(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

COURBEVOIE

(11) **N° de publication :** ***3 143 918***

(à n'utiliser que pour les commandes de reproduction)

(21) **N° d'enregistrement national :** **22 13324**

(51) Int Cl[8] : ***H 04 L 43/04*** *(2023.01), H 04 L 43/02, G 06 F 21/00, 21/55, G 06 N 3/02*

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt :** 14.12.22.

(30) **Priorité :**

(43) **Date de mise à la disposition du public de la demande :** 21.06.24 Bulletin 24/25.

(56) **Liste des documents cités dans le rapport de recherche préliminaire :** *Se reporter à la fin du présent fascicule*

(60) **Références à d'autres documents nationaux apparentés :**

( ) **Demande(s) d'extension :**

(71) **Demandeur(s) :** *COMMISSARIAT A L'ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES Etablissement public* — FR.

(72) **Inventeur(s) :** FRIJI Hamdi, OLIVEREAU Alexis et JANNETEAU Christophe.

(73) **Titulaire(s) :** COMMISSARIAT A L'ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES Etablissement public.

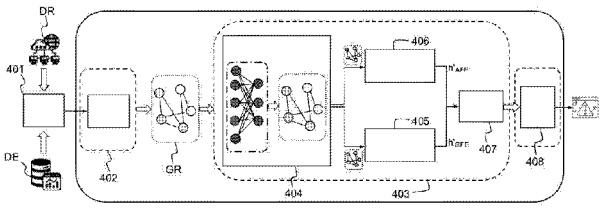
(74) **Mandataire(s) :** ATOUT PI LAPLACE.

(54) Méthode et système de détection d'intrusions dans un réseau informatique par apprentissage automatique.

(57) Méthode de détection de flux malveillants dans un réseau informatique comprenant les étapes de :

Capturer un ensemble de flux réseau, Créer un graphe dont les nœuds représentent chaque flux capturé, deux nœuds du graphe étant connectés si les deux flux associés sont émis par le même utilisateur ou sont transmis vers le même utilisateur, Exécuter un moteur d'intelligence artificielle entrainé pour détecter une intrusion dans le réseau informatique à partir des graphes et comprenant les sous-étapes de :Appliquer aux graphes au moins un réseau neuronal de graphes entrainé pour générer, pour chaque flux, un second vecteur de caractéristiques aptes à discriminer un flux autorisé d'un flux malveillant, l'entrainement dudit réseau neuronal de graphes exploitant au moins la topologie des graphes,Appliquer aux seconds vecteurs de caractéristiques un réseau de neurones artificiel classifieur entrainé pour classifier les flux parmi deux classes correspondant respectivement aux flux autorisés et aux flux malveillants.

Figure 4



FR 3 143 918 - A1

Description

Titre de l'invention : Méthode et système de détection d'intrusions dans un réseau informatique par apprentissage automatique

[0001] L'invention concerne le domaine de la cyber-sécurité des réseaux informatiques ou de télécommunications.

[0002] L'invention porte sur une méthode et un système de détection d'intrusions dans un réseau informatique au moyen d'une méthode de détection implémentant un modèle d'apprentissage automatique exploitant une représentation particulière des données sous forme de graphe.

[0003] Les systèmes de détection d'intrusion (IDS) (« Intrusion Detection Systems ») ont pour fonction la détection d'intrusion ou plus généralement la détection d'anomalies dans un réseau. Leur objectif est d'agir en complément de mécanismes de sécurité préventifs de type cryptographique ou contrôle d'accès. Il existe principalement deux types de systèmes IDS selon la méthode de détection d'attaque mise en œuvre. Deux types de méthodes de détection d'intrusion existent principalement, les méthodes de détection d'intrusion par signatures ou par anomalies.

[0004] La détection par signatures repose sur une base de signatures d'attaques. Les différents éléments observés par le système IDS (par exemple, des flux réseau dans le cas d'un système de détection d'intrusion sur réseau) sont comparés aux différentes entrées de cette base. Une correspondance signifie que le système sous surveillance est la cible de l'attaque dont la signature a correspondu à l'observation. Ces systèmes visent à détecter les attaques en comparant le trafic réseau à des modèles prédéfinis d'attaques déjà connus.

[0005] Les systèmes de détection par anomalies reposent quant à eux sur la détection d'une opération anormale, par rapport aux observations antérieures, dans l'élément observé. Ces systèmes présentent l'avantage de permettre la détection d'attaques inconnues, jamais observées auparavant et pour lesquelles des signatures n'ont pas encore pu être constituées. Ils sont cependant également caractérisés par des taux de faux positifs élevés pour la détection des attaques inconnues. On parle de faux positif quand un événement qui n'est pas une intrusion est identifié comme étant une anomalie. Ce cas arrive, par exemple, lors d'une défaillance du réseau qui provoque un phénomène anormal qui peut être détecté comme étant une intrusion.

[0006] Les systèmes IDS basés sur les signatures sont efficaces pour détecter les attaques connues avec un faible taux de faux positifs, mais ne parviennent pas à détecter les nouveaux types d'attaques. Par conséquent, c'est pour cela, au cours des dernières années, que des concepts de détection d'anomalie utilisant l'apprentissage automatique

(Machine Learning ML en anglais) et l'apprentissage profond (Deep Learning DL en anglais) ont été investigués pour identifier tout comportement anormal dans le réseau. Ces techniques montrent des résultats encourageants dans la détection des nouvelles attaques.

[0007] La demande de brevet américaine US20220021695 propose un procédé de détection d'intrusion adaptatif qui se base sur le déploiement d'un système de capture de trafic réseau et de collecte de traces de paquets réseau qui est utilisé par un outil d'audit de réseau afin d'extraire des caractéristiques des paquets réseau collectés. Ensuite, les caractéristiques extraites sont considérées comme des données non étiquetées et combinées à un ensemble d'apprentissage étiqueté capturant des exemples de trafic réseau malveillant à utiliser avec la nouvelle représentation apprise des données non étiquetées pour modifier l'ensemble d'apprentissage étiqueté et obtenir un nouvel ensemble d'apprentissage. Les nouvelles données d'apprentissage ainsi obtenues sont utilisées pour former un modèle d'apprentissage automatique de classification du trafic. Un inconvénient de cette approche est qu'elle ne considère que les données extraites des paquets pour classer le trafic et en particulier ne prend pas en compte la topologie du réseau, en particulier les informations spatiales liées aux attaques.

[0008] Le brevet américain US7260846 utilise un analyseur de réseau pour collecter des paquets de réseau et en extraire des données individuelles, ainsi qu'un enregistreur de trafic. Ensuite, un constructeur de vecteurs est configuré pour générer des vecteurs multidimensionnels à partir des caractéristiques collectées des champs des entêtes de paquets. Des modules de regroupement auto-organisés sont configurés pour traiter les vecteurs multidimensionnels afin de produire une carte auto-organisée de regroupements. Ensuite, un détecteur d'anomalies peut détecter des corrélations anormales entre des clusters individuels de la carte auto-organisée sur la base d'au moins une métrique de corrélation configurable. Enfin, un classifieur classe le trafic selon qu'il est malveillant ou normal en utilisant les corrélations anormales calculées.

[0009] La demande de brevet américaine US20200137083 concerne un procédé de détection de comportement de programme malveillant qui comprend l'exécution d'une vérification de programme basée sur des données d'activité système. Une analyse de données de programme non vérifiées identifiées à partir de la vérification de programme est réalisée pour détecter des événements anormaux, y compris l'analyse d'événements au niveau de l'hôte pour détecter des événements anormaux. L'analyse se base sur l'apprentissage d'une représentation de programme sous forme de graphe s'encastrant dans une architecture attentionnelle basée sur un graphe invariant entre différentes entités système, générant des résultats de détection basés sur l'analyse, et exécutant au moins une action corrective basée sur les résultats de détection.

[0010] La demande internationale WO202189196 concerne un procédé de détection

d'intrusion pour détecter des activités de menaces malveillantes dans un réseau composé de plusieurs profils d'utilisateurs interconnectés en apprenant un modèle de comportement pour chaque profil d'utilisateur sur la base de plusieurs événements d'activité, dans lequel la détermination du modèle de comportement est exécutée par un réseau de neurones récurrent. La différenciation entre les événements d'activité normaux et anormaux pour les utilisateurs cachés est effectuée par un réseau neuronal de type « Feed-Forward » entraîné en utilisant plusieurs modèles de comportement. Un inconvénient de cette méthode est qu'elle considère uniquement les données de comportement des utilisateurs dans la procédure d'apprentissage et ne considère aucune relation de l'utilisateur avec d'autres utilisateurs dans le réseau.

