

**DISPOSITIF DE COLLECTE DE DONNÉES POUR LA SURVEILLANCE DE
FLUX DANS UN RÉSEAU DE DONNÉES**

[0001] La présente invention concerne les techniques de surveillance de flux de communication dans des réseaux de données.

- 5 [0002] Dans le cadre de la surveillance de flux de données dans un réseau de données tel qu'Internet, les équipements sont confrontés à la multiplication des protocoles utilisés, à la masse des données transférées ainsi qu'à la multiplicité des émetteurs/récepteurs.

- 10 [0003] Il existe actuellement deux types de dispositifs pour la mise en œuvre de telles surveillances de flux. Des dispositifs d'un premier type collectent pendant une durée déterminée un ensemble de paquets puis, dans une étape ultérieure, analysent et traitent les paquets collectés. Un exemple de ce type de dispositif repose sur le logiciel distribué sous la marque Wireshark, autrefois la marque Ethereal.

- 15 [0004] Le second type de dispositif utilise des mécanismes matériels pour acquérir les données puis les transmettre à des dispositifs d'analyse. Très souvent la transmission d'information entre l'acquisition et les dispositifs d'analyse repose sur le protocole ipfix standardisé par l'IETF (Internet Engineering Task Force) sous la référence RFC 3917.

- 20 [0005] Dans "Incorporation of Application Layer Protocol Syntax into Anomaly Detection", Information Systems Security, Springer, pp. 188-202, décembre 2008, P. Düssel et al. décrivent l'incorporation de la syntaxe d'un protocole de couche applicative dans un processus de détection d'anomalies dans des réseaux de télécommunications. Un système de détection d'intrusion
25 appelé "Bro" capture des paquets sur le réseau et réassemble des segments TCP pour transmettre des paquets entrants à un analyseur de protocole utilisant un langage d'analyse syntaxique (yacc) appelé "binpac".

- [0006] La demanderesse a développé des technologies de surveillance de réseaux de type Internet. Elle a en particulier déposé les demandes de brevet
30 EP 1 722 509 A1 et EP 1 722 508 A1 décrivant des architectures multiniveaux

permettant d'examiner en temps réel des empilements de protocoles complexes afin d'en extraire les flux, c'est-à-dire l'ensemble structuré de données transmis entre un émetteur et un récepteur.

5 [0007] Au vu de la quantité d'informations à traiter au niveau des collecteurs de données, ceux-ci doivent se limiter à faire le traitement strictement indispensable pour transmettre l'information pertinente à un analyseur et seulement celle-ci.

10 [0008] Certaines solutions chargent les collecteurs de données d'opérations de traitement permettant un transfert d'informations structurées et pertinentes. Toutefois, ces solutions obligent à développer des architectures matérielles très performantes compte tenu de la masse d'informations à examiner en temps réel par le collecteur de données.

15 [0009] D'autres types de solutions limitent les traitements effectués par les collecteurs de données, en leur faisant transférer des données non structurées, y compris des données redondantes et/ou inutiles, ce qui entraîne d'autres contraintes au niveau de l'analyseur et de son interface avec le collecteur.

[0010] Une autre possibilité consiste à effectuer un traitement différé au niveau de l'analyseur. Il en résulte toutefois une moindre réactivité de l'application de surveillance de flux.

20 [0011] Il serait donc avantageux de proposer une technique de collecte de données qui permette de transférer à l'analyseur seulement des données structurées et pertinentes tout en limitant les ressources matérielles nécessaires aux traitements d'acquisition et de mise en forme de ces données.

25 [0012] Selon l'invention, il est proposé un dispositif de collecte de données pour la surveillance de flux dans un réseau de données utilisant un mode de transmission par paquets. Ce dispositif comprend:

- un extracteur de données contenues dans des paquets appartenant à un flux défini par un émetteur, un récepteur et un protocole;
- un analyseur syntaxique agencé pour:
 - 30 - recevoir les données en temps réel de l'extracteur;

- décomposer les données reçues en éléments en fonction de règles syntaxiques dudit protocole, les règles syntaxiques permettant de représenter lesdits éléments suivant une structure en arbre; et
- associer des indicateurs d'états arborescents respectifs à certains au moins desdits éléments, où l'indicateur d'état arborescent associé à un élément repère ledit élément dans la structure en arbre; et
- une interface pour transmettre à un analyseur de flux les indicateurs d'états arborescents accompagnés des éléments auxquels ils ont été associés.

