیوتر

Federated Learning & Privacy Preserving ML

نام دانشجویان:
حمید سالمی، سهند خوشدل، مهدی زارعی

نام استاد:
دکتر صیاد حقیقی

تیر ماه ۱۴۰۰

به نام خدا

فهرست مطالب

۱. چکیده

کمتر کسی است که اهمیت یادگیری ماشین در پیشرفت علوم و فناوری های امروزی بر او پوشیده باشد. علی رغم پیشرفت های بسیار زیادی که این حوزه در دهه های اخیر کرده است، یک چالش جدی همچنان توجه متخصصیان را به خود جلب می کند و آن کار با داده های بزرگ، پراکنده و حساس است. اینجاست که federated learning معرفی می شود. سیستمی که با چارچوب خاص خود متناسب با نیاز های این نوع داده ها طراحی شده که شاید بتوان گفت تامین امنیت در آن مهم ترین نقش را بازی می کند.

یکی از راه های غلبه بر این چالش ها استفاده از یادگیری فدرال ایمن [۱]است، این روش یادگیری برای اولین بار توسط شرکت Google در سال ۲۰۱۶ معرفی شد .

این گزارش خلاصه مطالعاتی است که گروه ما در مورد federated learning و مسئله امنیت داده حین آموزش همگانی انجام داده است.

کلمات کلیدی:

Federated Learning – Machine Learning – Data Security – Secret Sharing – GAN's – Gradient Flow – Deep Leakage – IoT, etc.

[۱] secure federated learning

مقدمه

در سال ۲۰۱۶ با پیشرفت هوش مصنوعی به خصوص مقابله با انسان در برخی بازی ها و غیره پتانسیل عظیمی در این شاخه دیده شد ؛ این امر آغازی بود برای گسترش این شاخه در صنایع پیشرفته و پیچیده تری نظیر اتومبیل های بدون سرنشین ، امور مالی، مراقبت های پزشکی و ...

اما با مطالعه ی دقیق تر عوامل پیشرفت این حوزه متوجه خواهیم بود که یکی از تاثیر گزارترین آن ها استفاده از داده هایی با حجم بالا [1]است. به صور مثال برای آموزش یک مدل بازی AlphaGo به چیزی در حدود 300,000 اطلاعات بازی نیاز داریم تا مدل بتواند به نتیجه ی مطلوب برسد.

در نتیجه ی آن ما به جمع آوری داده های زیادی برای آموزش مدل ها نیازمندیم.

اما در دنیای واقعی صنایع شرایط کمی متفاوت است، در بیشتر حوزه ها داده های کافی و یا با کیفیتی در اختیار محققین نیست؛ ولی آیا برای حل این مشکل میوانیم داده هارا از بخش های گوناگون در یک جا جمع آوری کنیم؟

چند مشکل اساسی برای جمع آوری داده ها در یک مرکز روبروی ماست.

۱. هزینه ی زیاد انتقال داده ها
۲. تفاوت انواع داده ها
۳. حفظ امنیت داده

در بیشتر موارد جدای از اینکه شرایط انتقال داده ها میسر نیست، این انتقال چه مبتی بر شبکه و چه به صورتی فیزیکی و مستقیم نیازمند هزینه ی زیادی خواهد بود.

اما علاوه بر تفاوت انواع داده نیز مشکل بزرگیست به عنوان مثال در یک سیستم پیشنهاد [2]یک فروشنده ی کالا اطلاعاتی نظیر انواع کالاها واطلاعات خرید های قبلی اورا دارد، امال اطلاعی از توانایی مالی کاربر ها نداردو این میتواند باعث پیشنهادات اشتباهی به کاربر شود.

[1] Big Data
[2] recommendation service

در بییشتر صنایع اطلاعات به صورت مجزا در اختیار بخش های مختلفی است همچنین به دلیل رقابت بین این بخش و یا شرکت ها و همچنین به دلیل امنیت اطلاعات ، به اشتراک گذاشتن آن ها تقریبا امری محال است.

در سال های گذشته حفظ امنیت اطلاعات کاربران به موضووعی حیاتی و مورد توجه بدل شده به صورتی که بعد از اعتراضات گشترده در جامعه و انتشار آن ها در رسانه ها دولت ها اقدام به تصویب قوانینی در این زمینه کرده اند.

در نتیجه ی تمام این محدودیت یادگیری فدرالی راه حلی به نسبت جدید تر برای عبور از این چالش هاست.در این روش سعی برا این است که آموزش مدل ها به صورت گسترده و و توزیع شده [1]صورت بگیرد و به جای انتقال داده ، داده در محل خود باقی مانده و اطلاعات و متغیر های لازم برای فرآیند یادگیری ارسال شوند.

مثلا فرض کنید در آموزش یک مدل دستیار صوتی در گوشی های همراه میخواهیم بدون استفاده از اطلاعات شخصی کاربر نظیر شماره و نام مخاطبین ، سیستم را آموزش دهیم برای این کار مدل را به گوشی همراه میفرستیم و از کاربر (گوشی هوشمند) میخواهیم با استفاده از اطلاعات محرمانه ی خود دقت و عملکرد شبکه را مورد ازمایش قرار دهید و نتیجه ی این دقت را به سرور اصلی بازگرداند تا مدل در سرور اصلی اصلاح شود.

۱-۱- انواع یادگیری فدرالی

به صورت معمول این یادگیری شامل دو روش کلی می باشد که تفاوت این روش ها در انواع داده های مورد استفاده است.

فرض کنیم داده ها در دو بخش A و B تقسیم شده اند. با توجه به نوع این داده ها از دو روش یادگیری عمودی و یادگیری افقی استفاده میکینم.