[0011] La publication scientifique "Unveilling the potential of graph neural networks for robust intrusion detection, David Pujol-Perich et al" est basée sur l'emploi de plusieurs algorithmes d'intelligence artificielle appliqués à des graphes ou « Graph Neural Network (GNN) » pour développer des systèmes de détection d'intrusions réseau (NIDS). Ce document propose une représentation à base d'un graphe qui représente les propriétés des flux de communication et leurs relations dans le réseau combiné à une nouvelle architecture de réseau de neurone GNN spécialement conçue pour apprendre et généraliser les informations précédentes structurées en graphes. Les auteurs créent un graphe de connexion entre les hôtes dont chaque nœud correspond à un utilisateur du réseau. Ils proposent d'abord de créer un hétérographe généreux où ils introduisent deux types de nœuds; le premier type représente les utilisateurs, et le second représente les flux.

[0012] De façon générale, il existe un besoin pour améliorer les méthodes de l'art antérieur afin de mieux mettre en évidence les motifs caractéristiques d'une intrusion dans un réseau.

[0013] L'invention propose une nouvelle méthode basée sur un type de graphe particulier dont la structure est créée de manière à donner des informations spatiales plus pertinentes sur les motifs des intrusions. En particulier les nœuds du graphe sont associés à des flux et non plus à des utilisateurs.

[0014] Cette nouvelle structure de graphe basée sur les flux permet au modèle d'apprentissage de faire la distinction entre les flux normaux et malveillants, y compris les flux générés par des attaquants camouflés (c'est-à-dire des attaquants essayant de dissimuler leurs activités malveillantes pour échapper à la détection). Tout d'abord, la première phase de la méthode proposée consiste à transformer le trafic capturé sur le réseau en un graphe puis des caractéristiques spatiales et non spatiales sont extraites du graphe par un ensemble de réseaux de neurones agissant en tant que modèle d'apprentissage machine.

[0015] Les avantages techniques de l'invention sont induits par les informations pertinentes

qui peuvent être extraites de la structure. La présente invention propose de modéliser les flux de communication sous la forme d'une nouvelle structure de graphe. Cette structure permet d'extraire les informations topologiques pertinentes permettant à un modèle d'intelligence artificielle d'apprendre une généralisation des motifs caractéristiques d'attaques lors d'une phase d'apprentissage et de reconnaître un comportement anormal tout en suivant la procédure de détection. La méthode proposée modélise les communications dans le réseau sous la forme d'un graphe où chaque nœud représente un flux. Les liens entre nœuds sont créés pour relier les flux provenant du même utilisateur ou dirigés vers le même utilisateur.

[0016] Un avantage d'une telle structure de graphe est que les nœuds du graphe correspondent à des flux de communication et non aux utilisateurs du réseau. En effet, la détection d'intrusion basée uniquement sur des informations liées aux utilisateurs peut facilement être contournée par usurpation d'identité. Les attaques dans un réseau ont un modèle comportemental qui modélise les étapes d'attaque effectuées par l'attaquant. Ces modèles sont les informations que le spécialiste de la cybersécurité recherche pour détecter une attaque. La structure de graphe propose que deux communications appartiennent au même schéma d'attaque si elles sont orientées vers ou générées par le même utilisateur. En d'autres termes, deux flux générés par le même utilisateur ou destinés au même utilisateur sont connectés dans la structure du graphe pour donner aux modèles d'intelligence artificielle la capacité d'extraire les communications qui pourraient appartenir à la même attaque. Ces informations peuvent être combinées avec les attributs des communications pour différencier les communications normales et les comportements malveillants.

[0017] La plupart des solutions de l'état de l'art représentent le réseau sous la forme d'un graphe où les nœuds sont les utilisateurs (ou machines / hôtes du réseau), chacun identifié avec une adresse IP seule ou combinée avec un numéro de port comme second identifiant. Cependant, cette représentation est trop simple et présente plusieurs inconvénients. Premièrement, en utilisant cette structure, le problème est transformé en une tâche de classification des arêtes du graphe, mais la détection d'intrusion consiste à capturer les flux malveillants et non les attaquants. En effet, détecter les utilisateurs avec un comportement malveillant est plus difficile que détecter les flux malveillants eux-mêmes et il peut être éludé facilement. Pour cela, les attaquants modifient généralement leurs adresses IP en utilisant des réseaux privés virtuels. De plus, la tâche de classification des arêtes n'est pas encore bien étudiée à l'aide de la théorie des réseaux de neurones basés sur des graphes GNN, et par conséquent, les résultats de l'état de l'art ne sont pas encourageants.

[0018] D'autre part, la présente invention est plus simple que la solution proposée par David Pujol-Perich et al. En effet, cette solution de l'état de l'art, précédemment introduite,

est caractérisée par un nœud pour chaque utilisateur de réseau. Au contraire, l'invention est basée sur un graphe dont les nœuds sont uniquement associés aux flux de communication. Par conséquent, l'invention est plus efficace en terme de consommation de ressources (utilisation de la mémoire et consommation énergétique/ électrique réduite).

[0019] L'invention exploite les graphes créés dans un premier temps au moyen d'un premier modèle d'intelligence artificielle entrainé pour déterminer une représentation discriminante des caractéristiques des flux de communication de sorte à maximiser une dissimilarité entre un flux légitime et un flux malveillant. Ensuite, un ou plusieurs modèles de réseaux neuronaux de graphes sont appliqués pour exploiter la topologie du graphe afin d'extraire des caractéristiques supplémentaires qui sont finalement transmises à module de décision entrainé pour classifier les flux entre flux légitimes et malveillants.

[0020] L'invention a pour objet une méthode, mise en œuvre par ordinateur, de détection de flux malveillants dans un réseau informatique comprenant les étapes de :

- Capturer un ensemble de flux réseau, chaque flux réseau étant associé à un premier vecteur de caractéristiques du flux mesurées et à un instant de mesure,
- Créer, pour chaque instant de mesure, un graphe dont les nœuds représentent chaque flux capturé, deux nœuds du graphe étant connectés si les deux flux associés sont émis par le même utilisateur ou sont transmis vers le même utilisateur, chaque nœud étant pourvu dudit premier vecteur de caractéristiques,
- Exécuter un moteur d'intelligence artificielle entrainé pour détecter une intrusion dans le réseau informatique à partir des graphes et comprenant les sous-étapes de :
  i. Appliquer aux graphes au moins un réseau neuronal de graphes entrainé pour générer, pour chaque flux, un second vecteur de caractéristiques aptes à discriminer un flux autorisé d'un flux malveillant, l'entrainement dudit réseau neuronal de graphes exploitant au moins la topologie des graphes,
  ii. Appliquer aux seconds vecteurs de caractéristiques un réseau de neurones artificiel classifieur entrainé pour classifier les flux parmi deux classes correspondant respectivement aux flux autorisés et aux flux malveillants.

[0021] Selon un aspect particulier de l'invention, l'étape d'exécuter le moteur d'intelligence artificielle comprend en outre une étape, préalable à l'application de l'au moins un réseau neuronal de graphes, consistant à appliquer aux premiers vecteurs de caractéristiques au moins un réseau neuronal artificiel entrainé pour réduire la dimension des premiers vecteurs par extraction de caractéristiques discriminants un flux autorisé d'un

flux malveillant, l'entrainement dudit réseau neuronal artificiel n'exploitant pas la topologie des graphes.

[0022] Selon un aspect particulier de l'invention, l'au moins un réseau neuronal artificiel est un perceptron multi-couches.

[0023] Dans une variante de réalisation, la méthode comprend une étape de prétraitement des premiers vecteurs de caractéristiques mesurées consistant au moins en une normalisation des valeurs et/ou un encodage catégoriel des valeurs.

[0024] Dans une variante de réalisation, la méthode comprend en outre une étape de calcul, pour chaque arête du graphe, d'un score de similarité entre les premiers vecteurs de caractéristiques des deux flux associés aux deux nœuds connectés par l'arête.

[0025] Dans une variante de réalisation, le score de similarité est calculé au moyen d'une mesure de similarité en cosinus.

[0026] Dans une variante de réalisation, l'au moins un réseau neuronal de graphes est un réseau neuronal convolutionnel de graphes.