[0013] Dans un mode de réalisation du dispositif, l'indicateur d'état arborescent associé à un élément comporte un identifiant d'un nœud père de la structure en arbre auquel se rattache un emplacement recevant ledit élément dans la représentation suivant la structure en arbre. L'analyseur syntaxique peut être agencé pour:

- affecter un identifiant de nœud à un élément courant s'il est déterminé selon les règles syntaxiques que l'élément courant est reçu, dans la représentation suivant la structure en arbre, à un emplacement situé à un nœud de la structure en arbre; et
- inclure l'identifiant de nœud affecté à l'élément courant dans l'indicateur d'état arborescent associé audit élément courant.

[0014] Dans une réalisation avantageuse, l'analyseur syntaxique est agencé pour ne conserver en mémoire, pour un flux donné après transmission d'un élément et de son indicateur d'état arborescent associé, que chaque identifiant de nœud rencontré en remontant la structure en arbre à partir de l'emplacement recevant ledit élément dans la représentation suivant la structure en arbre.

[0015] Un autre aspect de l'invention se rapporte à un système de surveillance de flux dans un réseau de données utilisant un mode de transmission par paquets. Ce système comprend un dispositif de collecte de données tel que défini ci-dessus et un analyseur de flux recevant les éléments

transmis avec les indicateurs d'états arborescents associés.

[0016] Un autre aspect encore de l'invention se rapporte à un procédé de surveillance de flux dans un réseau de données utilisant un mode de transmission par paquets. Le procédé comprend:

- 5 • extraire des données contenues dans des paquets appartenant à un flux défini par un émetteur, un récepteur et un protocole;
- décomposer en éléments les données reçues en temps réel, en fonction de règles syntaxiques dudit protocole, les règles syntaxiques permettant de représenter lesdits éléments suivant une structure en arbre;
- 10 • associer des indicateurs d'états arborescents respectifs à certains au moins desdits éléments, où l'indicateur d'état arborescent associé à un élément repère ledit élément dans la structure en arbre; et
- transmettre à un analyseur de flux les indicateurs d'états arborescents accompagnés des éléments auxquels ils ont été associés.

- 15 [0017] Le procédé peut comprendre en outre la construction d'une partie au moins de la structure en arbre au niveau de l'analyseur de flux à partir des éléments transmis avec les indicateurs d'états arborescents associés.

- [0018] D'autres particularités et avantages de la présente invention apparaîtront dans la description ci-après d'un exemple de réalisation non
20 limitatif, en référence aux dessins annexés, dans lesquels :

- la figure 1 est une vue schématique d'un système de surveillance dans un réseau de données selon un mode de réalisation de l'invention;
- la figure 2 est une vue schématique fonctionnelle d'un dispositif collecteur de données du système de surveillance de la figure 1;
- 25 - la figure 3 est un organigramme de fonctionnement du dispositif collecteur de la figure 2;
- la figure 4 est un exemple illustratif de décomposition en arbre dynamique d'une requête HTTP selon un mode de réalisation de l'invention; et

- la figure 5 montre un exemple de structure en arbre dans le cas d'un protocole de messagerie électronique.

[0019] En référence à la figure 1, un système de surveillance comporte un collecteur de données 1 branché directement sur un réseau de données 3, par exemple l'Internet, utilisant un protocole par paquets de type IP ("Internet Protocol") afin d'acquérir les paquets circulant sur ce réseau. Ce collecteur de données 1 utilise par exemple l'architecture détaillée dans la demande de brevet EP 1 722 509 A1 ou la demande de brevet EP 1 722 508 A1. Il transmet par une connexion de données 5 à un analyseur 7 des données structurées
10 extraites des paquets circulant sur le réseau de données 3.