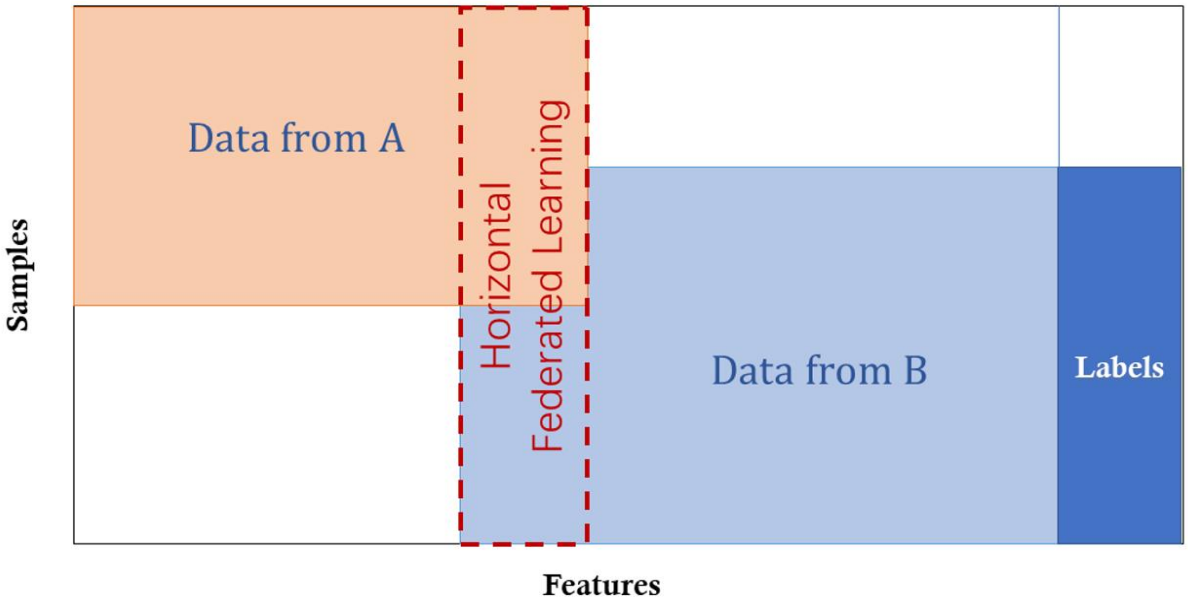
البته در شرایطی که داده در بخشی شرایط عمودی بودن و در بخش دیگری شرایط افقی بودن را داشته باشد همزمان از دو روش برای هر بخش مناسب استفاده میشود.

[1] distributed

۱-۱-۱- یادگیری فدرالی افقی[1]

در این روش داده ها شامل اطلاعات و ویژگی های مشابه هستند و تنها نمونه های مختلفی در هر دو بخش A و B وجود دارد برای مثال هم A و هم B بانک هستند و هرکدام اطلاعات مربوط به مشتریان خود را با ویژگی های یکسان نظیر میزان موجودی و ... در اختیار دارند البته در این روش A و B میتواند شامل نمونه های مشترک نیز باشند.

در این روش اطلاعات ارسالی از هر دو بخش برای ارسال داده با هم جمع شده و به یک بردار تبدیل میشود.



شکل ۱ - توزیع داده به صورت افقی

۱-۱-۲- یادگیری فدرالی عمودی[2]

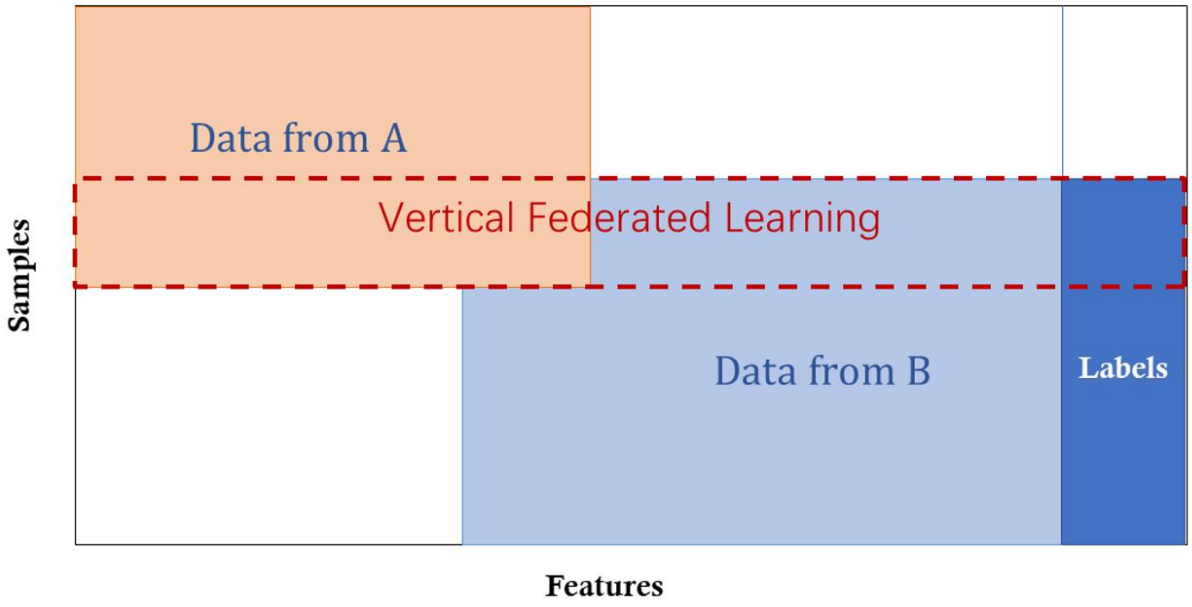
در این روش فرض بر آن است که نمنه های در اختیار هر دو بخش یکسان است ولی هر کدام اطلاعات و ویژگی های متفاوتی از نمونه ها در اختیار دارند. به طور مثال فرض کنید. A یک بانک و B یک شرکت بیمه است و هر

[1] Horizontal Federated Learning

[2] Vertical federated learning

دو به مشتریان یکسانی خدمت رسانی میکنند و هر کدام اطلاعات و ویژگی های مختلی از آن ها را در اختیار دارند.

در این روش اطلاعات منتشر شده از بخش های Aو B در کنار هم قرار داده میشود.یعنی به طور مثال اگر اطلاعات هر کدام شامل پنج ویژگی از مشتری باشد در نهایت یک بردار با بعد ۱۰ در اختیار داریم.



