Använda AI för att analysera dataloggar

Ingress:

Datasäkerhet har blivit allt viktigare i takt med att allt mer känslig information lagras digitalt. Datorprogram sparar digitala spår av vad som händer på system, och med denna information borde man kunna upptäcka dataintrång eller systemfel. Men mängden data är enorm och en människa hinner därför inte med. Vi har därför utvecklat en AI-lösning för denna analys.

Artikel:

När program körs på en dator eller server, sparar systemet information om interna händelser i textfiler, så kallade loggar. Dessa loggar innehåller detaljerad information om vad som händer under programmets körning, vilket innebär att även hackerattacker lämnar spår efter sig.

Dagens datorsystem är komplexa och genererar enorma mängder loggdata. Detta gör att viktig information ofta drunknar i loggar om helt vanliga händelser. Mängden data är helt enkelt för stor för att en människa ska hinna analysera den, samtidigt som nya loggar ständigt genereras. Det finns därför ett tydligt behov av att automatisera processen att hitta avvikelser i loggar. Vi har utvecklat en AI-modell som kan lära sig att hitta dessa avvikelser på ett mycket snabbare och effektivare sätt än vad en människa skulle kunna göra.

Vi har utgått från ett tidigare examensarbete som påbörjat processen med att automatisera detta. Genom att använda deras arbete som grund har vi vidareutvecklat och integrerat en mer avancerad AI-modell. Detta har lett till ett bättre resultat och vi kan med högre träffsäkerhet identifiera avvikelser i loggfilerna. Det betyder att en slutprodukt som bygger på vår modell med högre sannolikhet kan upptäcka dataintrång.

AI-modellen som vi har använt oss av är en så kallad Transformer-modell. Det är en relativt ny modell som har visat sig särskilt framgångsrik inom textbaserade problem och att förstå innehållet samt meningen i text.

För att möjliggöra denna analys har vi använt AI för att, väldigt förenklat, omvandla loggdata till punkter i ett rum. AI:n är instruerad att placera vanliga loggar som punkter nära mitten och ovanligare loggar som avviker från det normala längre ifrån mitten. Datorn bestämmer sedan ifall en logg är en avvikelse eller inte beroende på hur nära mitten den är placerad. Genom detta kan en dator snabbt avgöra när det behöver larmas om ett pågående fel.

```
081109 203741 181 INFO dfs.DataNode$PacketResponder:
Received block blk_-8165149451366912526 of size
67108864 from /10.251.90.239

081109 203741 182 INFO dfs.DataNode$PacketResponder:
PacketResponder 1 for block blk_-8165149451366912526
terminating
```