[0020] Les transmissions sur le réseau de données 3 utilisent de façon classique un empilement de protocoles de transmission. Ainsi, par exemple, dans le cadre de la consultation de sites web, un paquet IP contient une structure de données du protocole TCP ("Transmission Control Protocol")
15 contenant elle-même une structure de données HTTP ("HyperText Transfer Protocol"). Dans le cadre de la messagerie électronique (e-mail), la structure de données TCP contient une structure de données SMTP ("Simple Mail Transfer Protocol"), etc. En fonction du type de flux considéré, de nombreux autres protocoles de niveau applicatif ou transport sont employés, comme il est
20 bien connu dans le domaine des communications sur réseau IP.

[0021] Les informations circulant dans le cadre d'un flux donné sont rarement contenues dans un unique paquet IP. Ces informations sont au contraire segmentées pour être transmises dans plusieurs paquets successifs. Ces paquets IP circulent sur le réseau, l'en-tête du paquet contenant des
25 informations pour permettre son routage, notamment les adresses IP de la source et de la destination du paquet. Il peut s'intercaler entre eux d'autres paquets relevant d'autres flux. Les paquets peuvent suivre des routes différentes au sein du réseau, et/ou être reçus d'un un ordre différent de celui dans lequel ils ont été émis.

30 [0022] Pour surveiller un flux donné, il convient d'identifier les paquets relevant du flux en question, d'en extraire les informations pertinentes et de les

analyser en fonction des besoins de l'application de surveillance considérée. Dans l'architecture de la figure 1, l'identification et l'extraction incombent au collecteur de données 1 tandis que l'application de surveillance proprement dite est exécutée par l'analyseur 7.

- 5 [0023] En référence à la figure 2, le collecteur de données 1 comporte un extracteur de données 11 connecté à un nœud ou un lien du réseau 3 afin d'observer le trafic. L'extracteur de données 11 a une interface réseau adaptée aux couches protocolaires basses utilisées par le réseau 3 au niveau du nœud ou lien où l'extracteur est connecté. Il extrait les données contenues dans les
- 10 paquets appartenant à un flux spécifié, le flux étant défini par un émetteur, un récepteur et un protocole (par exemple, SMTP pour du courrier électronique, HTTP pour de la navigation web, etc.).

- [0024] Le collecteur de données 1 comporte en outre un analyseur syntaxique 13 qui reçoit en temps réel des données extraites par l'extracteur 11
- 15 pour un flux donné. L'analyseur syntaxique 13 dispose de ressources logiques incorporant des règles syntaxiques du protocole pertinent pour le flux considéré. Il décompose les données extraites du flux en éléments, encore appelés événements, respectant la syntaxe du protocole du flux. Les éléments ainsi obtenus sont fournis à une interface 15 qui se charge de les transmettre à
- 20 l'analyseur externe 7 avec des indicateurs d'états arborescents.

- [0025] La décomposition des données du flux en éléments est effectuée par l'analyseur syntaxique 13 en fonction des règles syntaxiques applicables au protocole, qui permettent de représenter ces éléments suivant une structure en arbre. En pratique, la structure en arbre n'est pas mémorisée de manière
- 25 complète au niveau du collecteur de données 1. Elle est reconstituée par l'analyseur de données structurées 7 si l'application le requiert. Le collecteur de données 1 se contente de mémoriser des états arborescents dont des indicateurs sont respectivement associés aux éléments résultant de la décomposition opérée par l'analyseur syntaxique 13.

- 30 [0026] L'indicateur d'état arborescent associé à un élément repère cet élément dans la structure en arbre pour que l'analyseur de données

structurées 7 puisse le situer dans le flux surveillé.

[0027] Par arbre, il est entendu une structure de données au sens de la théorie des graphes comportant une donnée racine et des données filles de telle sorte qu'une donnée fille n'est reliée qu'à une seule donnée parente de
5 niveau supérieur. Une donnée comportant des liens vers des données filles est appelée nœud de l'arbre et une donnée sans donnée fille est appelée feuille de l'arbre. Un arbre est donc un graphe ne comportant pas de boucle.