شکل ۲ - توزیع داده به صورت عمودی

۲. مراحل راه اندازی یک سیستم مبتنی بر FL

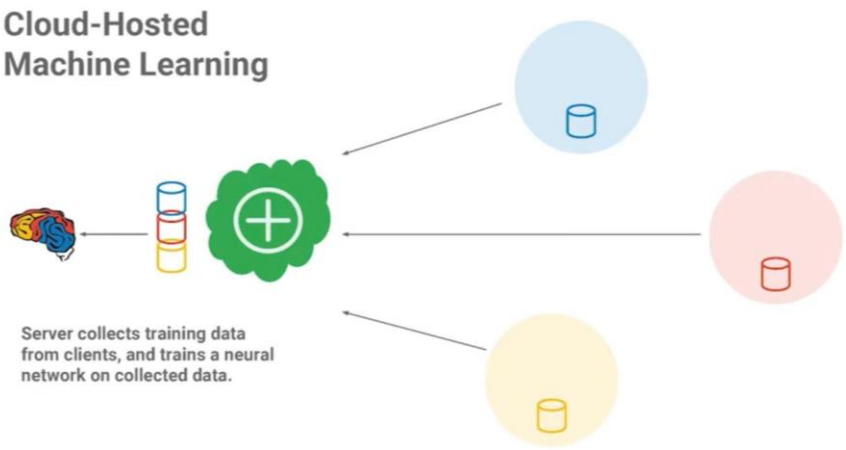
هدف یک سیستم مبتنی بر FL، آموزش یک مدل همگانی (Global)، روی داده های محلی (Local) مربوط به کاربران و پایگاه داده (Database) های مجزا از هم است، به شکلی که حریم شخصی هر یک از کاربران محفوظ بماند و دسترسی مستقیم به داده ممکن نباشد.

این مسیر را بر اساس طراحی و عیب یابی مرحله مرحله، به صورت برداشتی آزاد از ارائه تیم Federated Learning گوگل در کنفرانس ACM سال ۲۰۱۷، طی خواهیم کرد. (Practical Secure Aggregation for Privacy Preserving Machine Learning, 2017)

۳. جمع آوری داده ها در یک مرکز و آموزش یک مدل همگانی

اولین راهی که به ذهن می رسد جمع آوری داده ها در یک مرکز و آموزش دادن مدل بر اساس تمامی داده ها به صورت یکجا است.

راهکار مذکور به Cloud-Hosted ML معروف است. یک مثال حالت تک کاربره آن وقتی است که هر یک از ما از سرویس Google Collaboratory و GPU های مستقر بر سرور های گوگل برای اجرای برنامه هایمان استفاده میکنیم.



علی رغم آنکه این سیستم در دنیای امروزه استفاده می شود اما برای تحقق هدف ما، به چند دلیل مناسب نیست. از جمله:

- دسترسی مستقیم هر حمله کننده ای که وارد شبکه شود به داده، از طریق شنود
- دسترسی کاربران به داده های همدیگر از طریق جا زدن خود به جای سرور
- دسترسی سرور به داده های تمامی کاربران
- بار محاسباتی بالا به دلیل تکیه سیستم به یک سرور برای آموزش

نقض محرمانگی (confidentiality) و دست نخوردگی (integrity)

معمولا سیستم هایی که داده های به شکل مرکز در آن ها ذخیره شده یا جمع آوری می شوند از نظر امنیتی ریسک بالاتری دارند چرا که با حمله به یکی از node های سیستم می توان به صورت یکجا به تمام داده های سیتسم دسترسی پیدا کرد

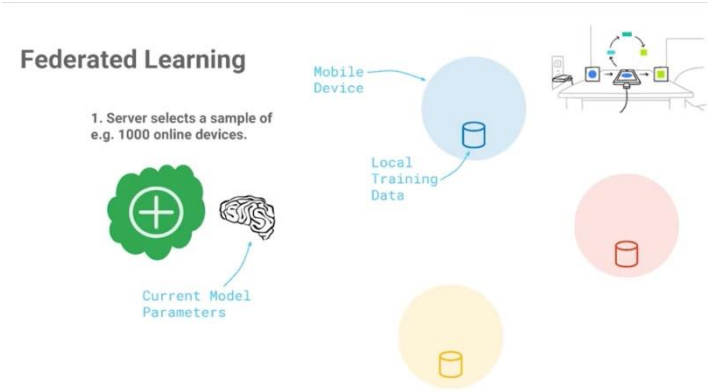
۴. آموزش محلی (Local) و ارسال گرادیان داده ها برای پرورش مدل همگانی (Global)

برای حفظ حریم شخصی کاربران یا باید داده را به صورت رمز شده ارسال کرد و یا به جای داده چیز دیگری که حاوی اطلاعات و خصوصیت مورد نیار داده برای آموزش مدل باشد را ارسال کرد. ابتدا به بررسی مسیر دوم می پردازیم.

- **حین آموزش داده ها در یک شبکه عصبی چه چیزی در اصل به روز رسانی می شود؟**

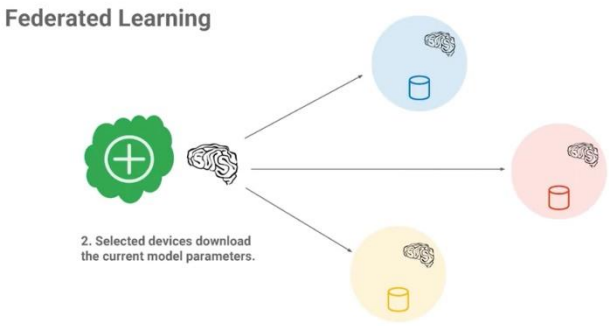
گرادیان تابع هدف نسبت

مراحل این روش به شرح زیر است:

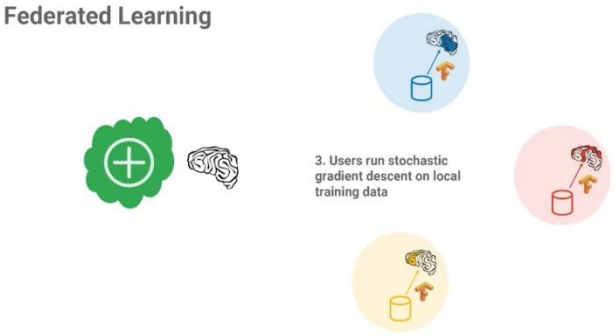


۱- سرور تعدادی از کاربران را که برای ارسال مدل ، به طور تصادفی انتخاب میکند.