[0027] Dans une variante de réalisation, le réseau neuronal convolutionnel de graphes comporte plusieurs couches configurées pour mettre en œuvre un calcul de convolution du type :

$$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{\frac{1}{2}}H^{(l)}W^{(l)}\right),$$

$H^{(l)}$, $H^{(l+1)}$ sont les sorties respectives des couches d'indices l, l+1

$\hat{A} = A + I$ avec A une matrice d'adjacence dont les entrées (i,j) valent 0 si les nœuds i et j du graphe ne sont pas connectés et valent le score de similarité si les nœuds i et j sont connectés,

$\hat{D}$ est une matrice diagonale dont les valeurs sont égales aux degrés de chaque nœud,

$W^{(l)}$ correspond aux paramètres de la couche d'indice l qui sont appris,

σ est une fonction d'activation, par exemple une fonction ReLu

[0028] Selon un aspect particulier de l'invention, l'au moins un réseau neuronal de graphes est un réseau d'attention de graphes.

[0029] Selon un aspect particulier de l'invention, le réseau d'attention de graphes comporte un réseau neuronal de graphes et une couche d'attention entrainée au moyen d'un modèle d'attention configuré pour apprendre des valeurs d'attention pour chaque arête du graphe et calculer une caractéristique pour chaque nœud à l'aide d'une pondération des caractéristiques des nœuds voisins en fonction des valeurs d'attention.

[0030] Selon un aspect particulier de l'invention, plusieurs réseaux neuronaux de graphe sont exécutés en parallèle et la méthode comporte une étape de concaténation des seconds vecteurs de caractéristiques générés par chaque réseau neuronal de graphe pour fournir un vecteur concaténé au réseau de neurones artificiel classifieur.

[0031] Selon un aspect particulier de l'invention, le réseau de neurones artificiel classifieur est un perceptron multi-couches.

[0032] Selon un aspect particulier de l'invention, le premier vecteur de caractéristiques mesurées contient des informations caractéristiques des flux parmi : les adresses sources et destination, les protocoles utilisés, des statistiques relatives aux paquets transmis via ce flux.

[0033] Selon un aspect particulier de l'invention, le moteur d'intelligence artificielle est entrainé de manière supervisée ou non supervisée.

[0034] L'invention a aussi pour objet un système de détection de flux malveillants dans un réseau de télécommunications comprenant au moins une sonde apte à acquérir un flux réseau ou un ensemble de trames réseau transitant dans ledit réseau et au moins un dispositif de détection de flux malveillants configuré pour mettre en œuvre les étapes de la méthode de détection flux malveillants selon l'invention, à partir du flux réseau ou de l'ensemble de trames réseau.

[0035] D'autres caractéristiques et avantages de la présente invention apparaîtront mieux à la lecture de la description qui suit en relation aux dessins annexés suivants.

[0036] [Fig.1] représente un schéma général d'un dispositif de détection d'intrusions dans un réseau informatique,

[0037] [Fig.2] représente un schéma d'une structure de graphe modélisant un ensemble de flux réseaux selon un mode de réalisation de l'invention,

[0038] [Fig.3] représente un schéma de plusieurs structures de graphes modélisant une intrusion dans un réseau,

[0039] [Fig.4] représente un schéma d'une méthode de détection d'intrusions selon un mode de réalisation de l'invention,

[0040] La [Fig.1] représente un exemple d'implémentation d'un système de détection d'intrusion NIDS selon un mode de réalisation de l'invention. Un tel système NIDS agit au niveau réseau pour classifier les flux provenant d'un réseau informatique en flux légitimes ou malveillants. Les flux malveillants sont reportés vers un administrateur réseau AR qui est responsable de l'analyse des flux malveillants et la mise en œuvre d'actions permettant de prévenir des attaques potentielles. Les flux classés comme malveillants par le système de détection NIDS sont sauvegardés dans une base de données spécifique BD.

[0041] Le système de détection NIDS collecte les flux transitant dans le réseau informatique, via l'intermédiaire d'un pare-feu PF. Les flux peuvent provenir d'un réseau local RL ou d'un réseau externe RE, par exemple le réseau Internet.

[0042] Le système de détection NIDS est configuré pour exécuter une méthode de détection d'intrusions selon un mode de réalisation de l'invention. A cet effet, il est apte à capturer les paquets réseaux transmis au sein du réseau et les transformer en flux de

communications identifiés par différentes caractéristiques parmi lesquelles le type de protocole (IP), le numéro de port source/destination, l'adresse IP source/destination, un horodatage (timestamp) ainsi que toute autre information statistique extraite des entêtes des paquets ou calculées à partir d'informations extraites des entêtes des paquets.

[0043] La [Fig.2] représente, sur deux schémas comparatifs, une structure de graphe selon l'art antérieur et une structure de graphe selon l'invention appliquée à un réseau informatique.

[0044] Sur la gauche de la [Fig.2], on a représenté un premier exemple de graphe GR-1 selon une représentation utilisée dans l'art antérieur. Le graphe GR-1 est structuré de telle sorte que les nœuds IP1, IP2, IP3, IP4 du graphe correspondent à des utilisateurs du réseau, par exemple identifiés par leur adresse IP, et les arêtes du réseau F1, F2, F3, F4, F5 correspondent à des flux de communication entre un nœud source et un nœud destination. Par exemple, le flux F1 est entre la source IP2 et la destination IP1.

[0045] Sur la droite de la [Fig.2], on a représenté, pour le même exemple de réseau informatique utilisé pour le graphe GR-1, un exemple de structure de graphe GR-2 proposée par l'invention. Le graphe GR-2 est structuré de sorte que les nœuds du graphe correspondent cette fois aux flux de communication F1, F2, F3, F4, F5. Deux nœuds du graphe GR-2 sont connectés par une arête si les flux associés proviennent de la même source ou sont destinés à la même destination.

[0046] Dans l'exemple de la [Fig.2], les flux F1, F2 et F4 sont connectés ensemble car ils partagent la même source IP2. Les flux F2 et F3 sont connectés car ils partagent la même destination IP3. Les flux F4 et F5 sont connectés car ils partagent la même destination IP4.

[0047] La structure du graphe GR-2 présente l'avantage d'une plus grande efficacité en termes d'informations fournies par rapport à la structure de l'art antérieur GR-1. Le graphe GR-2 permet de fournir à des modèles d'intelligence artificielle basés sur des réseaux de graphes neuronaux des informations structurelles pertinentes afin de mieux distinguer les flux de communications malveillants des flux autorisés.

[0048] Par exemple, dans le cas d'une attaque de reconnaissance (ou « footprinting attack » en anglais) ou d'une attaque DDoS par déni de service distribué, on obtient une structure de graphe où tous les flux malveillants sont interconnectés. Les modèles de réseaux de graphes neuronaux peuvent apprendre rapidement cette interconnexion, améliorant ainsi la détection des flux malveillants.

[0049] En outre, dans un mode de réalisation particulier de l'invention, on attribue à chaque arête du graphe GR-2 un score de similarité entre les deux nœuds connectés par l'arête. Le score de similarité $\zeta_{i,j}$ est un indicateur de la similarité entre les deux flux F_i et F_j. Chaque arête du graphe GR-2 est ainsi munie d'un poids qui permet d'améliorer la re-

connaissance des formes des modèles basés sur les réseaux de graphes neuronaux en fournissant des informations sur les deux nœuds connectés par cette arête.

[0050] En effet, les flux de communications impliqués dans un trafic d'attaque présentent en général une certaine similitude, par exemple lorsque l'attaque correspond à un sondage, une charge utile DDoS ou à une attaque unique qui se propage. Le score de similitude entre deux flux peut notamment augmenter si deux flux partagent une même source ou une même destination, si deux flux sont proches du point de vue de leurs spécificités par exemple en termes de durée, volume, ou motif de trafic. Le score de similitude peut également augmenter de manière inversement proportionnelle au temps écoulé entre l'apparition de ces deux flux.

[0051] Le score de similarité est par exemple un score de similarité en cosinus défini par la relation :

[0052]
$$c(d_i, d_j) = \frac{d_i^T d_j}{\|d_i\| \|d_j\|} = \frac{\sum_{k=1}^{N_f} v_k^i v_k^j}{\sqrt{\sum_{k=1}^{N_f} (v_k^i)^2} \sqrt{\sum_{k=1}^{N_f} (v_k^j)^2}}, \quad (1)$$

[0053] d_i,d_j sont des vecteurs contenant les caractéristiques v_k^i, v_k^j mesurées pour les flux de communication F_i,F_j, N_f .