[0028] Le fonctionnement général du collecteur 1 est illustré par le diagramme de la figure 3. Dans une étape de configuration 21, les paramètres
10 nécessaires à l'extracteur 11 lui sont fournis, en particulier:

- l'émetteur, le récepteur et le protocole pertinents pour un flux surveillé;
- une indication de types de données à extraire (toutes les données, certains messages seulement, données respectant telle ou telle condition configurable, ...). Dans une configuration simple, toutes les données du
15 flux sont extraites, les éventuels filtrages étant effectués au niveau de l'analyseur 7. Spécifier que certains types de données seulement sont à extraire permet, de manière facultative, de réduire la charge du collecteur 1 et la quantité d'information à faire transiter sur la connexion 5;
- l'identification d'une instance de l'analyseur syntaxique 13 en charge du
20 flux considéré puisqu'en pratique, l'extracteur 11 et l'analyseur 13 peuvent examiner plusieurs flux à la fois.

[0029] À l'étape 23, l'extracteur 11 examine les paquets au fur et à mesure de leurs transmissions sur le réseau 3. Pour ce faire, il utilise par exemple le procédé de reconnaissance et d'analyse de protocoles décrit dans la demande
25 de brevet WO 2004/017595 A2. Il est ainsi capable d'extraire, à l'étape 25, les données appartenant au protocole spécifié lors de la configuration et circulant de l'émetteur spécifié vers le récepteur spécifié, et de transmettre ces données en temps réel à l'analyseur syntaxique 13.

[0030] Afin d'illustrer le fonctionnement de l'analyseur syntaxique 13 à
30 l'étape de décomposition 27 indiquée sur la figure 3, on considère le cas particulier d'un flux obéissant au protocole HTTP entre un émetteur et un

récepteur donnés. Dans cet exemple illustratif, l'extracteur 11 analyse les paquets et extrait dans deux paquets successifs, après suppression des en-têtes, l'information indiquée au Tableau I.

Numéro paquet	Contenu HTTP du paquet
P1	GET /search?hl=fr&
P2	q=brevet HTTP/1.0\nHost: google.fr\n

Tableau I

5 [0031] Dans ce cas particulier, la requête HTTP complète:

GET /search?hl=fr&q=brevet HTTP/1.0\nHost: google.fr\n

a été segmentée en deux paquets pour qu'elle puisse être transportée tout en respectant les limites imposées par les protocoles de niveau inférieur (TCP, IP, couches 1 et 2) en termes de taille de la charge utile.

10 [0032] Sémantiquement, ce message HTTP se décompose en une commande de requête "GET" et un identifiant de ressource URI ("Uniform Resource Identifier"). L'URI se décompose lui-même en éléments ou événements syntaxiquement pertinents. Ces éléments sont agencés dans les données placées dans les paquets conformément aux règles syntaxiques
15 propres au protocole considéré (ici HTTP).

[0033] Par exemple, les données du paquet P1 peuvent se décomposer comme l'indique le Tableau II,

Numéro d'événement	contenu
1.1	request
1.2	uri=/search?hl=fr&
1.3	uri_param
1.4	name = hl
1.5	value = fr

Tableau II

tandis que les données du paquet P2 peuvent se décomposer comme l'indique le Tableau III.

Numéro d'événement	contenu
2.1	uri = q=brevet
2.2	uri_param
2.3	name = q
2.4	value = brevet
2.5	server = google.fr

Tableau III

- 5 [0034] L'événement 1.1 correspond à la détection de la commande "GET". Ensuite, l'analyseur syntaxique 13 détermine que le reste du paquet P1 fait partie de l'identifiant de ressource URI, ce qui constitue l'événement suivant 1.2. Cet événement "uri=/search?hl=fr&" doit lui-même être décomposé en événements car il fait apparaître une liste de paramètres qui peut être
- 10 incomplète. Cette nouvelle décomposition fait descendre d'un niveau dans la structure en arbre pour ajouter l'événement 1.3 qui annonce un paramètre d'URI. Ensuite, les événements 1.4 et 1.5 donnent les nom et valeur de ce

premier paramètre. À réception du paquet P2, il apparaît que l'URI se poursuit (événement 2.1) avec un autre paramètre (2.2) dont les nom et valeur sont donnés par les événements 2.3 et 2.4. Finalement, après le numéro de version du protocole HTTP, qui peut être ignoré dans cet exemple, le dernier
5 événement se rapporte au serveur interrogé, ici "google.fr".

