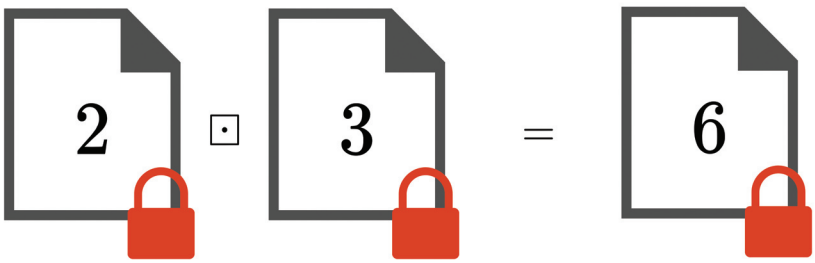
**Auteur:** Dieuwke van der Ende is werkzaam bij het ministerie van Defensie als Cyber Security Researcher. Ze heeft dit onderzoek gedaan bij TU Delft in samenwerking met TNO. Dieuwke is bereikbaar via dieuwkevdende@gmail.com.


# Onderzoek: evalueren beslisboom met privé input data

Op 6 oktober 2021 vond de uitreiking van de Joop Bautz Information Security Award plaats, waar ik de eer had om samen met Bhaskar Dercon, Jeroen Gaiser en Dilara Toprakhisar onze scripties te presenteren. Ik was erg verrast en blij toen ik hoorde dat ik de award had gewonnen! Voor mijn scriptie heb ik onderzoek gedaan naar het evalueren van een beslisboom terwijl de input data, die afkomstig is van meerdere partijen, privé blijft. Hiervoor heb ik gebruik gemaakt van een techniek genaamd homomorfische encryptie.

$$2 \cdot 3 = 6$$




*Afbeelding 1 - Homomorfische vermenigvuldiging.*

Beslisbomen kennen vele toepassingen. Denk bijvoorbeeld aan de detectie van malware, spam, fraude of het doen van de juiste productaanbevelingen (Portugal et al., 2018; Gibert et al., 2020). Daarnaast kunnen beslisbomen worden toegepast bij toegangscontroles, door te toetsen of iemand toegang mag hebben tot bepaalde systemen of documenten. Een belangrijk voorbeeld daarvan is de medische wereld, waar het voor een goede behandeling van patiënten cruciaal is dat artsen toegang hebben tot hun medische dossiers. Sommige artsen pleiten zelfs voor één Elektronisch Patiënten Dossier (EPD), waarin alle medische dossiers van patiënten te vinden zijn (Pieterman, 2020). Voor een goede implementatie is het van essentieel belang dat er een toegangscontrole is die de huidige context in beschouwing neemt en ervoor zorgt dat er alleen toegang wordt verleend aan de juiste artsen. Alleen zij mogen bij de bestanden die daadwerkelijk nodig zijn voor de huidige behandeling van de patiënt. Het gebruik van beslisbomen is hier een mogelijke oplossing.

## Gevoelige data

Beslisbomen hebben voor het maken van keuzes toegang nodig tot data die vaak gevoelig of privé is. Deze data is meestal afkomstig van verschillende organisaties voor wie het delen van gevoelige informatie buiten de organisatie vaak moeilijk of zelfs onmogelijk is. Banken kunnen samenwerken en hun financiële dossiers combineren voor het detecteren van fraude (Sangers et al., 2019) en daarbij gebruikmaken van bijvoorbeeld beslisbomen. Samen kunnen ze meer conclusies trekken dan alleen, maar het delen van financiële data is vaak moeilijk. Ook voor het juist functioneren van een beslisboom als toegangscontrole tot medische dossiers, is gevoelige data nodig van bijvoorbeeld ziekenhuizen, huisartsen en/of spoedposten. Dit kan data zijn over bijvoorbeeld artsen, patiënten en inhoud van de dossiers. Het gebruik van beslisbomen in deze toepassingen kan dus alleen plaatsvinden wanneer de privacy van de input data wordt gewaarborgd.

## Collaboratieve setting

Er zijn meerdere onderzoeken gedaan naar het evalueren van beslisbomen waarbij de vertrouwelijkheid van de input data wordt gewaarborgd (Tai et al., 2017; Tueno et al., 2020). Echter, geen van deze oplossingen kan worden gebruikt wanneer de input data vanuit meerdere organisaties afkomstig is. Dit noemen we een collaboratieve setting. Samenwerking bij het evalueren van beslisbomen is steeds meer nodig en is vaak alleen mogelijk als de privacy van de data van alle partijen wordt gewaarborgd. Ons onderzoek zet de eerste stap in de richting van het evalueren van een beslisboom terwijl de input data, dat afkomstig is uit meer dan één bron, privé blijft.