(انتخاب تمامی کاربران در شبکه های بزرگ، باعث بروز مشکلاتی از جمله بار محاسباتی بالا، مشکلات مربوط به congestion control و traffic flow و ... می شود)

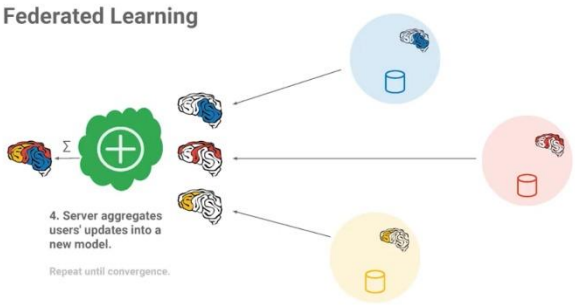


۲- در مرحله بعد، سرور یک مدل همگانی اولیه را که پارامتر های آن را initialize کرده برای هر یک از کاربران ارسال میکند و پارامتر های مدل توسط کاربران دانلود می شود



۳- حال کاربران با اجرای الگوریتم SGD (گرادیان نزولی آماری) ، مدل خود را آموزش می دهند و گرادیان های اپدیت شده را به عنوان معرف مدل خود برای ارسال به سرور آماده می کنند.

۴- سرور با توجه به اینکه پارامتر های مدل آن ها مشترک و یا متفاوت است، بردار گرادیان همگانی را از حاصل جمع و یا کنار هم قرار دادن گرادیان های محلی تشکیل می دهد و با اپدیت پارامتر های مدل همگانی، مدل را برای توزیع مجدد آماده می کند. سپس سیستم دوباره در مرحله ۱ قرار گرفته و این فرایند تا رسیدن به همگرایی کافی ادامه پیدا می کند.



استفاده از SGD و یا BGD به جای GD هم بار محاسباتی را هنگام محاسبه گرادیان کاهش می دهد و در نهایت به سرعت سیستم کمک می کند و هم به نحوی باعث ایجاد regularization و جلو گیری از فرابرازش (overfit) مدل به داده های محلی کاربران می شود.

شمای کلی یک سیستم مبتنی بر FL، به صورت زیر است:

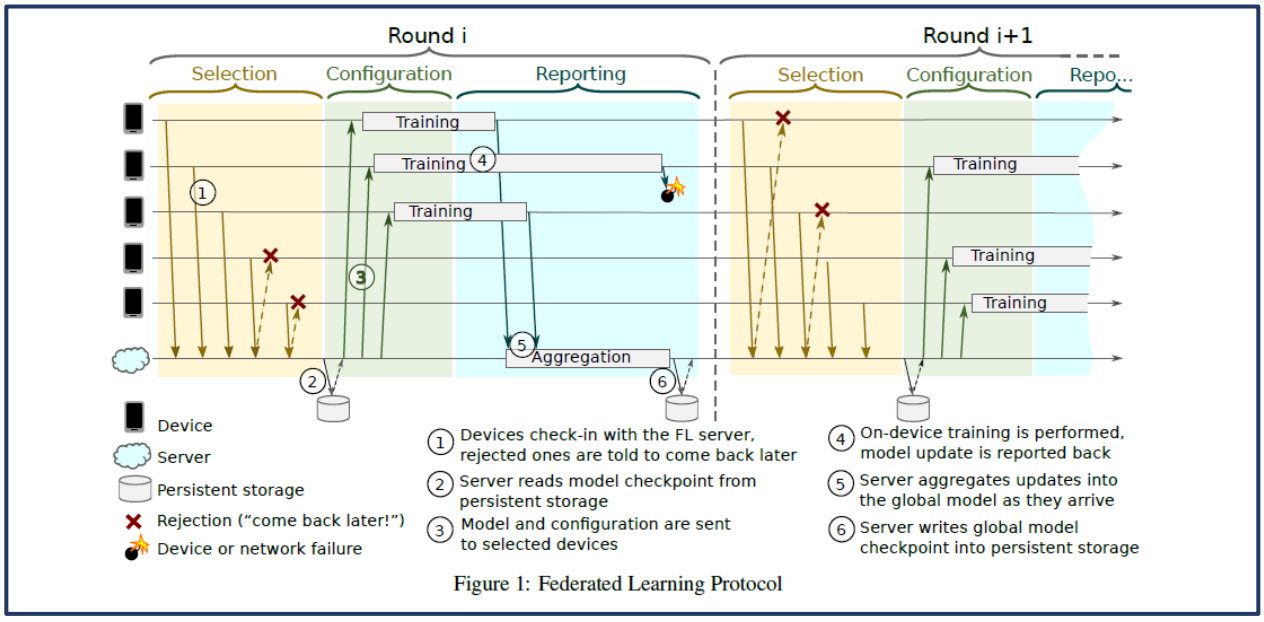
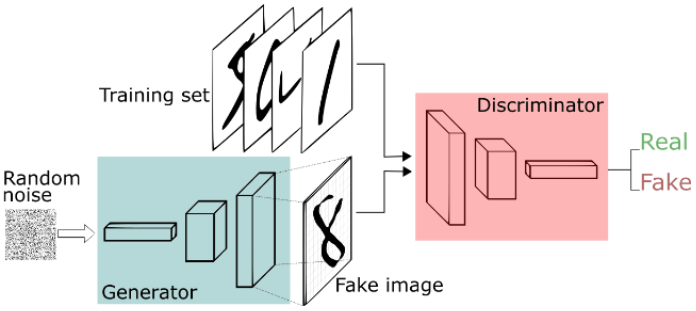


Figure 1: Federated Learning Protocol

مشکلی که برای این سیستم ها به وجود آمد، این بود که گرادیان ها هم امنیت را به

۵. مشکلات امنیتی و حملات مبتنی بر اطلاعات گرادیان



یکی از موضوعاتی که در سال های اخیر به شدت مورد توجه قرار گرفته استفاده از شبکه های GAN در یادگیری عمیق است. کار اصلی این شبکه تولید داده هایی مشابه داده هایی است که به عنوان داده های آموزشی به یک مدل اصلی داده می شود تا با دادن این داده های مصنوعی به شبکه قدرت تشخیص آن را بالا ببرد. شبکه اصلی را discriminator می نامند و شبکه فریب دهنده را Generator می نامند که همان چیزی است که اکنون تحت عنوان Generative Adversarial Network می شناسند