[0035] L'analyseur syntaxique 13 effectue une décomposition de ce type à l'étape 27 en fonction de la syntaxe du protocole analysé, pour convertir les données brutes fournies par l'extracteur 11 en événements. Chaque protocole applicatif (SMTP, SIP, RTP, etc.) utilise une syntaxe propre, généralement
10 différente de celle du protocole HTTP donné en exemple. L'analyseur syntaxique peut avoir une architecture modulaire lui permettant d'analyser des protocoles différents sur différents flux. Ainsi un nouveau protocole est analysable par l'analyseur syntaxique moyennant le développement d'un module spécifique à sa syntaxe.

15 [0036] On souhaite transmettre à l'analyseur externe 7 les données suffisantes pour lui permettre de reconstituer le flux, notamment suivant la structure en arbre, mais avec aussi peu d'information additionnelle que possible. Si le collecteur de données 1 ne transmet que les contenus d'événements indiqués dans les colonnes de droite des Tableaux II et III, il
20 demeure des ambiguïtés. Par exemple, l'analyseur 7 ne pourra pas savoir si l'événement 1.2, 2.1 ou 2.5 concerne la même requête que l'événement 1.1. Pour éviter ces ambiguïtés sans avoir à transmettre des informations relevant de couches protocolaires plus basses (comme par exemple des numéros de séquence de segments TCP), les contenus d'événement sont transmis par le
25 collecteur de données à l'étape 31 sous une forme structurée avec les indicateurs d'états arborescents associés par l'analyseur syntaxique 13 à l'étape 29.

[0037] Les états arborescents peuvent être déterminés de la manière illustrée par la figure 4 dans l'exemple déjà discuté ci-dessus. Sur cette figure
30 4, la colonne de gauche montre la décomposition des paquets P1, P2 extraits du flux en éléments ou événements tels que rapportés aux Tableaux II et III. La colonne centrale de la figure 4 montre les événements tels que transmis par le

collecteur de données 1 à l'analyseur externe 7 à l'étape 31, accompagnés de leurs indicateurs d'états arborescents respectifs, et la colonne de droite montre la manière dont l'analyseur 7 est capable de reconstituer l'arbre représentatif du flux surveillé. Dans cette partie droite, les caractères gras indiquent les éléments que l'analyseur 7 ajoute à l'arbre au fur et à mesure de leur réception depuis le collecteur 1.

[0038] L'arbre peut être construit par l'analyseur 7 au fur et à mesure des événements reçus, en commençant par un nœud racine correspondant au flux analysé. Ce nœud racine est désigné par "http" sur la figure 4 et porte par exemple un identifiant de nœud (id) égal à 15C4. Dans cet exemple, l'arbre comporte déjà deux requêtes reçues avant celle portée par les paquets P1 et P2, désignées en partie supérieure de la colonne de droite par les deux nœuds "request" dont les ramifications de sont pas représentées pour éviter de surcharger le dessin.

[0039] À réception de l'événement 1.1 indiquant une nouvelle requête pour le flux considéré, un nouveau nœud "request" est ajouté à l'arbre, ayant pour nœud père la racine "http". Comme il s'agit d'un événement dont l'analyseur syntaxique 13 sait, grâce aux règles du protocole HTTP, qu'il aura des données filles, il lui affecte un identifiant de nœud, dans cet exemple id = 38FA. L'indicateur d'état arborescent associé à l'événement 1.1 par l'analyseur syntaxique 13 à l'étape 29 comporte dans ce cas l'identifiant de nœud 38FA qu'il vient de lui affecter et l'identifiant (p) de son nœud père dans l'arbre, ici p = 15C4.