[0054] La [Fig.3] schématise plusieurs exemples comparatifs de graphes selon l'art antérieur et selon l'invention, appliqués à des scénarios d'attaques. Les graphes selon la méthode de l'art antérieur GR-1-1, GR-1-2, GR-1-3, GR-1-4 sont représentés sur la gauche de la figure tandis que les graphes selon l'invention GR-2-1, GR-2-2, GR-2-3, GR-2-4 sont représentés sur la droite de la figure.

[0055] Les scénarios d'attaques considérés sur l'exemple de la [Fig.3] correspondent à la séquence temporelle suivante :

- A l'instant t_1, l'attaquant IP0 collecte des informations simultanément sur plusieurs équipements du réseau IP1, IP2, IP3, IP4, il génère ainsi des flux F1, F2, F3, F4. Cette étape est représentée par les graphes GR-1-1 et GR-2-1 ;
- A l'instant t_2, l'attaquant met en œuvre une injection SQL pour exploiter une faille de sécurité dans l'un des équipements IP1. Pour cela, l'attaquant IP0 génère les flux F_1^1, F_1^2, F_1^3 vers la cible IP1. Cette étape est représentée par les graphes GR-1-2 et GR-2-2 ;
- A l'instant t_3, l'attaquant tente de craquer un mot de passe de l'utilisateur IP4 pour obtenir un accès au réseau. Pour cela, il génère les flux F_4^1, F_4^2, F_4^3 , F_4^4, F_4^5 vers la cible IP4. Cette étape est représentée par les graphes GR-1-3 et GR-2-3 ;
- A l'instant t_4, l'attaquant usurpe l'identité de IP4 après cinq tentatives de connexion. Pour cela, il génère les flux F'_4^1, F'_4^2, F'_4^3, F'_4^4, F'_4^5 vers la cible

IP4. Cette étape est représentée par les graphes GR-1-4 et GR-2-4.

[0056] Par souci de simplification, toutes les arêtes n'ont pas été représentées sur la [Fig.3], en particulier les arêtes reliant les nœuds du graphe GR-2-3 aux flux F1, F2, F3.

[0057] Les représentations des graphes GR-2-1, GR-2-2, GR-2-3, GR-2-4 selon l'invention permettent de faire ressortir certains motifs particuliers dans la structure des graphes qui sont caractéristiques des attaques précitées. Egalement, on peut remarquer qu'il existe un lien entre le graphe GR-2-1 qui représente l'étape de collecte d'informations et les autres graphes GR-2-2, GR-2-3, GR-2-4 qui correspondent aux étapes liées aux attaques.

[0058] L'invention met en œuvre un moteur d'intelligence artificielle particulier qui exploite la structure particulière du graphe proposé pour apprendre à reconnaitre des motifs caractéristiques de comportements associés à des attaques ou plus généralement à des flux malveillants et à les différencier de flux autorisés.

[0059] La [Fig.4] schématise la mise en œuvre de la méthode de détection de flux malveillants selon un mode de réalisation de l'invention.

[0060] La méthode proposée met en œuvre un modèle d'intelligence artificielle entrainé pour classifier les flux capturés dans un réseau informatique en deux catégories : flux malveillants ou flux autorisés. Ce modèle peut être entrainé de façon supervisée ou non-supervisée.

[0061] A cet effet, la méthode débute par une étape de capture de données réseau DR. Les flux réseau sont capturés en collectant les trames réseau puis en identifiant chaque flux à partir de son adresse source et son adresse destination et en collectant un ensemble de caractéristiques relatives au flux parmi lesquelles des caractéristiques relatives au protocole utilisé, comme le type et la version du protocole (4 ou 6), la longueur de l'en-tête IP, le type de service, la longueur totale du paquet ou datagramme, le numéro d'identification, les indicateurs ou flags, le fragment offset, la durée de vie, le numéro du protocole, le champ somme de contrôle (checksum en anglais), les adresses IP source et destination ou encore d'autres options du protocole qui peuvent être extraites de l'entête IP. Les caractéristiques peuvent aussi être déterminées à partir d'informations prélevées dans la trame. Ces caractéristiques concernent, par exemple, des informations d'horodatage ou de temps entre deux trames consécutives. Elles peuvent concerner les activités relatives à différents protocoles, SSL, HTTP, DNS.

[0062] Plus généralement, d'autres caractéristiques peuvent également être obtenues à partir de l'analyse d'un flux ou d'une connexion, sans nécessité d'analyser chaque trame réseau. En effet, un analyseur ou une sonde réseau présente en général cette capacité d'analyse et de restitution de caractéristiques associées à un flux ou une connexion. Des caractéristiques statistiques peuvent également être considérées, par exemple des statistiques relatives à la distribution de certains paramètres d'un flux comme la

longueur des paquets, la durée d'un flux, le nombre de paquets dans un flux, la durée entre deux paquets, le nombre d'occurrences d'un indicateur (flag) dans une trame d'un flux, la taille d'un entête, le nombre de paquets par seconde, cette liste n'étant pas exhaustive. Des opérations statistiques peuvent être appliquées à chacune de ces caractéristiques, par exemple la moyenne, médiane, minimum, maximum ou encore la déviation standard.

[0063] A l'étape 401, on effectue un prétraitement des données capturées, c'est-à-dire des caractéristiques mesurées pour chaque flux afin de mettre en forme ces données en vue de leur traitement par un moteur d'apprentissage automatique. La phase de prétraitement peut comporter une étape de normalisation des données afin de les ramener à une échelle de valeurs communes. Elle peut aussi comporter une étape d'encodage des données non numériques, par exemple à l'aide d'une méthode d'encodage « one-hot » afin de fournir en entrée des algorithmes d'intelligence artificielles des données uniquement numériques.

[0064] La phase de prétraitement 401 est appliquée à la fois aux données collectées sur le réseau à analyser et à d'éventuelles données d'entrainement DE dans le cas où on utilise des méthodes d'entrainement supervisé. Dans ce second cas, les données d'entrainement sont constituées de flux de communications autorisés et illicites qui sont stockés dans une base de données qui peut, par exemple, être alimentée par la méthode de classification selon l'invention elle-même. Dans le cas d'un entrainement supervisé, la phase de prétraitement 401 peut comporter une étape de sous-échantillonnage des données d'entrainement afin de limiter la proportion des données correspondant à des flux autorisés par rapport à la proportion des données correspondant à des flux malveillants.

[0065] L'étape 402 consiste ensuite en la création d'un graphe à partir des flux capturés. Un graphe est généré pour chaque instant temporel de mesure.

[0066] En sortie de l'étape 402, on obtient donc une séquence temporelle de graphes, chaque graphe GR étant composé de nœuds représentant les flux de communications capturés et d'arêtes. Les nœuds du réseau sont associés à des vecteurs de caractéristiques de chaque flux. Dans un mode de réalisation de l'invention, les arêtes du réseau sont associées à des scores de similarité.

[0067] Ensuite, on applique à cette séquence temporelle de graphes un moteur d'intelligence artificielle 403 entrainé pour extraire, pour chaque flux de communication, un vecteur de caractéristiques pertinentes qui permettent de distinguer un flux malveillant d'un flux autorisé.

[0068] Le moteur d'intelligence artificielle 403 se compose de deux étapes. La première étape 404 consiste en l'application d'un réseau de neurones artificiels entrainé pour réduire la dimension des vecteurs de caractéristiques initiales et fournir en sortie, pour

chaque flux, un vecteur de caractéristiques de dimension réduite et qui soit plus représentatives des motifs caractéristiques des flux malveillants ou autorisés. Une particularité de cette première étape 404 est qu'elle n'exploite pas d'information relative à la structure du graphe, c'est-à-dire la topologie des nœuds et des arêtes, mais uniquement les valeurs des caractéristiques associées à chaque nœud.

[0069] Le réseau de neurones artificiels mis en œuvre à l'étape 404 est, par exemple, un perceptron multi-couches entrainé pour générer des caractéristiques ou plongements (de l'anglais « embedding ») discriminatifs. Le modèle de réseau de neurones est entrainé pour apprendre à réaliser une correspondance entre l'espace des données initiales (de dimension égale au nombre de caractéristiques initiales) vers un espace plus petit et plus pertinent. Autrement dit, le réseau de neurones de l'étape 404 est entrainé pour apprendre à générer une représentation discriminante entre les données extraites des flux de communications légitimes et malveillants. Cette première étape 404 permet d'une part de réduire la taille des données à traiter et de renforcer les différences entre flux légitimes et malveillants.