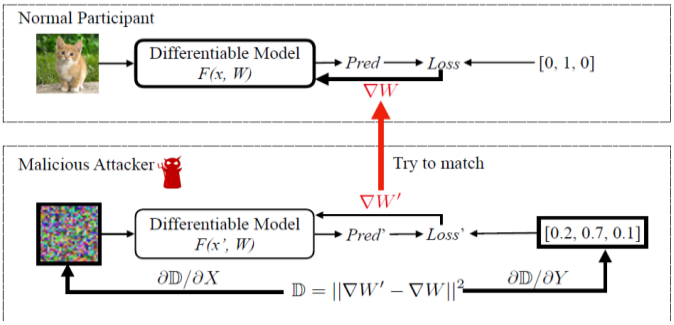
- پس این شبکه ها ذاتا برای مقاصد غیر قانونی و بر هم زدن امنیت استفاده نمی شوند و هدف آن ها افزایش قدرت تشخیص مدل های classifier است. با این وجود می توان با وارد کردن یک شبکه GAN به سیستم های مبتنی بر FL، داده های مصنوعی به سرور تحویل داد و علاوه بر تاثیر گذاری مدل های محلی هر کدام از کاربران هم تحت تاثیر قرار بگیرند و به این صورت یک اختلال کلی در آموزش پیش آید.

شبکه های GAN به شکل واضحی اصل دست نخوردگی امنیت را نقض می کنند.

در سال های اخیر، مقاله های متعددی به بررسی نشر اطلاعات و استنباط داده ها از روی اطلاعات گرادیان پرداخته اند. یکی از مهم ترین مقالات در این زمینه مقاله یک دانشجو PHD دانشگاه MIT بود که به کمک مدلی که ارائه داد نشان داد می توان داده های کاربران را در صورت استفاده از GD به طور کامل و در صورت استفاده از SGD , BGD از روی بردار گرادیان به طور تقریبا کامل بازیابی کرد. در این قسمت به بررسی اجمالی روش و نتایج این مقاله تحت عنوان "Deep leakage form gradients" می پردازیم. (Deep Leakage from Gradients, Dec 2019)

- **مسئله شناسایی داده ها** از روی گرادیان در این مقاله به صورت یک **مسئله بهینه سازی** ساده می شود. به این صورت که ابتدا یک بردار گرادیان dummy تعریف می شود و تابع هزینه فاصله این بردار گرادیان dummy که دست حمله کننده است از بردار گرادیان مشاهده شده در کانال است. با کمینه

کردن این تابع روی آرگومانی که همان داده آموزش است می توان به داده های آموزش رفته رفته نردیک شد. این کار مانند آن است که از روی جهت سع بعدی حرکت یک فرد در یک زمین و مقدار هر گام او رفته رفته بتوان شکل زمین را تخمین زد.



- فرض می کنیم گرادیان t امین اپدیت از کاربر k ام را در دست داریم. آن را با $\nabla W_{t,k}$ نشان می دهیم.

$$\nabla W_{t,1} = \frac{\partial L(F(x_{t,1}, W_t),\ y_{t,1})}{\partial W_t}$$ → User #1

$$\nabla W_{t,i} = \frac{\partial L(F(x_{t,i}, W_t),\ y_{t,i})}{\partial W_t}$$ — User #i

$$\nabla W_{t,N} = \frac{\partial L(F(x_{t,N}, W_t),\ y_{t,N})}{\partial W_t}$$ → User #N

بردار های گرادیان در آپدیت t ام

- مراحل بازیابی داده در روش Deep Gradient Leakage به صورت زیر است:

۱- به صورت رندوم یک بردار تصادفی و برچسب تصادفی به عنوان داده اولیه اختیار می کنیم آن ها را $(x_{i,k}, y_{i,k})$ نام گذاری می کنیم.

۲- مدل همگانی فعلی را روی داده ها اعمال میکنیم و گرادیان را نسبت به پارامتر های مدل محاسبه میکنیم.

$$\nabla W' = \frac{\partial L(F(x', W), y')}{\partial W}$$

۳- تابع هزینه را مطابق عبارت مقابل تعریف می کنیم تا با تکرار این مراحل به داده واقعی نزدیک تر شویم:

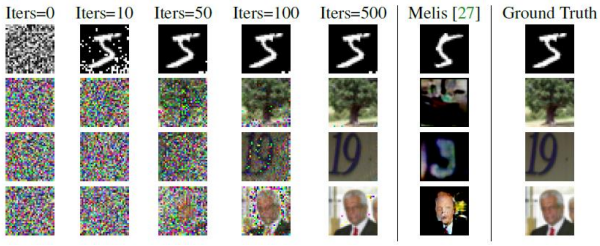
$$L = \|\nabla W' - \nabla W\|^2$$

۴- عبارت زیر بازیابی داده ها تحت مدل ارائه شده را توصیف می کند.

$$x'^*, y'^* = argmin_{x', y'} \|\nabla W' - \nabla W\|^2$$

$$= argmin_{x', y'} \left\| \frac{\partial L(F(x_{t,i}, W_t), y_{t,i})}{\partial W_t} - \nabla W \right\|^2$$

نویسندگان این مقاله از فرایند بالا تحت عنوان Gradient Matching یاد کرده اند. نتایج شگفت انگیز این مدل در تصاویر زیر مشخص است:



احتمال موفقیت حمله در صورت استفاده کابران از SGD و BGD چقدر کاهش خواهد یافت ؟

تقریبا هیچ ! تعداد تکرار های لازم برای بازیابی داده بیشتر می شود و ترتیب ها هم ممکن است به دلیل عدم آگاهی حمله کننده از اینکه گرادیان مربوط به کدام داده است عوض شود اما تصاویر تقریبا کامل بازیابی شده اند

نتایج استفاده از SGD , BGD روی مجموعه داده CIFAR به شکل زیر است:

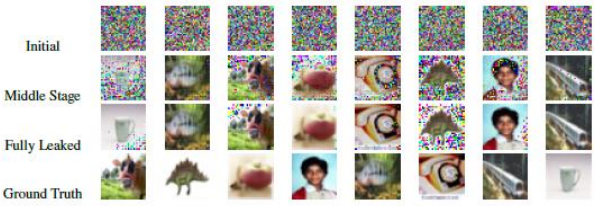


Figure 4: Results of deep leakage of batched data. Though the order may not be the same and there are more artifact pixels, DLG still produces images very close to the original ones.