[0040] À réception de l'événement 1.1 avec l'indicateur (id = 38FA / p = 15C4), l'analyseur 7 pourra rajouter le nouveau nœud à l'arbre, dont il garde en mémoire l'identifiant 38FA en rapport avec sa position (fils du nœud racine 15C4).

[0041] L'événement suivant (2.2) résultant de l'étape de décomposition 27 se voit également affecter à l'étape 29 un indicateur d'état arborescent comportant l'identifiant de son nœud père dans l'arbre, en l'occurrence p = 38FA. Comme l'analyseur syntaxique 13 peut déterminer que cet

événement correspond à une feuille de l'arbre, il n'a pas besoin de lui affecter un identifiant de nœud et son indicateur d'état arborescent peut donc se limiter à l'identifiant $p = 38FA$ de son nœud père désignant la requête HTTP en cours.

5 [0042] Le troisième événement 1.3 annonce que la suite contient des paramètres de l'URI. Il peut donc être représenté comme un nœud fils du nœud "request" (38FA) qui aura des liens avec d'autres nœuds fils. En conséquence de quoi, l'analyseur syntaxique 13 lui affecte un identifiant unique, par exemple $id = 8B53$, de sorte que son indicateur d'état arborescent associé est ($id = 8B53 / p = 38FA$) .

10 [0043] Les événements suivants (1.4, 1.5) qui donnent les nom et valeur de paramètres de l'URI sont attachés sous forme de feuilles au nœud 8B53. Leurs indicateurs d'états arborescents sont simplement $p = 8B53$, sans spécifier de nouvel identifiant de nœud.

15 [0044] Pour les événements 2.1-2.5 extraits du paquet P2, la logique de construction de l'arbre par l'analyseur 7, représenté en partie basse de la figure 4, est semblable à celle qui vient d'être décrite. L'analyseur syntaxique 13 du collecteur 1 affecte donc à la volée des nouveaux identifiants aux nœuds qui apparaissent (sans avoir à le faire pour les feuilles de l'arbre), et il associe à chaque événement transmis un indicateur d'état arborescent comportant
20 l'identifiant unique de son nœud père et, s'il est lui-même représenté par un nœud ayant une descendance, l'identifiant nouvellement affecté à ce nœud.

[0045] Il est à noter que l'URI, qui dans l'exemple a été tronqué dans le paquet P1 par suite de la segmentation des paquets, est complété après réception de l'événement 2.1 par l'analyseur 7. Celui-ci peut en effet constater
25 d'après la précédente feuille "uri=/search?hl=fr&" que l'URI est tronqué de sorte qu'à réception d'un nouvel événement de type URI (2.1), il procédera par concaténation sur la feuille correspondante plutôt que par création d'une nouvelle feuille. Ce traitement par l'analyseur 7 ne requiert pas de signalisation particulière de la part du collecteur 1.

30 [0046] L'ensemble du traitement de construction de l'arbre à l'aide des indicateurs d'états arborescents est visualisable sur la figure 4. La requête

HTTP analysée est représentable sous forme d'un arbre tel que représenté en bas et à droite de la figure 4.

[0047] Cette transmission permet avantageusement de ne transmettre que les informations nécessaires tout en préservant la structuration, ou le contexte, des données transmises grâce aux identifiants de nœuds. Elle est effectuée au fur et à mesure de la réception des paquets par le collecteur 1 et elle permet à l'analyseur 7 de reconstituer les flux surveillés, immédiatement si nécessaire.

[0048] Un autre avantage est que le collecteur de données 1, qui est soumis aux plus grandes contraintes de rapidité de traitement, n'a besoin de conserver en mémoire que des informations dynamiques minimales, correspondant aux identifiants des nœuds qu'on rencontre dans l'arbre en remontant jusqu'à la racine à partir de l'événement courant. Les règles syntaxiques du protocole considéré permettent, en examinant la suite du flux, de se rattacher au nœud approprié. Les informations relatives aux nœuds frères ou cousins de l'événement courant dans l'arbre n'ont pas besoin d'être conservées par le collecteur 1.