[0070] Le premier réseau de neurones artificiels 404 est suivi d'au moins un second réseau neuronal de graphes 405,406. Un réseau neuronal de graphes ou « Graph Neural Network » désigne une catégorie de techniques d'intelligence artificielle qui sont prévues pour traiter des données sous forme de graphes structurés. Un modèle de réseau neuronal de graphes contient en général plusieurs couches qui interfacent entre elles par une représentation mise à jour d'un graphe. Chaque couche du réseau produit en sortie un vecteur de caractéristiques pour chaque nœud du graphe qui est ensuite utilisé comme entrée de la couche suivante. Un objectif du modèle que vise à représenter un réseau neuronal de graphes est de construire des plongements pour chaque nœud du graphe en intégrant à la fois les caractéristiques initiales de chaque nœud et des informations topologiques sur le voisinage du nœud dans le graphe.

[0071] Un réseau neuronal de graphes implémente plusieurs types de tâches, parmi lesquelles des tâches exécutées au niveau des nœuds et d'autres tâches exécutées au niveau des arêtes et d'autres exécutées au niveau de l'ensemble du graphe. Un principe du fonctionnement d'un réseau neuronal de graphes est qu'il met à jour chaque nœud du graphe en agrégeant des caractéristiques des nœuds voisins de façon itérative. Un réseau neuronal de graphes présente l'avantage de permettre l'extraction d'informations structurelles et de détecter des motifs dans la topologie des graphes qui vont au-delà des caractéristiques associées à chaque nœud. Ainsi, un réseau neuronal de graphes exploite la topologie des graphes pour son apprentissage à la différence du réseau de neurones artificiels utilisé à l'étape 404 qui n'exploite pas cette topologie.

[0072] Un ou plusieurs réseaux neuronaux de graphes, de types différents, peuvent être appliqués en sortie de l'étape 404.

[0073] Dans un mode de réalisation de l'invention, un premier réseau neuronal de graphes convolutionnel ou « Graph convolutional network » GCN en anglais est appliqué à l'étape 405. Le réseau neuronal de graphes convolutionnel 405 a pour fonction d'extraire des caractéristiques discriminantes mais en exploitant la topologie du graphe, c'est-à-dire les informations spatiales, par exemple les relations entre nœuds voisins.

[0074] Un exemple de réseau neuronal de graphes convolutionnel ou réseau GCN est décrit dans la publication « Convolutional Neural Networks on Graphs with fast localized spectral filtering », M. Defferrard et al. 30th Conference on Neural Information Processing Systems (NIPS 2016).

[0075] Le réseau GCN met en œuvre des opérations de convolution sur les graphes fournit en entrée et met à jour itérativement les valeurs des caractéristiques attachées à chaque nœud en agrégeant les caractéristiques des nœuds voisins et en prenant en compte les scores de similarités des arêtes entre nœuds voisins.

[0076] L'opération de convolution mise en œuvre par le réseau GCN peut être définie par la relation suivante :

[0077] $$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{\frac{1}{2}}H^{(l)}W^{(l)}\right), \quad (2)$$

[0078] $H^{(l)}$, $H^{(l+1)}$ sont les sorties respectives des couches d'indices l, l+1

[0079] $\hat{A} = A + I$ avec A une matrice d'adjacence dont les entrées (i,j) valent 0 si les nœuds i et j du graphe ne sont pas connectés et valent le score de similarité si les nœuds i et j sont connectés,

[0080] $\hat{D}$ est une matrice diagonale dont les valeurs sont égales aux degrés de chaque nœud,

[0081] $W^{(l)}$ correspond aux paramètres de la couche d'indice l qui sont appris,

[0082] σ est une fonction d'activation, par exemple une fonction ReLu

[0083] Le réseau GCN délivre en sortie un vecteur de caractéristiques de dimension N, noté $h'_{SFE}=[h'_1, h'_2, h'_N]$ pour chaque nœud du graphe correspondant à chaque flux de communication.

[0084] Par exemple, le réseau GCN mis en œuvre à l'étape 405 comporte deux couches de convolution qui sont entrainées pour apprendre des représentations de dimension réduite qui permettent de capturer la topologie du graphe, les relations nœud à nœud, les motifs temporels et d'autres informations permettant de discriminer des comportements associés à des flux de communication autorisés de ceux associés à des flux de communication malveillants.

[0085] Un avantage à l'utilisation d'un réseau de graphes est qu'il permet de prendre en compte des informations spatiales relatives aux connexions entre nœuds voisins d'un graphe. Par ailleurs, l'équation de convolution (2) prend en compte les valeurs de si-

milarités associées aux arêtes des graphes pour exploiter les similarités entre flux de communication dans l'apprentissage des caractéristiques discriminantes.

[0086] Alternativement, ou parallèlement à l'application du réseau GCN 405, un second réseau de graphes 406 de type réseau d'attention de graphes ou « Graph Attention Network » GAT en anglais peut être appliqué afin de générer un second vecteur de caractéristiques pour chaque nœud.

[0087] Le réseau d'attention de graphes GAT peut être implémenté tel que décrit dans les publications « Graph Attention Networks », P Velickovic et al, ICLR 2018 ou « How attentive are graph attention networks ? », S. Brody et al, ICLR 2022.

[0088] Un réseau d'attention de graphes est composé d'un réseau neuronal de graphes suivi d'une couche d'attention de graphe ou « Graph Attention Layer » en anglais. La couche d'attention de graphe prend en entrée un vecteur de caractéristiques $h=[h_1, h_2, h_N]$ fourni en sortie du réseau neuronal de graphes et produit en sortie un vecteur de caractéristiques transformées $h'_{AFE}=[h'_1, h'_2, h'_{N'}]$.

[0089] La couche d'attention de graphe introduit la notion de valeur d'attention h_{ij} entre deux nœuds du graphe i et j, qui est définie par la relation suivante :

[0090] $$h_{ij} = a\left(W'h_i, W'h_j, W''\zeta_{i,j}\right) \quad (3)$$

[0091] a() est une fonction de projection qui prend deux vecteurs en argument et délivre une valeur réelle en sortie. W' et W'' sont deux matrices de transformation linéaire. $\zeta_{i,j}$ est le score de similarité associé à l'arête entre le nœud i et le nœud j.

[0092] La valeur d'attention h_{ij} représente l'importance relative de l'arête reliant les nœuds i et j du graphe. Cette valeur d'attention peut être utilisée pour estimer l'importance du nœud i. Le modèle d'attention est entrainé pour apprendre des paramètres ou poids d'attention pour chaque arête du graphe et pour collecter des informations sur les nœuds voisins d'un nœud afin de déterminer les caractéristiques de sortie à l'aide de la relation suivante :

[0093] $$h'_i = \sigma\left(\sum_{j\in\Omega_i} a_{ij} W' h_j\right)$$

[0094] a_{ij} est un coefficient qui traduit l'attention relative d'une arête par rapport à un voisinage dans le graphe. Ce coefficient peut être calculé à l'aide de la relation suivante :

[0095] $$a_{ij} = softmax_j\left(h_{ij}\right) = \frac{\exp(h_{ij})}{\sum_{n\in\Omega_i}\sum_{r\in R_{in}}\exp(h_{in})}$$

[0096] Ω_i est l'ensemble des nœuds voisins du nœud i

[0097] R_{in} est l'ensemble des arêtes qui connectent les nœuds i et n.

[0098] σ est une fonction d'activation, par exemple la fonction ReLU

[0099] Un objectif de la couche d'attention est d'estimer une importance relative à différents nœuds voisins du nœud considéré plutôt que de les considérer tous avec le même

poids.

[0100] Lorsque plusieurs réseaux neuronaux de graphes sont utilisés, par exemple deux réseaux 405,406 comme sur l'exemple de la [Fig.4], un module de recombinaison 407 est appliqué pour concaténer les vecteurs de caractéristiques h'_{AFE} et h'_{SFE} en un seul vecteur, pour chacun des nœuds du réseau. Cette étape 407 est optionnelle si un seul réseau neuronal de graphes est utilisé.

[0101] Enfin, une étape 408 finale de classification des flux est appliquée à chaque vecteur de caractéristiques associé à chaque flux pour déterminer si le flux est autorisé ou malveillant. L'étape 408 de classification peut être mise en œuvre au moyen d'un réseau de neurones artificiels tel qu'un perceptron multi-couches. Par exemple, le réseau de neurones 408 peut comprendre plusieurs couches entrainées pour distinguer les flux légitimes des flux malveillants à partir de l'analyse des caractéristiques reçues en entrée.