	BS=1	BS=2	BS=4	BS=8
ResNet-20	270	602	1173	2711

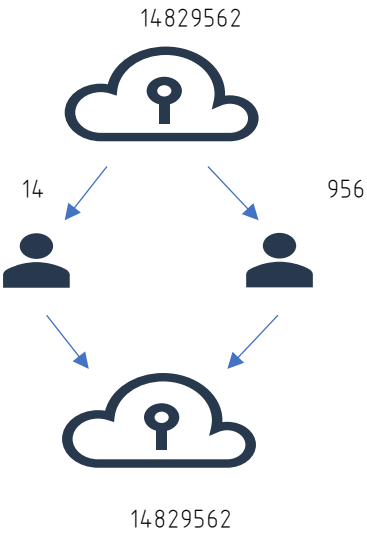
Table 1: The iterations required for restore batched data on CIFAR [21] dataset.

ایده ای که برای رفع مشکل گرادیان توسط گوگل داده شد، بر مبنای موضوع بنیادی در امنیت تحت عنوان "به اشتراک گذاری راز" (Secret Sharing) است. ابتدا به بررسی چگونگی به اشتراک گذاشتن یک secret در فضای چند جمله ای ها به کمک یک شهود هندسی می پردازیم.

۶. Secret Sharing – روش های به اشتراک گذاری راز ها

اولین راهی که برای توزیع یک secret ارائه می شود تولید کپی های متعدد از آن و دادن هر کپی به یک user دارای مجوز است. اما مسائلی که secret sharing برای آن ها استفاده می شود معمولا مسائلی هستند که یک نفر به صورت تنها نباید اجازه دخالت در آن یا آگاه شدن از آن را داشته باشد. دو نمونه از این دست مسائل پسورد های معاملات بزرگ بانکی و کد های پرتاب موشک هستند.

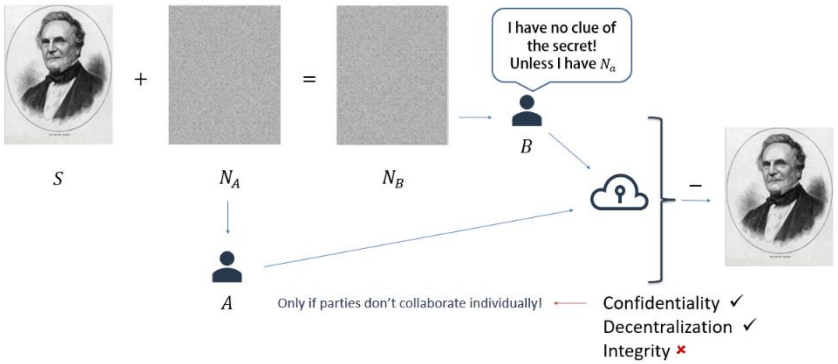
راه بعدی که برای به اشتراک گذاری یک secret به ذهن می رسد چند تکه کردن آن و توزیع آن تکه ها بین افراد مختلف است. بدین شکل هیچ یک از افراد به شکل کاملی به secret دسترسی ندارند اما اگر یک منبع قابل اطمینان از تمامی کاربران درخواست سهم خودشان را از secret بکند، می تواند با سرهم کردن قسمت های مختلف secret را بازیابی کند.



این راه هم همانند راه اول مشکلات زیادی دارد که مهمترین آن این است که:

فضای جستجوی لازم برای brute force کردن secret کاهش می یابد.

از روش های دیگر برای مخفی کردن یک secret اضافه کردن نویز به آن است. به این صورت که اگر برای تصویری مثل تصویر S، یک نویز جمع شونده رندوم تولید کنیم و سپس خود نویز را به شخصA و تصویر حاصل را به شخص B بدهیم هیچ یک با در دست داشتن تصویر دیگری نمی توانند تصویر اصلی یعنی S را بازیابی کنند.

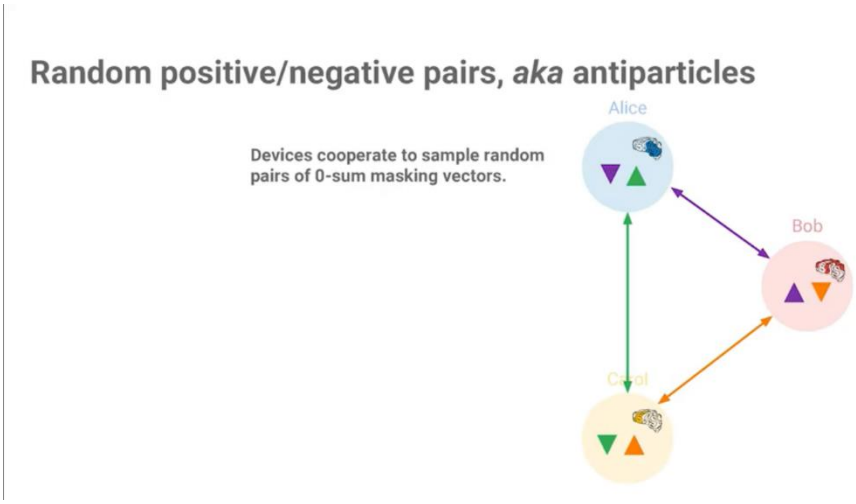


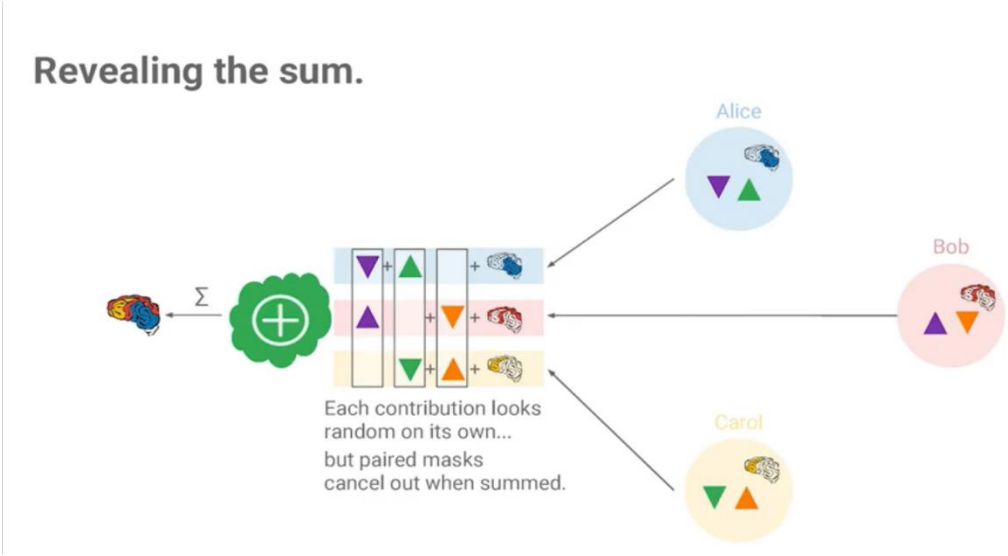
به این ترتیب می توانیم یکی از فاکتور های امنیت که محرمانگی secret می باشد با فرض عدم همکاری افراد با هم تامین کنیم.

