

# TrustNet: eller hur man håller chatt rumsren med hjälp av tillit och algoritmer

Alexander Cobleigh

June 2020

Kan vi ta vara på våra relationer till varandra för att delegera ansvar även i digitala miljöer? Hur kan vi använda tillit för att skapa ett slags digitalt immunförsvar?

Varje dag förlitar vi oss på varandra, liksom på samhällets institutioner. Vi förlitar oss dagligen på vår tillit till de mänskliga och naturliga system som finns omkring oss. Hittills har däremot denna tillit varit bristfällig i den digitala domänen. Visserligen finns det en viss tillit när internetbetalningar görs, eller när vi förlitar oss på att tjänster som dagens marknadsekonomiska jättar tillhandahåller inte ska göra intrång i vår integritet, och att de ska fortsätta fungera. Men den sortens tillit som vi gör oss till nytta av när vi delegerar uppgifter i arbetslivet, eller litar på vänners rekommendationer, är svårfunnen i det digitala.

En domän där denna sortens tillit skulle komma väl tillhands är chattsystem, där det är väldigt lätt för en ondsint aktör, som till exempel en spammare, att ha utomordentligt stor påverkan på övriga deltagare. Ett vanligt problem är att så kallade moderatorer (de som sorterar bort ondsinta meddelanden) antingen blir överbelastade av sitt volontärarbete, eller att maktpositionen missbrukas.

Det visar sig vara möjligt att använda våra sociala tillitsband till varandra för att kunna åstadkomma samma typ av bortsortering av ondsinta aktörer som med moderatorer. Sättet detta görs är genom att uttrycka vem vi litar på, hur mycket vi litar på deras omdöme, samt vilket område tilliten begränsas till. Genom att uttrycka tillit på det här sättet, för fyra till fem kompisar, skapas ett slags tillitsnätverk. Detta tillitsnätverk gör det möjligt att delegera ansvar, som att gömma elakartade meddelanden, till de som vi litar på allra mest. Skulle tilliten utnyttjas är det lika möjligt att ångra den, och samtidigt häva någon effekt som den numera opålitliga parten orsakat. Vad vi får till slut är ett slags digitalt immunförsvar, där du, dina kompisar, och dina kompisars kompisar, alla arbetar tillsammans för att försvara varandra från ondsinta aktörer.

Sättet detta fungera på är att vi med hjälp av tillitsnätverken kan rangordna de mest pålitliga personerna, sett från en person och deras perspektiv. Det är nämligen så att tillitsnätverket för varje person med stor sannolikhet kommer att se annorlunda ut; vi har ju olika kompisar som vi litar olika mycket på, och det samma gäller för deras kompisar och deras kompisars kompisar. Med hjälp

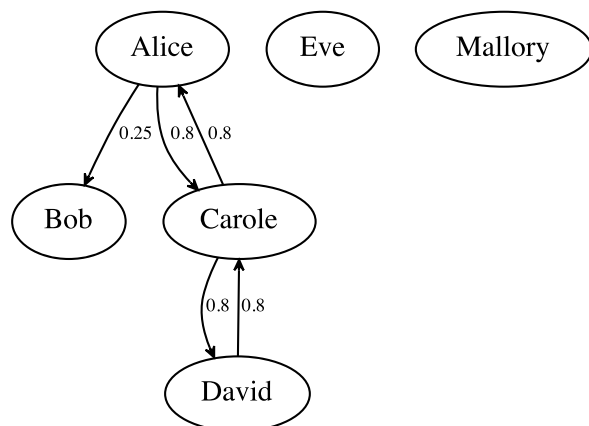


Figure 1: En enkel tillitsgraf för ett hypotetiskt chattsystem. Eve och Mallory har inte uttryckt någon tillit.

av denna rangordning är det sedan möjligt att göra ett snitt i den, så att vi kan sortera bort de med för lite tillit för att kunna vara pålitliga moderatorer, samtidigt som vi behåller resten och anser dem vara moderatorer.

Vad som kan vara överraskande är att ett moderationssystem som bygger på ovanstående rangordnade tillitsnätverk skulle i praktiken innebära att olika deltagare kan se olika meddelanden, allt beroende på vem som litas på. Jag och mina kompisar, och deras kompisar, skulle kunna ha en vy på chatten - medan en annan kompisgrupp skulle kunna ha en annan. Det är däremot sannolikt att vyerna skulle överlappa, då det som främst sorteras bort är ondsinta aktörer: spam. Detta eftersom de som missbrukar sin makt förlorar tilliten från andra.

Detta sätt att utforma ett moderationsystem är dessutom ypperligt lämpat för framtidens chattsystem, som kommer att fungera utan konstant internetuppkoppling. Detta görs redan idag genom att, istället för att rådfråga ett centralt datorsystem om de senaste meddelandena, helt enkelt direkt fråga datorerna för de människor du chattar med. Detta har innebörden att de meddelanden som faktiskt visas i chatten kommer helt från den egna datorn och dess lokala lagring. Varför ett typiskt moderationssystem inte skulle fungera för den här typen av chattsystem är på grund av att när centraldatorn försvinner, så försvinner också möjligheten att fråga den om vilka som har rättigheter att agera som moderator. Det beskrivna chattsystemet fungerar däremot bra med den tillitsbaserade moderationen eftersom ett tillitsbaserat system är helt subjektivt, alltså baserat på det egna perspektivet, men även eftersom alla meddelanden (inklusive de som beskriver vem som litas på vem) finns lokalt; det vill säga, all information vi behöver för att veta vem som kan agera som moderator å våra vägnar finns tillgänglig.