[0049] Il faut noter que les procédés et dispositifs selon l'invention ne se limitent nullement au protocole HTTP qui n'a été mentionné que pour illustrer en termes concrets un exemple de réalisation.

[0050] Pour donner un autre exemple, on peut considérer un protocole applicatif de messagerie, tel SMTP. Un flux de ce protocole entre un émetteur et un récepteur donne lieu à des messages successifs (*email* dans la représentation en arbre de la figure 5) ayant chacun un objet (*subject*), un émetteur (*sender*) ayant une adresse de messagerie (*sender_email*) et un pseudonyme éventuel (*sender_alias*), un ou plusieurs destinataires (*receiver*) ayant chacun une adresse de messagerie (*receiver_email*) et un pseudonyme éventuel (*receiver_alias*), un corps de message (*message*) et d'autres champs éventuels (indicateurs d'urgence, de confidentialité, pièce(s) jointe(s), etc.). Ces différents champs d'un message sont repérables par l'analyseur syntaxique 13 si on lui fournit les règles syntaxiques appropriées du protocole employé. Ils sont souvent transmis dans des paquets IP différents en fonction

de la segmentation opérée au niveau de l'émetteur.

[0051] La figure 5 détaille les ramifications dans l'arbre que peut construire l'analyseur de données structurées 7 pour représenter un exemple de message simple ayant deux destinataires. On suppose que les éléments placés dans le cadre en traits interrompus sont extraits d'un premier paquet P1' et que ceux placés dans le cadre en traits mixtes sont extraits d'un deuxième paquet P2' (qui peut être suivi d'un ou plusieurs autres paquets contenant une suite du corps du message, des pièces jointes, etc.).

[0052] Dans cet exemple, l'étape de décomposition 27 donne lieu à sept événements 1.1-1.7 pour le premier paquet P1' et quatre événements 2.1-2.4 pour le deuxième paquet P2'. À l'étape 29, l'analyseur syntaxique 13 affecte des identifiants aux nœuds successivement détectés (représentés par des ellipses sur la figure 5) et rattache ces nœuds ainsi que les feuilles (représentées par des rectangles) à des nœuds pères respectifs. Ces identifiants et ces rattachements sont transmis en tant qu'indicateurs d'états arborescents à l'analyseur externe 7 qui reconstituera tout ou partie de l'arbre représenté sur la figure 5 si l'application le requiert.

[0053] On observe qu'après avoir traité le premier paquet P1' qui se termine par l'adresse d'un destinataire (événement 1.7), l'analyseur syntaxique 13 du collecteur de données 1 n'a besoin de conserver en mémoire que l'identifiant du nœud père (1.6) de cet événement, pour le cas où un paquet ultérieur contiendrait un pseudonyme du destinataire en question, l'identifiant du nœud grand-père (1.1), pour le cas où le paquet suivant du flux commencerait par le corps du message, la date, une pièce jointe ou autre et l'identification du nœud racine pour le cas où un nouveau message apparaîtrait dans le flux.

[0054] La méthode ci-dessus de représentation arborescente des données transmises dans un flux et de transmission d'une description structurelle minimale des éléments pouvant se placer dans l'arbre est applicable à tout type de protocole à partir d'une description des règles syntaxiques pertinente pour le protocole considéré.

[0055] L'invention a été illustrée et décrite en détail dans les dessins et la description précédente. Celle-ci doit être considérée comme illustrative et donnée à titre d'exemple et non comme limitant l'invention à cette seule description. De nombreuses variantes de réalisation sont possibles.