۷. Masking

Vectors – ارسال گرادیان به کمک secret sharing

بر پایه روشی که معرفی شد بردار هایی به نام masking vector شامل نویز های قرینه تولید می شوند که به هر زوج کاربر اختصاص داده می شود. کاربران با xor کردن (mask کردن) این بردار تصادفی با بردار گرادیان خودشان به جای ارسال خود گرادیان ها گرادیان های نویزی را ارسال می کنند. بازم به ذکر است که این بردار ها باید روی کانال امن مبادله شوند وگرنه امنیت روش زیر سوال خواهد رفت.



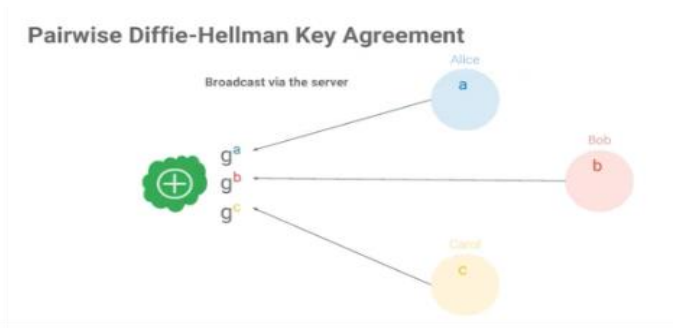


سرور با در دست داشتن بردار های گرادیان نویز ارسال شده درصورتی که گرادیان هارا جمع کند و به نحوی aggregation را انجام دهد نویز ها با هم خنثی می شوند و حاصل جمع گرادیان ها بدون نویز به وجود می آید. از آنجا محرمانگی در این روش تامین می شود به آن secure aggregation می گویند.

اما این روش هم خالی از اشکال نیست و مهم ترین اشکال آن این است که:

Masking Vector ها به شدت طولانی اند (به اندازه خود بردار گرادیان) و overheadبالایی به سیستم یادگیری ما اعمال می کنند. شبکه های عصبی که با

راه حلی که برای این مشکل در نظر گرفته شده استفاده از Pairwise Diffie Hellman است. بدین صورت یک Diffie Hellman share میان هر دو کاربر به اشتراک گذاشته می شود که کمک میکند تا این بردار طولانی یک بار کمتر مبادله شود. سپس با استفاده از یک PRNG و دادن این share به عنوان ورودی به آن یک رشته تصادفی تولید می شود که با داده جمع شده و روی کانال قرار می گیرد.



۸. کاربردها و چالش‌ها

۲-۱- کاربرد در زمینه‌ی اینترنت اشیاء و شهر هوشمند

همانطور که تا اینجا به آن اشاره شده است، هدف اصلی و عمده‌ی FL، حفظ بیشتر حریم شخصی کاربران می‌باشد. در زمینه‌ی شهر هوشمند (و به طور کلی، اینترنت اشیاء)،دستگاه‌های مختلف نیتز به ارتباط مداوم و یا لحظه‌ای با یکدیگر دارند. با کمک Federated Learning، می‌توان ضریب امنیت و همچنین احتمال به خطر افتادن داده‌ها را کاهش داد. این نکته یک رکن اساسی برای استفاده از FL در زمینه‌ی اینترنت اشیاء محسوب می‌شود. در طراحی یک الگوریتم برای انجام محاسبات، باید به این نکته نیز توجه داشت که این الگوریتم قرار است توسط چه دستگاهی و با چه سرعت و فرکانسی پردازش شود. در حقیقت طراح، باید قدرت پردازشی دستگاه هدف را مد نظر داشته باشد و الگوریتم طراحی شده را بر اساس آن بهینه کند. همچنین برای ارتباطات حساس می‌توان از بلاک‌چین نیز استفاده کرد.

۱-۳- کاربرد در زمینه‌ی وسایل شخصی

یکی از کاربردهای FL، مربوط به کاربردهای شخصی می‌شود. به عنوان مثال، کیبورد گوگل، Gboard، برای پیش‌بینی کلمات و یا auto-correct از Federated Learning استفاده می‌کند. استفاده از FL، باعث می‌شود که داده‌ای بین کاربر و سرور جابه‌جا نشود؛ بلکه فقط الگوریتم بین آن دو انتقال یابد. برنامه‌های دیگری مانند دستیاران صوتی مانند Siri نیز در حال استفاده از FL برای مدل کردن تابع خود هستند.

۱-۴- کاربرد در زمینه‌ی حمل و نقل

۱-۴-۱- حمل و نقل زمینی

در بحث حمل و نقل زمینی، ارتباط خودروها با یکدیگر و با محیط (کاربرهای شهر هوشمند نیز محسوب می‌شود) بسیار مهم است. در حال حاضر، بسیاری از شرکت‌های مطرح خودروسازی، به سمت استفاده از خودروهای خودران و همچنین خودروهای الکتریکی پیش می‌روند. در زمینه‌ی خودروهای خودران، خودرو نیاز حیاتی به ارتباط با محیط اطراف خود دارد. حال، در صورت استفاده از FL به عنوان روش برقراری ارتباط با محیط، می‌توان امنیت آن را افزلیش داد. همچنین در حوزه‌ی خودروهای الکتریکی، می‌توان با این روش یک مدل جامعی از مصرف انرژی در خودروها بدست آورد و با استفاده از آن، عمر باطری خودرو را بهبود داد. همچنین می‌توان ایستگاه‌های شارژ خودروها را بر اساس میزان شلوغی در ساعات مختلف شبانه‌روز بدست آورد. با این کار، خودروها در ایستگاه‌های مختلف پراکنده می‌شوند. توجه داشته باشید که فرآیند شارز خودروهای برقی به مراتب طولانی‌تر از پر کردن باک یک ماشینی است که از سوخت فسیلی استفاده می‌کند.