[0102] Chaque couche peut exécuter un calcul de convolution du type donné par la relation suivante :

[0103] $$Z_l = f\left(W_m^l Z_{l-1} + b_l + \omega^l e\right)$$

[0104] Z_l, Z_{l-1} sont les sorties respectives des couches d'indice l et l-1,

[0105] W_m^l et b_l sont les poids synaptiques et biais de la couche d'indice l,

[0106] e est l'erreur de reconstruction et ω^l son poids correspondant.

[0107] La dernière couche du réseau de classification 408 comporte une seule sortie qui donne un score P_m de probabilité de l'aspect légitime ou malveillant d'un flux. Si ce score est supérieur à un seuil donné, le flux est considéré comme malveillant.

[0108] L'ensemble des modèles d'intelligence artificielle utilisés pour réaliser le moteur d'intelligence artificielle 403 peuvent être entrainés de façon supervisée ou non supervisée.

[0109] Un entrainement supervisé suppose l'utilisation de données d'entrainement DE comportant des flux de communication légitimes et malveillants. Ces données d'entrainement sont utilisées pour l'apprentissage des modèles 404,405,406 conjointement.

[0110] Dans le cas où l'entrainement est non supervisé, des algorithmes d'intelligence artificielle non supervisés sont utilisés pour réaliser les étapes 404,405 et 406. Par exemple les algorithmes décrits dans les publications « Unsupervised universal self attention network for graph classification », Dai Quoc Nguyen et al, ICLR 2020 Conference ou « Toward unsupervised graph neural network : interactive clustering and embedding via optimal transport », Liang Yang, 2020 IEEE International Conference on Data Mining, peuvent être utilisés.

[0111] L'invention peut être mise en œuvre au sein d'un système de détection d'intrusion

NIDS comme décrit à la [Fig.1]. Le système de détection d'intrusion peut comprendre une sonde réseau pour capturer des flux réseaux ou des trames réseaux transitant dans un réseau informatique. Le système de détection d'intrusion comporte un processeur configuré pour exécuter les étapes de la méthode de détection d'intrusion selon l'un quelconque des modes de réalisation de l'invention.

[0112] L'invention peut être mise en œuvre en tant que programme d'ordinateur comportant des instructions pour son exécution. Le programme d'ordinateur peut être enregistré sur un support d'enregistrement lisible par un processeur.

[0113] La référence à un programme d'ordinateur qui, lorsqu'il est exécuté, effectue l'une quelconque des fonctions décrites précédemment, ne se limite pas à un programme d'application s'exécutant sur un ordinateur hôte unique. Au contraire, les termes programme d'ordinateur et logiciel sont utilisés ici dans un sens général pour faire référence à tout type de code informatique (par exemple, un logiciel d'application, un micro logiciel, un microcode, ou toute autre forme d'instruction d'ordinateur) qui peut être utilisé pour programmer un ou plusieurs processeurs pour mettre en œuvre des aspects des techniques décrites ici. Les moyens ou ressources informatiques peuvent notamment être distribués ("*Cloud computing*"), éventuellement selon des technologies de pair-à-pair. Le code logiciel peut être exécuté sur n'importe quel processeur approprié (par exemple, un microprocesseur) ou cœur de processeur ou un ensemble de processeurs, qu'ils soient prévus dans un dispositif de calcul unique ou répartis entre plusieurs dispositifs de calcul (par exemple tels qu'éventuellement accessibles dans l'environnement du dispositif). Le code exécutable de chaque programme permettant au dispositif programmable de mettre en œuvre les processus selon l'invention, peut être stocké, par exemple, dans le disque dur ou en mémoire morte. De manière générale, le ou les programmes pourront être chargés dans un des moyens de stockage du dispositif avant d'être exécutés. L'unité centrale peut commander et diriger l'exécution des instructions ou portions de code logiciel du ou des programmes selon l'invention, instructions qui sont stockées dans le disque dur ou dans la mémoire morte ou bien dans les autres éléments de stockage précités.

[0114] Le dispositif de détection d'intrusions NDIS peut être implémenté sur un calculateur basé, par exemple, sur un processeur embarqué. Le processeur peut être un processeur générique, un processeur spécifique, un circuit intégré propre à une application (connu aussi sous le nom anglais d'ASIC pour « Application-Specific Integrated Circuit ») ou un réseau de portes programmables in situ (connu aussi sous le nom anglais de FPGA pour « Field-Programmable Gate Array »). Le dispositif de calcul peut utiliser un ou plusieurs circuits électroniques dédiés ou un circuit à usage général. La technique de l'invention peut se réaliser sur une machine de calcul reprogrammable (un processeur ou un micro-contrôleur par exemple) exécutant un programme comprenant une

séquence d'instructions, ou sur une machine de calcul dédiée (par exemple un ensemble de portes logiques comme un FPGA ou un ASIC, ou tout autre module matériel).

[0115] L'invention permet d'améliorer les performances de détection de flux malveillants par rapport aux techniques de l'art antérieur qui sont basées sur des graphes plus classiques. En particulier la précision de détection est améliorée, c'est-à-dire le ratio entre le nombre de flux malveillants correctement détectés et l'ensemble des flux malveillants détectés (y compris ceux qui correspondent à des faux positifs).

Revendications

[Revendication 1] Méthode, mise en œuvre par ordinateur, de détection de flux malveillants dans un réseau informatique comprenant les étapes de :

- Capturer un ensemble de flux réseau, chaque flux réseau étant associé à un premier vecteur de caractéristiques du flux mesurées et à un instant de mesure,
- Créer (402), pour chaque instant de mesure, un graphe (GR) dont les nœuds représentent chaque flux capturé, deux nœuds du graphe étant connectés si les deux flux associés sont émis par le même utilisateur ou sont transmis vers le même utilisateur, chaque nœud étant pourvu dudit premier vecteur de caractéristiques,
- Exécuter (403) un moteur d'intelligence artificielle entrainé pour détecter une intrusion dans le réseau informatique à partir des graphes et comprenant les sous-étapes de :
  i. Appliquer aux graphes au moins un réseau neuronal de graphes (405,406) entrainé pour générer, pour chaque flux, un second vecteur de caractéristiques aptes à discriminer un flux autorisé d'un flux malveillant, l'entrainement dudit réseau neuronal de graphes exploitant au moins la topologie des graphes,
  ii. Appliquer (408) aux seconds vecteurs de caractéristiques un réseau de neurones artificiel classifieur entrainé pour classifier les flux parmi deux classes correspondant respectivement aux flux autorisés et aux flux malveillants.

[Revendication 2] Méthode de détection de flux malveillants dans un réseau informatique selon la revendication 1 dans laquelle l'étape d'exécuter (403) le moteur d'intelligence artificielle comprend en outre une étape (404), préalable à l'application de l'au moins un réseau neuronal de graphes (405,406), consistant à appliquer aux premiers vecteurs de caractéristiques au moins un réseau neuronal artificiel (404) entrainé pour réduire la dimension des premiers vecteurs par extraction de caractéristiques discriminants un flux autorisé d'un flux malveillant, l'entrainement dudit réseau neuronal artificiel n'exploitant pas la topologie des graphes.

[Revendication 3] Méthode de détection de flux malveillants dans un réseau informatique selon la revendication 2 dans laquelle l'au moins un réseau neuronal artificiel (404) est un perceptron multi-couches.

[Revendication 4] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes comprenant une étape de prétraitement (401) des premiers vecteurs de caractéristiques mesurées consistant au moins en une normalisation des valeurs et/ou un encodage catégoriel des valeurs.

[Revendication 5] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes comprenant en outre une étape de calcul (402), pour chaque arête du graphe (GR), d'un score de similarité entre les premiers vecteurs de caractéristiques des deux flux associés aux deux nœuds connectés par l'arête.

[Revendication 6] Méthode de détection de flux malveillants dans un réseau informatique selon la revendication 5 dans laquelle le score de similarité est calculé au moyen d'une mesure de similarité en cosinus.

[Revendication 7] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes dans lequel l'au moins un réseau neuronal de graphes est un réseau neuronal convolutionnel de graphes (405).

[Revendication 8] Méthode de détection de flux malveillants dans un réseau informatique selon la revendication 7 en combinaison avec la revendication 5 ou 6 dans laquelle le réseau neuronal convolutionnel de graphes (405) comporte plusieurs couches configurées pour mettre en œuvre un calcul de convolution du type :

$$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{\frac{1}{2}}H^{(l)}W^{(l)}\right),$$

$H^{(l)}$, $H^{(l+1)}$ sont les sorties respectives des couches d'indices l, l+1

$\hat{A} = A + I$ avec A une matrice d'adjacence dont les entrées (i,j) valent 0 si les nœuds i et j du graphe ne sont pas connectés et valent le score de similarité si les nœuds i et j sont connectés,

$\hat{D}$ est une matrice diagonale dont les valeurs sont égales aux degrés de chaque nœud,

$W^{(l)}$ correspond aux paramètres de la couche d'indice l qui sont appris,

σ est une fonction d'activation, par exemple une fonction ReLu.

