**《A little is enough: Circumventing defenses for distributed learning》**

**文章的主要思想是：**

1. 现有防御算法通过对客户端提交的参数进行分析，从而区别恶意客户端和良性客户端。这些防御算法（methods – Krum、Trimmed Mean）的假设前提是：各个客户端所提交的更新参数是服从正态分布的。

2. 基于1中的假设，本文的思想是：

1）攻击者可以利用所控制的客户端来掌控更新参数值的中位数，从而使得攻击能够绕过防御算法。

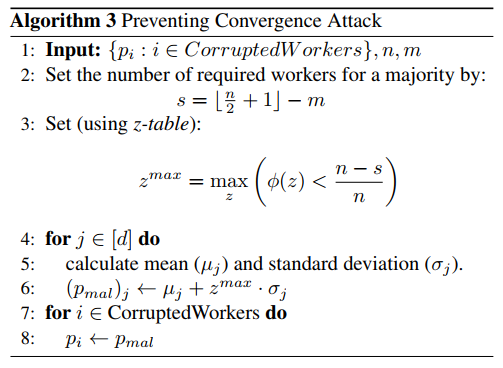
2）利用正态分布的特点，通过确定最大的z值，来保证攻击者所提交的更新在保证隐蔽性的前提下，最大化攻击效果。

**文章的优势：**

可以使用同一组参数设定对多种防御算法进行规避。

**攻击目标与手段：**

1. 使模型无法收敛：



攻击者最大化z值，然后对更新的每个维度的值进行“+”的偏移。

举例：

客户端总数量：50

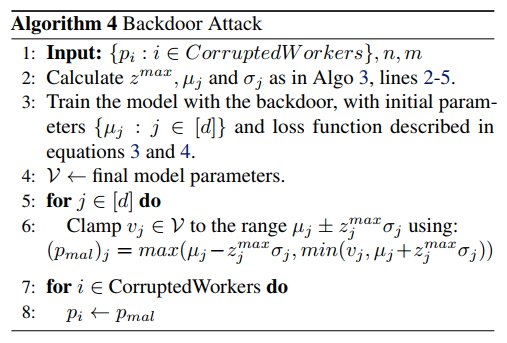
恶意客户端数量：24

需要“seduce”的客户端数量：

最大化z值：查找z-table，取最大的z，使得，查得。

攻击操作：对每个恶意客户端，对每一个服从的参数，设置其为。

1. 后门攻击：



基础算法是参考How to backdoor federal leaning 中的基于scale的攻击方法，利用公式（3）中的Loss对本地恶意模型进行训练。在保证后门能够成功注入的情况下，尽可能“小”地设置α以避免攻击者提交的参数过于异常。

本文贡献在于，攻击者可以利用每个参数的知识，而不是直接使用任何距离来表示.

**防御实验：**

本文提出的方法对Krum、TrimmedMean和Bulyan这三种防御方法都进行了有效的攻击，相较于这三种防御方法，针对本文攻击最有效的方法是No Defense——不采取任何防御措施。在No Defense下，本文的方法对模型总体只产生了微弱的偏移影响。

No Defense是不现实的——对于No Defense的情况，只需在本文的基础上单独设置一个针对No Defense的攻击客户端即可。

**文章特点：**

1. 基于scale的攻击。

2. 利用防御算法对参数的假定——服从正态分布，进而保证攻击者大概率掌控每轮全部更新的中位数以逃避防御。

3. 不基于欧氏距离，而是基于概率分布设置了以保证新参数更接近原始参数。