۱-٤-۲- حمل و نقل هوایی

در این بخش، ابتدا به موضوع تشخیص خطا می‌پردازیم. تشخیص دادن خطا برای یک سیستم حمل و نقل هوایی که داده‌های زیادی تولید می‌کند، با توجه به توان پردازشی آن بسیار مشکل است. بنابراین، می‌توان بر اساس یک سیستم Online Decision Tree این عملیات را انجام داد. برقراری ارتباط از نوع FL می‌باشد. این عمل، باعث کاهش نیاز به توان محاسباتی بالا می‌شود. یک بخش دیگر، مربوط به سیستم‌های هوایی بدون سرنشین می‌باشد. در این دسته از سیستم‌ها، انتقال اطلاعات حیاتی‌ترین بخش مسئله می‌باشد. بیشتر پهنای باند مشغول

فرستادن اطلاعات از سیستم به سرور (کاربر) می‌باشد. بت استفاده از Federated Learning، نه تنها می‌توان امنیت بالاتری بین سیستم و سرور بدست آورد؛ بلکه باعث می‌شود که پهنای باند کمتری نیز اشغال شود.

۱-٤-۳- چالش‌ها

نکته‌ای که باید به آن توجه داشت، این است که ما در حال صحبت درباره‌ی سستم‌هایی هستیم که متحرک می‌باشند. از این رو، توان پردازشی این سیستم‌ها به مراتب از سیستم‌های ثابت کمتر است. بنابراین باید در طراحی الگوریتم به این موضوع توجه ویژه‌ای داشته باشیم؛ که انجام دادن این دسته از عملیات‌های پردازشی، موجب نشود که زمان پاسخ‌گویی افزایش یابد. یکی دیگر از موضوعات، این است که در انتقال بسیار زیا مدل بین کاربر و سرور، ممکن است اطلاعات حساس قابل کشف باشند. بنابراین باید یک روش دینامیکی طراحی کرد که این ریسک را کاهش دهد. در نهایت بحث انرژی مصرف شده برای انجام عملیات مطرح است. باید الگوریتم به نحوی باشد که یک تعادل بین مصرف انرژی و کارایی دستگاه را شاهد باشیم.

۱-۵- کاربرد در زمینه‌ی بانک‌داری و مالی

۱-۵-۱- کاربردها

در زمینه‌ی بانکی و مالی، با دو کاربرد مهم روبه‌رو هستیم. کاربرد اول مربوط به تشخیص کلاهبرداری است. اید در ابتدا نیز به این نکته توجه داشت که بانک‌ها مایل نیستند تطلاعاتی از حساب مشتریان خود را با بقیه موسسات مالی در ارتباط قرار دهند. ما می‌توانیم با استفاده از FL، آن‌ها را به استفاده از Model Learning تشویق کنیم؛ زیرا در این حالت اطلاعاتی بین بانک‌ها منتقل نمی‌شود و فقط مدل جابه‌جا می‌شود. حال می‌توانیم کدل خود را به گونه‌ای train کنیم که در صورت وقوع هر گونه کلاهبرداری، آن را تشخیص دهد. یکی از دلایل موفقیت بیشتر این روش، استفاده از جامعه آماری بیشتری است که FL مهیا می‌کند. کاربرد مهم دیگر این است؛ که FL می‌تواند به شرکت‌های بیمه‌ای کمک کند تا ریسک را بهتر مدیریت کنند.

۱-۵-۲- چالش‌ها

یکی از چالش‌های پیش‌رو، عدم یکسان بودن جامعه‌ی آماری بین موسسات مالی می‌باشد. این ناهمگونی ممکن است باعث شود که مدلی که بدست می‌آید برای تمام شرکت‌ها قابل استفاده نباشد. یکی دیگر از چالش‌ها، همانطور که در بالاتر توضیح داده شده، عدم تمایل موسسات مالی بزرگ نسبت به اشتراک گذاشتن اطلاعات به هر روشی

می‌باشد. در حقیقت FL باید در زمینه‌ی امنیت و حریم شخصی به یک تکاملی دست یابد تا بتواند نظر موسسات کلان را به خود جذب کند.

۱-٦- کاربرد در زمینه‌ی پزشکی

۱-٦-۱- حریم شخصی

بسیاری از کشورها، قوانین سخت‌گیرانه‌ای نسبت به اشتراک گذاشتن اطلاعات یک بیمار دارند. حال، با وجود Federated Learning، می‌توان یک مدل خوب نسبت به نحوه‌ی درمان یک بیماری بدون به اشتراک گذاشتن تطلاعات یک بیمار بدست آورد. در حقیقت، استفاده از FL نقض قوانین داخلی کشورها محسوب نمی‌شود و می‌توان باعث افزایش موفقیت روند درمان شود.

۱-٦-۲- تحقیق در ارتباط با داروها

روند انجام آزمایشات بالینی نیز، تحت قوانین سختگیرانه‌ی محلی قرار دارد. با استفاده از FL می‌توان بدون نشر اطلاعات یک کاربر، باعث کاهش زمان تولید یک دارو شد. در طی سال گذشته، روند توسعه دارو و واکسن بیماری کووید-۱۹ را نیز با استفاده از این روش شاهد بودیم.

۱-٦-۳- پیش‌بینی کردن وقوع یک بیماری

با استفاده از FL، می‌توان در صورت وقوع یک بیماری در یک منطقه، به سرعت آن را شناسایی کرد. این روش باعث می‌شود پیش‌بینی سریع‌تر رخ دهد و زمان بیشتری به واحدهای مدیریتی به منظور کاهش خطر بدهد.

٩. مراجع

1. (Practical Secure Aggregation for Privacy Preserving Machine Learning, 2017) - K. Bonawitz & Google AI team – ACM conference 2017

2. (Deep Leakage from Gradients, Dec 2019) – L. Zhu, Z. Liu, S. Han

3. Federated Learning; Concept and Application, Q.Yang, Y.Liu, 2019

4. Towards Federated Learning at Scale, K. Bonawitz & Google AI team, 2019