[Revendication 9] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes dans lequel l'au

moins un réseau neuronal de graphes est un réseau d'attention de graphes (406).

[Revendication 10] Méthode de détection de flux malveillants dans un réseau informatique selon la revendication 9 dans lequel le réseau d'attention de graphes (406) comporte un réseau neuronal de graphes et une couche d'attention entrainée au moyen d'un modèle d'attention configuré pour apprendre des valeurs d'attention pour chaque arête du graphe et calculer une caractéristique pour chaque nœud à l'aide d'une pondération des caractéristiques des nœuds voisins en fonction des valeurs d'attention.

[Revendication 11] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes dans lequel plusieurs réseaux neuronaux de graphe (405,406) sont exécutés en parallèle et la méthode comporte une étape de concaténation (407) des seconds vecteurs de caractéristiques générés par chaque réseau neuronal de graphe pour fournir un vecteur concaténé au réseau de neurones artificiel classifieur (408).

[Revendication 12] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes dans laquelle le réseau de neurones artificiel classifieur (408) est un perceptron multicouches.

[Revendication 13] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes dans laquelle le premier vecteur de caractéristiques mesurées contient des informations caractéristiques des flux parmi : les adresses sources et destination, les protocoles utilisés, des statistiques relatives aux paquets transmis via ce flux.

[Revendication 14] Méthode de détection de flux malveillants dans un réseau informatique selon l'une quelconque des revendications précédentes dans laquelle le moteur d'intelligence artificielle (403) est entrainé de manière supervisée ou non supervisée.

[Revendication 15] Système de détection de flux malveillants dans un réseau de télécommunications comprenant au moins une sonde apte à acquérir un flux réseau ou un ensemble de trames réseau transitant dans ledit réseau et au moins un dispositif de détection de flux malveillants (NIDS) configuré pour mettre en œuvre les étapes de la méthode de détection flux malveillants selon l'une des revendications précédentes, à partir du flux réseau ou de l'ensemble de trames réseau.

[Fig. 1]

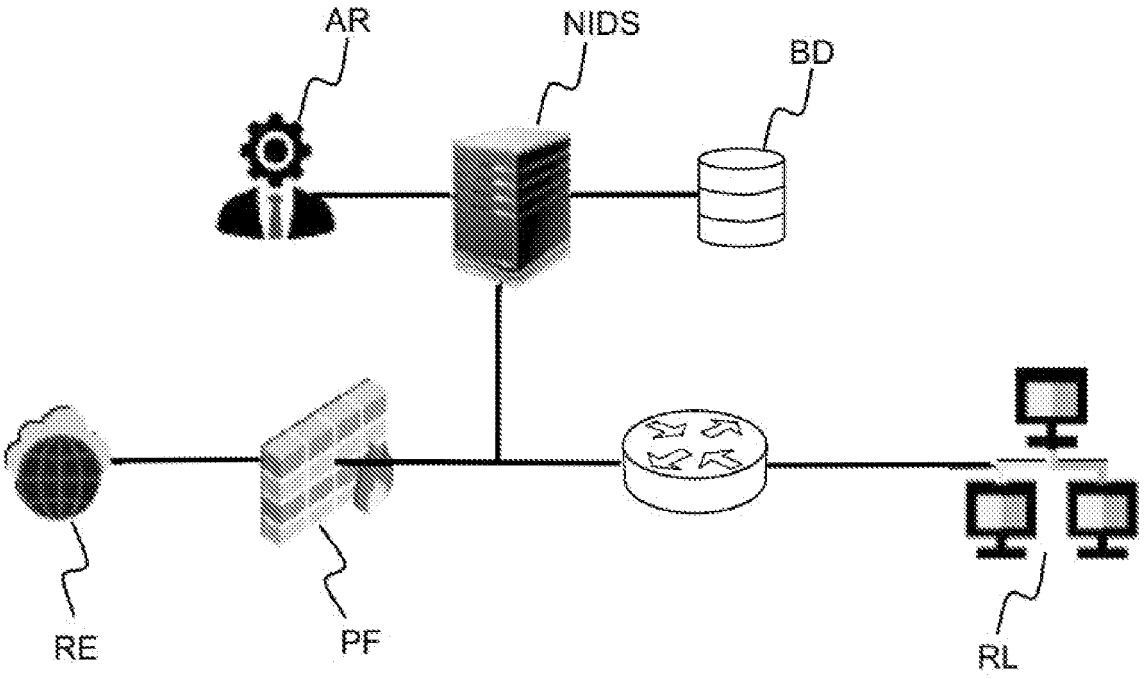


FIG.1

[Fig. 2]

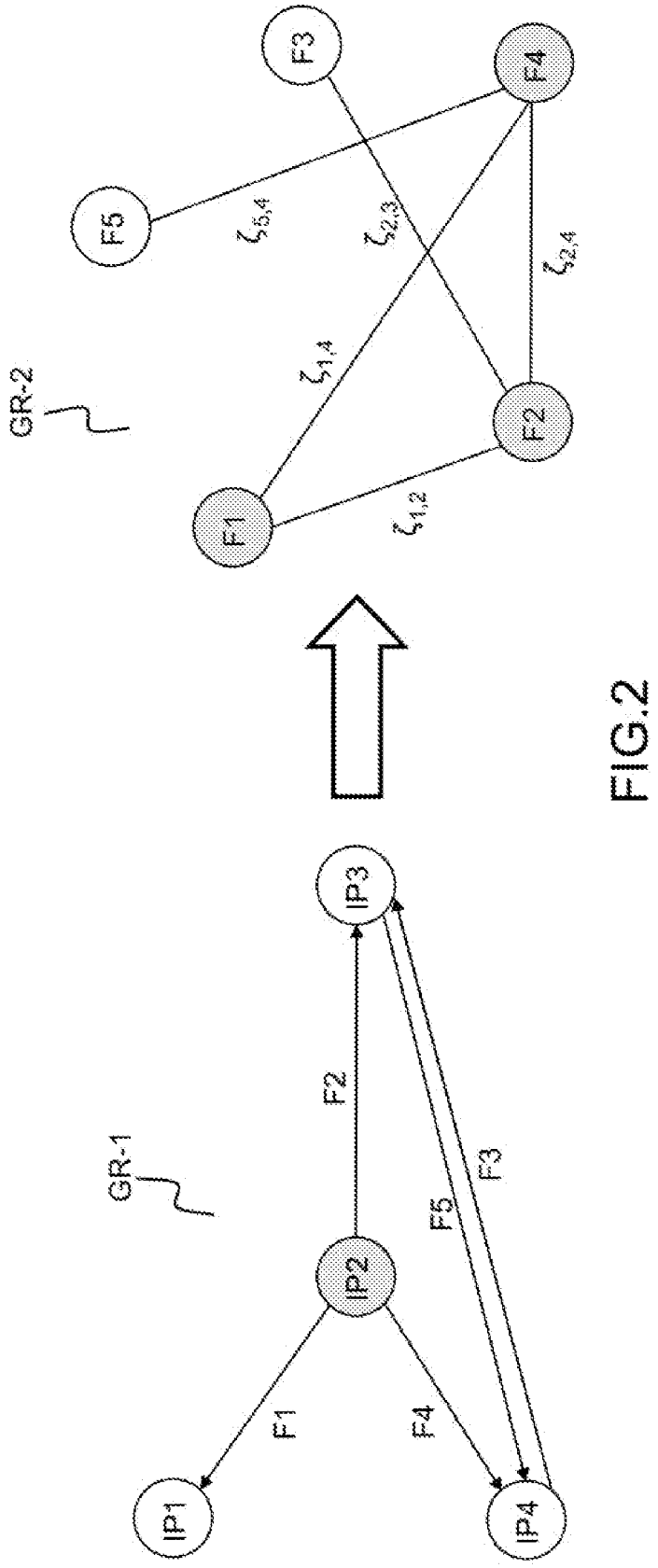

GR-2
F1
F2
F3
F4
F5
$\zeta_{1,2}$
$\zeta_{1,4}$
$\zeta_{2,3}$
$\zeta_{2,4}$
$\zeta_{5,4}$
GR-1
IP1
IP2
IP3
IP4
F1
F2
F3
F4
F5