REVENDICATIONS

1. Dispositif de collecte de données pour la surveillance de flux dans un réseau de données (3) utilisant un mode de transmission par paquets, le dispositif comprenant:
 - 5 • un extracteur (11) de données contenues dans des paquets appartenant à un flux défini par un émetteur, un récepteur et un protocole;
 - un analyseur syntaxique (13) agencé pour:
 - recevoir les données en temps réel de l'extracteur;
 - décomposer les données reçues en éléments en fonction de règles
10 syntaxiques dudit protocole, les règles syntaxiques permettant de représenter lesdits éléments suivant une structure en arbre; et
 - associer des indicateurs d'états arborescents respectifs à certains au moins desdits éléments, où l'indicateur d'état arborescent associé à un élément repère ledit élément dans la structure en
15 arbre; et
 - une interface (15) pour transmettre à un analyseur de flux (7) les indicateurs d'états arborescents accompagnés des éléments auxquels ils ont été associés.
2. Dispositif selon la revendication 1, dans lequel l'indicateur d'état
20 arborescent associé à un élément comporte un identifiant d'un nœud père de la structure en arbre auquel se rattache un emplacement recevant ledit élément dans la représentation suivant la structure en arbre.
3. Dispositif selon la revendication 2, dans lequel l'analyseur
syntaxique (13) est en outre agencé pour:
 - 25 - affecter un identifiant de nœud à un élément courant s'il est déterminé selon les règles syntaxiques que l'élément courant est reçu, dans la représentation suivant la structure en arbre, à un emplacement situé à un nœud de la structure en arbre; et

- 17 -

- Inclure l'identifiant de nœud affecté à l'élément courant dans l'indicateur d'état arborescent associé audit élément courant.

4. Dispositif selon la revendication 3, dans lequel l'analyseur syntaxique (13) est agencé pour ne conserver en mémoire, pour un flux donné
5 après transmission d'un élément et de son indicateur d'état arborescent associé, que chaque identifiant de nœud rencontré en remontant la structure en arbre à partir de l'emplacement recevant ledit élément dans la représentation suivant la structure en arbre.
5. Système de surveillance de flux dans un réseau de données (3)
10 utilisant un mode de transmission par paquets, le système comprenant un dispositif de collecte de données (1) selon l'une quelconque des revendications précédentes et un analyseur de flux (7) recevant les éléments transmis avec les indicateurs d'états arborescents associés.
6. Système selon la revendication 5, dans lequel l'analyseur de flux (7)
15 est agencé pour construire une partie au moins de la structure en arbre à partir des éléments transmis avec les indicateurs d'états arborescents associés.
7. Procédé de surveillance de flux dans un réseau de données (3) utilisant un mode de transmission par paquets, le procédé comprenant:
 - extraire (25) des données contenues dans des paquets appartenant à un
20 flux défini par un émetteur, un récepteur et un protocole;
 - décomposer (27) en éléments les données reçues en temps réel, en fonction de règles syntaxiques dudit protocole, les règles syntaxiques permettant de représenter lesdits éléments suivant une structure en arbre;
 - 25 • associer (29) des indicateurs d'états arborescents respectifs à certains au moins desdits éléments, où l'indicateur d'état arborescent associé à un élément repère ledit élément dans la structure en arbre; et
 - transmettre (31) à un analyseur de flux (7) les indicateurs d'états arborescents accompagnés des éléments auxquels ils ont été associés.

- 18 -

8. Procédé selon la revendication 7, dans lequel l'Indicateur d'état arborescent associé à un élément comporte un identifiant d'un nœud père de la structure en arbre auquel se rattache un emplacement recevant ledit élément dans la représentation suivant la structure en arbre.
- 5 9. Procédé selon la revendication 8, comprenant en outre:
- affecter un identifiant de nœud à un élément courant s'il est déterminé selon les règles syntaxiques que l'élément courant est reçu, dans la représentation suivant la structure en arbre, à un emplacement situé à un nœud de la structure en arbre; et
- 10 • Inclure l'identifiant de nœud affecté à l'élément courant dans l'Indicateur d'état arborescent associé audit élément courant.
10. Procédé selon la revendication 9, dans lequel, pour un flux donné après transmission d'un élément et de son indicateur d'état arborescent associé, n'est conservé en mémoire que chaque identifiant de nœud rencontré
- 15 en remontant la structure en arbre à partir de l'emplacement recevant ledit élément dans la représentation suivant la structure en arbre.
11. Procédé selon l'une quelconque des revendications 7 à 10, comprenant en outre la construction d'une partie au moins de la structure en arbre au niveau de l'analyseur de flux (7) à partir des éléments transmis avec
- 20 les Indicateurs d'états arborescents associés.