## Homomorfische encryptie

Ons werk maakt gebruik van homomorfische encryptie. Dit is een type encryptie dat ervoor zorgt dat berekeningen – wis-

kundige operaties –, gedaan kunnen worden over versleutelde data. Het resultaat is een encryptie van de waarde die de berekeningen zouden hebben over de normale, niet geëncrypte data. Zoals te zien in afbeelding 1 maakt homomorfische encryptie het mogelijk om een vermenigvuldiging van 2 en 3 te doen wanneer deze waarden zijn versleuteld (en dus niet zichtbaar).

Deze versleuteling, of encryptie, is in de afbeelding aangegeven met het rode slot. Het resultaat van deze homomorfische vermenigvuldiging, is een encryptie van de waarde 6. De daadwerkelijke waarde 6 is alleen zichtbaar als iemand over de juiste sleutel beschikt die `het rode slot kan openmaken', ofwel de encryptie kan ontcijferen.

Aangezien beslisbomen bestaan uit meerdere berekeningen of vergelijkingen van stukjes data, kunnen we de input data versleutelen en de berekeningen van de beslisboom homomorfisch uitvoeren. De moeilijkheid hierbij is dat voor het oplossen van onze probleemstelling, nu alle stukjes data versleuteld worden door verschillende partijen. Homomorfische berekeningen, zoals hierboven omschreven, kunnen alleen gedaan worden met data die versleuteld is met dezelfde sleutel.

## Drie protocollen

In ons werk zijn drie protocollen voorgesteld om dit wél mogelijk te maken. Deze protocollen maken gebruik van de techniek genaamd `Multi-Key Fully Homomorphic Encryption' of de techniek `Fully Homomorphic Encryption' (Peikert & Shiehian, 2016). Deze eerste techniek maakt het mogelijk om homomorfische berekeningen te doen op data die is versleuteld met verschillende sleutels.
De tweede techniek maakt het mogelijk deze data te combineren door een extra partij te introduceren van wie de encryptiesleutel wordt gebruikt, maar verder geen kennis vergaart wat betreft de input data of de beslisboom.
In het derde protocol wordt een zogenaamde `sleutelwisseling' voorgesteld die ervoor zorgt dat de protocollen minder afhankelijk zijn van deze extra partij.

## Conclusie

Alle protocollen zijn geïmplementeerd en met elkaar vergeleken voor wat betreft de complexiteit, runtijd en benodigde communicatie tussen de verschillende partijen. Daaruit kwam naar voren dat de techniek `Multi-Key Fully Homomorphic Encryption' erg complex is, wat resulteert in te hoge, en daarom niet praktische runtijden. De andere twee protocollen zijn daarom het meest haalbaar. Onze implementatie, als we aannemen dat de computaties in parallel gedaan kunnen worden, gaf een hoogst haalbare runtijd in de ordegrootte van dagen. Gelukkig hangt de efficiëntie van onze protocollen direct af van de efficiëntie van de onderliggende encryptieschema's, dus verbetering is niet uitgesloten. Met dit werk is de eerste stap gezet naar mogelijke oplossingen om in een collaboratieve setting een beslisboom privé te evalueren, waarvoor er vele interessante toepassingsgebieden bestaan.

Ben je na het lezen van bovenstaand stuk nieuwsgierig geworden naar de inhoud? Via deze link is de scriptie te downloaden: http://resolver.tudelft.nl/uuid:50073f62-cf87-40d1-bef3-e407b5a5b949.

## Referenties


Gibert, D., Mateu, C., & Planes, J. (2020, March 1). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. Journal of Network and Computer Applications, 153. 10.1016/j.jnca.2019.102526

Peikert, C., & Shiehian, S. (2016, October 21). Multi-key FHE from LWE, Revisited. Theory of Cryptography Conference, 217-238. 10.1007/978-3-662-53644-5_9

Pieterman, H. (2020, December 10). Geef huisarts spilfunctie in epd. Medisch Contact. https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/geef-huisarts-spilfunctie-in-epd.htm

Portugal, I., Alencar, P., & Cowan, D. (2018, May 1). The use of machine learning algorithms in recommender systems: A systematic review. Expert Systems with Applications, 97, 205-227. 10.1016/j.eswa.2017.12.020

Sangers, A., van Heesch, M., Attema, T., & Veugen, T. (2019). Secure Multiparty PageRank Algorithm for Collaborative Fraud Detection. Financial Cryptography and Data Security, 605-623.

Tai, R. K. H., Ma, J. P. K., Zhao, Y., & Chow, S. S. M. (2017, August 12). Privacy-Preserving Decision Trees Evaluation via Linear Functions. European Symposium on Research in Computer Security, 494-512. 10.1007/978-3-319-66399-9_27

Tueno, A., Boev, Y., & Kerschbaum, F. (2020, June 18). Non-interactive Private Decision Tree Evaluation. IFIP Annual Conference on Data and Applications Security and Privacy, XXXIV, 174-194. 10.1007/978-3-030-49669-2_10