FIG.2

[Fig. 3]

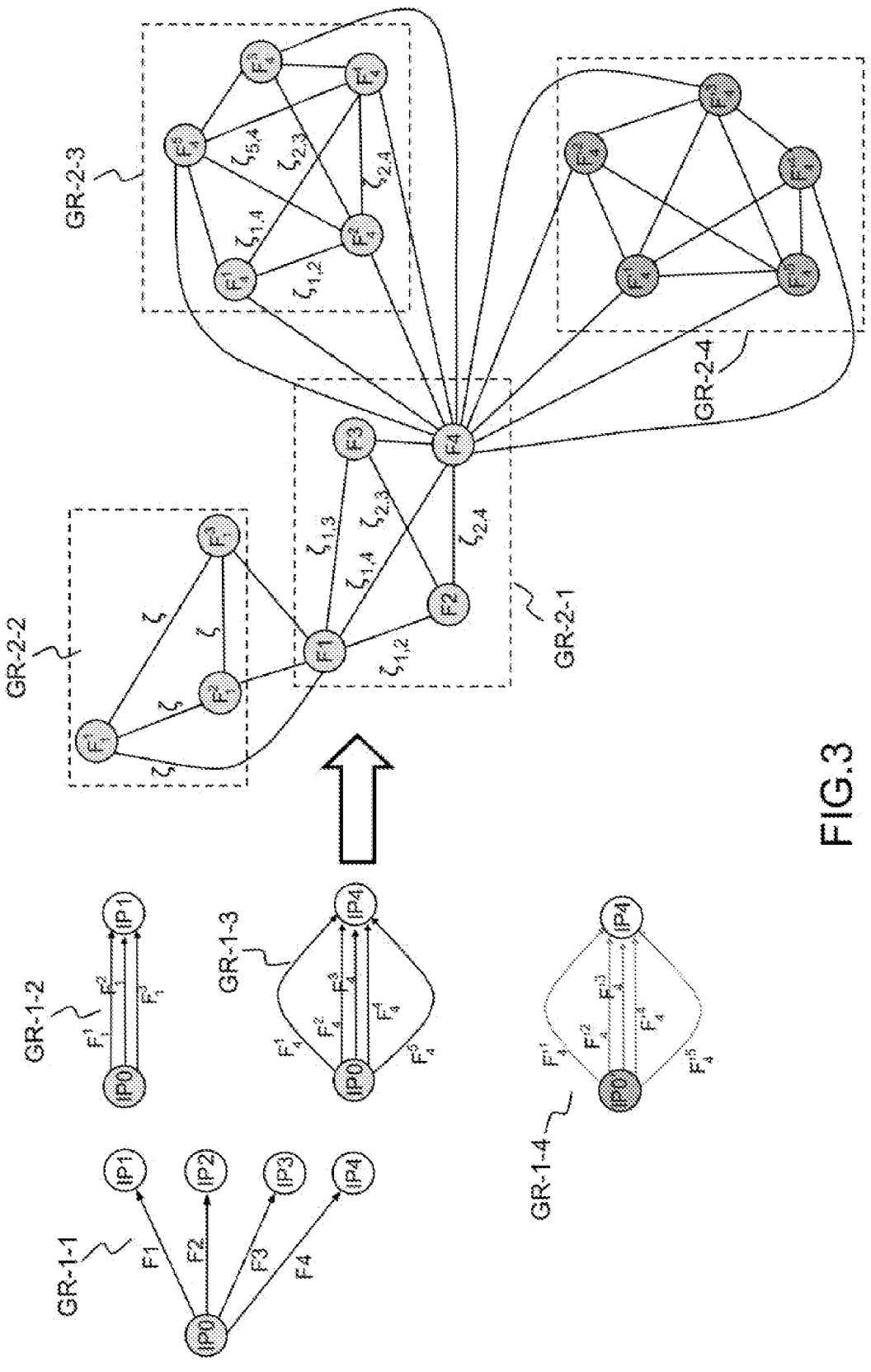

GR-2-3
GR-2-4
GR-2-1
GR-2-2
GR-1-2
GR-1-3
GR-1-4
GR-1-1
F1
F2
F3
F4
IP0
IP1
IP2
IP3
IP4

FIG.3

[Fig. 4]

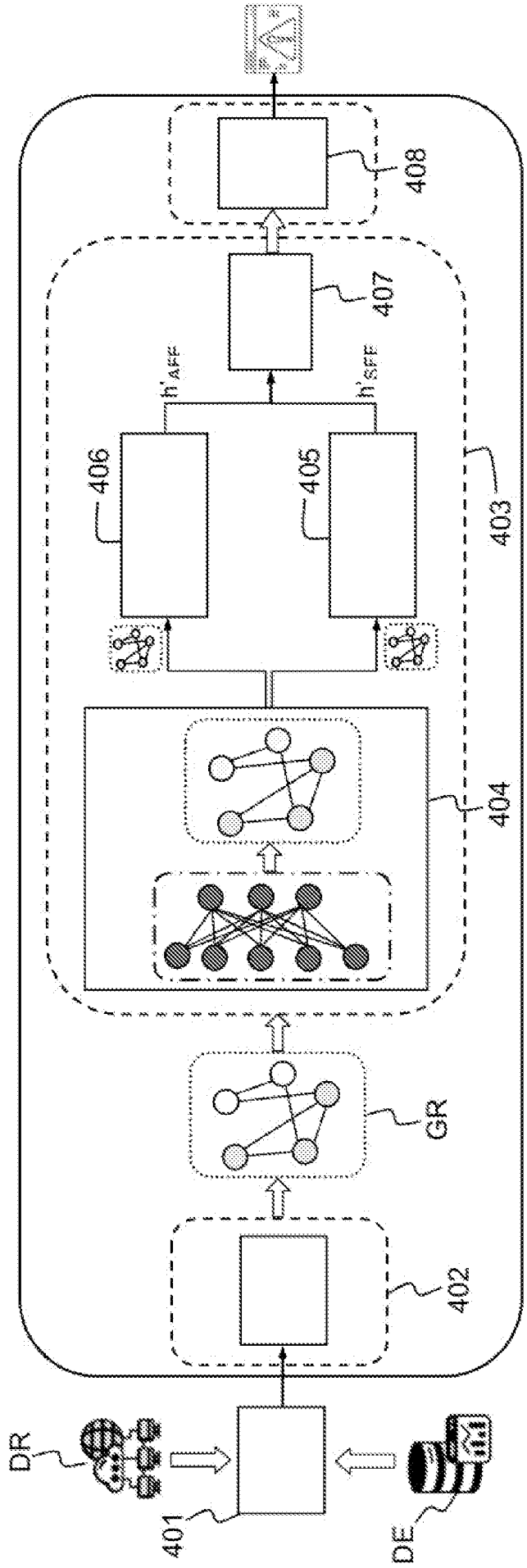


FIG.4

RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

RAPPORT DE RECHERCHE PRÉLIMINAIRE

établi sur la base des dernières revendications déposées avant le commencement de la recherche

N° d'enregistrement national

FA 915695
FR 2213324

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	WAI WENG LO ET AL: "E-GraphSAGE: A Graph Neural Network based Intrusion Detection System for IoT", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 6 janvier 2022 (2022-01-06), XP091125880,	1,4, 8-11,15	H04L43/04 H04L43/02 G06F21/00 G06F21/55 G06N3/02
Y	* le document en entier *	6	

Y	US 2021/248443 A1 (SHU XIAOKUI [US] ET AL) 12 août 2021 (2021-08-12) * alinéa [0145] *	6	

DOMAINES TECHNIQUES RECHERCHÉS (IPC)

G06N
H04L

Date d'achèvement de la recherche	Examinateur
29 juin 2023	Padilla Serrano, M

CATÉGORIE DES DOCUMENTS CITÉS

X : particulièrement pertinent à lui seul
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie
A : arrière-plan technologique
O : divulgation non-écrite
P : document intercalaire

T : théorie ou principe à la base de l'invention
E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.
D : cité dans la demande
L : cité pour d'autres raisons
.....
& : membre de la même famille, document correspondant

EPO FORM 1503 12.99 (P04C14)

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 2213324 FA 915695

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.
Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **29-06-2023**
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 2021248443 A1	12-08-2021	CN 115039098 A	09-09-2022
		EP 4100856 A1	14-12-2022
		JP 2023512507 A	27-03-2023
		US 2021248443 A1	12-08-2021
		WO 2021155971 A1	12-08-2021

EPO FORM P0465